

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



**الجامعة الإسلامية - غزة  
شؤون البحث العلمي والدراسات العليا  
كلية تكنولوجيا المعلومات  
ماجستير تكنولوجيا المعلومات**

## **Arabic SMS Spam Detection Based on Semantic Classification**

**كشف الرسائل العربية القصيرة المزعجة اعتماداً على التصنيف  
الدلالي**

**Ayman M. N. Abu Ouda**

**Supervised by**

**Dr. Rebhi Baraka**

**Associate professor of Computer Science**

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

**March/2017**

## إقرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

# Arabic SMS Spam Detection Based on Semantic Classification

## كشف الرسائل العربية القصيرة المزعجة اعتمادا على التصنيف الدلالي

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

### Declaration

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

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

Student's name:	أيمن محمد نجيب أبو عودة	اسم الطالب:
Signature:	أيمن أبو عودة	التوقيع:
Date:	21/03/2017	التاريخ:



## نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحث/ ايمن (محمد نجيب) اسماعيل ابو عودة لنيل درجة الماجستير في كلية تكنولوجيا المعلومات برنامج تكنولوجيا المعلومات وموضوعها:

### كشف الرسائل العربية القصيرة المزعجة اعتماداً على التصنيف الدلالي

### Arabic SMS Spam Detection Based on Semantic Classification

وبعد المناقشة التي تمت اليوم الثلاثاء 21 رجب 1438هـ، الموافق 2017/04/18م الساعة الحادية عشر صباحاً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

.....

مشرفاً و رئيساً

د. ربحي سليمان بركة

.....

مناقشاً داخلياً

د. توفيق سليمان برهوم

.....

مناقشاً خارجياً

د. يوسف نبيل أبو شعبان

وبعد المداولة أوصت اللجنة بمنح الباحث درجة الماجستير في كلية تكنولوجيا المعلومات / برنامج تكنولوجيا المعلومات.

واللجنة إذ تمنحه هذه الدرجة فإنها توصيه بتقوى الله ولزوم طاعته وأن يسخر علمه في خدمة دينه ووطنه.

والله ولي التوفيق،،،

نائب الرئيس لشئون البحث العلمي والدراسات العليا



د. عبدالرؤف علي المناعمة

## **Abstract**

Short Messaging Service (SMS) Spam is unwanted messages sent over the web or mobile system to mobile phone devices. SMS is attractive for spammers due to its cheap services, easily deciding the destination country and its higher response rates than email. Existing solutions to this issue are no longer adequate as they are either costly in terms of resources, inefficient, most of the existing detection techniques for SMS spam have been adapted from other contexts such as email spam detection methods. Spammers are constantly developing more sophisticated tactics causing previous methods for spam detection as ineffective. Additionally, when it comes to Arabic SMS messages, most SMS spam filtering system based on English language.

This research presents an Arabic SMS spam detection and classification approach using ontology with semantic rules. An Arabic SMS spam ontology with a support of Arabic WordNet is built by defining spam classes and hierarchy and adding a collection of various spam messages as instances creating a knowledge base reflecting the domain. To enable the detection and classification of messages based on the knowledge base, a set of SWRL rules were written. These rules are used by the reasoner to filter out messages as spam or legitimate. Based on the enriched knowledge base, an SMS spam detection system is built. It consists of several modules such as query module, reasoning module, synonym module, SMS module and finally classifier module.

The approach is evaluated based on its ability to classify and detect SMS messages as spam or legitimate. A number of performance measures are used for this purpose. The evaluation resulted in an accuracy of 96.5% and in a f-measure of 90.5% which are better than those achieved using a traditional classifier such as Naïve Bayes.

**Keywords:** SMS spam filtering, Arabic SMS spam ontology, text classification, semantic rules, reasoning.

## الملخص

تعرف رسائل الـ (Spam) بأنها الرسائل الإلكترونية المزعجة والغير مرغوب فيها والتي ترسل من خلال الانترنت (الويب) أو من وإلى أجهزة الهاتف المحمول ، وقد جذبت الرسائل القصيرة المزعجة المتطفلين لاستخدامها بشكل أكبر من البريد الإلكتروني المزعج ؛ وذلك لعدة أسباب منها رخص هذه الطريقة في الاحتيال على الزبائن ؛ حيث باستطاعتهم إرسال مجموعات هائلة من الرسائل الموجهة كإرسالها لمشتري بلد معين، بالإضافة إلى أن استجابة المشتركين للرسائل القصيرة أكبر من استجابتهم للبريد الإلكتروني وذلك لاعتقادهم بالموثوقية العالية للرسائل القصيرة. معظم الحلول المطروحة حالياً لعلاج قضية الرسائل المزعجة لم تعد كافية حيث أنها مكلفة في بعض الأحيان، أو غير فعالة في أحيان أخرى، فمعظم التقنيات الحالية لكشف الرسائل القصيرة المزعجة قد تجاوزها المتطفلون، علاوة على تجدد أساليبهم في إرسال الرسائل القصيرة المزعجة، مما يحد من فعالية هذه الحلول، ناهيك عن أن معظم تلك الحلول موجهة لخدمة لغات غير اللغة العربية.

نقترح - في هذه الرسالة - طريقة جديدة لكشف الرسائل القصيرة العربية المزعجة اعتماداً على التصنيف الدلالي، وذلك باستخدام أنطولوجيا صممناها خصوصاً لهذا المجال (SMS spam domain ontology) بالإضافة إلى استخدام القواعد الدلالية (semantic rules). تم إنشاء أنطولوجيا للرسائل القصيرة العربية المزعجة بدعم من WordNet العربية من خلال تحديد فئات الرسائل المزعجة والتسلسل الهرمي وإضافة مجموعة من الرسائل القصيرة المزعجة المختلفة لإنشاء قاعدة المعرفة التي تعكس المجال. ولتمكين الكشف وتصنيف الرسائل استناداً إلى قاعدة المعرفة، تمت كتابة مجموعة من القواعد الدلالية بلغة SWRL. بحيث يتم استخدام هذه القواعد من قبل المسبب (reasoner) لتصنيف الرسائل كرسائل غير مرغوب فيها أو مشروعة. استناداً إلى قاعدة المعرفة المعززة هذه، تم إنشاء نظام ويب للكشف عن الرسائل غير المرغوب فيها (SMS spam) وهو يتألف من عدة وحدات مثل وحدة الاستعلام، وحدة المنطق، وحدة المرادفات، وحدة إرسال الرسائل وأخيراً وحدة المصنف.

تم تقييم النظام بناءً على قدرته على تصنيف وكشف الرسائل القصيرة كرسائل غير مرغوب فيها أو مشروعة. حيث تم استخدام عدد من مقاييس الأداء لهذا الغرض. وأسفر التقييم عن دقة (Accuracy) بلغت 96.5% وفي مقياس (f-measure) 90.5% وهي أفضل من تلك التي تحققت باستخدام مصنف تقليدي مثل (Naïve Bayes).

**كلمات مفتاحية:** الرسائل القصيرة المزعجة، أنطولوجيا، القواعد الدلالية، تصنيف النص، التسبيب.

## Dedication

*To my beloved parents*

*To my brothers and sisters*

*To my dear wife and children*

*To those who gave me support*

*To all of them I dedicate this work*

## **Acknowledgment**

Thanks and praise be to Allah Almighty for guidance and help to complete this research. I would like to express my deepest gratitude to Dr. Rebhi Baraka for his patience, support, and great ideas without which, this study would not have been possible.

Special thanks to my father and my mother for their prayer, patience, motivation, and continued support. I am very grateful to my dear wife without her encouragement I could not do this work.

Last but not least, I would like to express my gratitude to my family and friends who supported me during this work.

**Ayman Mohammed Abu Ouda**

**March 2017**

## Table of Contents

<b>Declaration .....</b>	<b>I</b>
<b>Abstract.....</b>	<b>II</b>
<b>الملخص.....</b>	<b>III</b>
<b>Dedication .....</b>	<b>IV</b>
<b>Acknowledgment.....</b>	<b>V</b>
<b>Table of Contents .....</b>	<b>VI</b>
<b>List of Tables .....</b>	<b>IX</b>
<b>List of Figures.....</b>	<b>X</b>
<b>List of Abbreviations .....</b>	<b>XII</b>
<b>Chapter 1 Introduction .....</b>	<b>2</b>
1.1 Background and Context .....	2
1.2 Statement of The Problem .....	4
1.3 Objectives .....	4
1.3.1 Main Objective.....	4
1.3.2 Specific Objectives .....	5
1.4 Significance of The Research .....	5
1.5 Scope and Limitations .....	5
1.6 Research Methodology .....	6
1.7 Overview of The Research .....	8
<b>Chapter 2 Theoretical and Technical Foundation.....</b>	<b>11</b>
2.1 Short Message Services .....	11
2.1 Spam in Different Media .....	13
2.3 SMS Spam .....	13
2.4 SMS Spam Filtering Methods.....	14
2.5 Semantic Web .....	15
2.6 Ontology .....	15
2.7 Ontology Building .....	17
2.7.1 Determine The domain of Ontology and Scope.....	17
2.7.2 Reuse Existing Ontologies .....	17
2.7.3 List Important Terms in The Ontology .....	18
2.7.4 Define The Classes and Subclasses .....	18
2.7.5 Define The Properties .....	18
2.7.6 Define The Facets of The Slots.....	18
2.7.7 Create Instances .....	18



2.8 Resource Description Language (RDF).....	19
2.9 SPARQL .....	20
2.10 DL Query .....	21
2.11 Protégé .....	21
2.12 Reasoning.....	21
2.13 Semantic Rules .....	22
2.14 WordNet.....	23
2.15 Evaluation Method.....	23
2.16 Summary .....	26
<b>Chapter 3 Related Works.....</b>	<b>28</b>
3.1 SMS Spam Detection Methods.....	28
3.2 Using Ontology in Email Spam Detection .....	32
3.3 Using Ontology in SMS Spam Detection .....	35
3.4 Summary .....	36
<b>Chapter 4 Arabic SMS Spam Ontology.....</b>	<b>38</b>
4.1 Determining The domain of Ontology and Scope .....	38
4.2 Reuse Existing Ontologies.....	40
4.3 Overview of The Ontology .....	40
4.4 List The Important Terms in SMS Spam.....	41
4.5 Define Classes and Subclasses of SMS Spam .....	42
4.6 Define The Properties of Classes.....	44
4.7 Define The Facets of The Slots.....	47
4.8 Create Instances of Spam Words .....	49
4.9 Evaluate Ontology .....	49
4.10 Summary .....	54
<b>Chapter 5 Arabic SMS Spam Detection .....</b>	<b>57</b>
5.1 Overall Structure of The Approach .....	57
5.2 Functionality of The SMS Spam Detection Approach .....	59
5.3 Data Collection .....	61
5.4 Data Preprocessing .....	61
5.5 Word Extraction, Matching with WordNet .....	63
5.6 Building The Ontology .....	65
5.7 Create Semantic Rules .....	65
5.8 Apply Ontology Reasoner .....	69
5.9 Adding New Relations to Ontology.....	71
5.10 Querying .....	73

5.11 Classification .....	74
5.12 Summary .....	75
<b>Chapter 6 Results and Discussion .....</b>	<b>76</b>
6.1 Experiments .....	77
6.2 Evaluation Metrics .....	78
6.3 Evaluation Results .....	78
6.4 Comparison with Other Works .....	84
6.5 Summary .....	85
<b>Chapter 7 Conclusions and Future Work .....</b>	<b>87</b>
<b>References .....</b>	<b>90</b>

## List of Tables

Table (2.1): Confusion Matrix .....	24
Table (4.1): Ontology metrics.....	41
Table (4.2): The Arabic SMS ontology classes and subclasses.....	42
Table (4.3): Object properties of the ontology classes .....	45
Table (4.4): Data properties of the ontology classes .....	45
Table (6.1): SMS data set .....	77
Table (6.2): Confusion Matrix results .....	78
Table (6.3): Experimental Results for different measures.....	82
Table (6.4): Comparison with the Naïve Bayes classifier .....	85

## List of Figures

Figure (2.1): SMS network basic scheme (Ortiz & Prieto, 2004) .....	12
Figure (2.2): SMS spam asking to update Facebook account through fishing URL .....	14
Figure (2.3): Layers of languages used for the semantic web (Sugumaran & Gulla, 2011) .....	16
Figure (2.6): RDF triples showing relationships between an employee, book, and price (Sajja & Akerkar, 2012).....	19
Figure (4.1): Main classes in the SMS spam ontology .....	40
Figure (4.2): The class hierarchy of the Arabic SMS spam ontology .....	44
Figure (4.3): An Example of data property “عدد_الرسائل_المحظوره” .....	46
Figure (4.4): Object properties shown in Protégé.....	46
Figure (4.5): Data properties shown in Protégé.....	46
Figure (4.6): Creating data restriction.....	47
Figure (4.7): Example of different data types .....	48
Figure (4.8): Example of allowed values of slots .....	48
Figure (4.9): Example of cardinality.....	48
Figure (4.10): List of some ontology instances .....	49
Figure (4.11): Query for all spam words related to the verb “اربح” (win) .....	50
Figure (4.12): Justification result of query spam words related to the verb “اربح” (win).....	51
Figure (4.13): Query for all political synonym spam words of “مسيره” (march) ...	51
Figure (4.14): Justification result of query political individual have relation with “له_معنى” (has_synonym) with “مسيره” (march) .....	52
Figure (4.15): Query for all SMS classified as spam.....	52
Figure (4.16): Justification result of query classified as spam .....	53
Figure (4.17): Query for all synonym spam words of “تجمع” .....	53
Figure (4.18): Query for all blocked sender name.....	54
Figure (5.1): Structure of the SMS spam detection and classification approach... 57	57
Figure (5.2): User interface for sending SMS messages .....	59
Figure (5.3): SMS spam detection .....	60
Figure (5.3): Part of the SMS dataset .....	61
Figure (5.4): Replacing some characters in the process of tokenization .....	62
Figure (5.5): Examples of stop words.....	62
Figure (5.6): Setting the probability weight for spam words.....	64
Figure (5.7): Getting synonym of words from Arabic WordNet to enrich the SMS spam Knowledge base .....	64
Figure (5.8): Adding synonym of a word from Arabic WordNet to the ontology	65
Figure (5.9): Some rules to classify messages using ontology terms and reasoning .....	68
Figure (5.10): Writing a rule in JENA.....	69
Figure (5.11): Results of reasoner use semantic rules .....	70
Figure (5.12): Explanation of the inferred message classification with the used semantic rule .....	70
Figure (5.13): Adding rank (data property) for sender name .....	71
Figure (5.14): Data property “عدد_الرسائل_المحظورة” .....	71

Figure (5.15): User interface for SMS provider to adding new words to the SMS ontology .....	72
Figure (5.16): Adding new words and relations to the ontology using object and data properties.....	72
Figure (5.17): Interface to show all blocked sender names in the knowledge base	73
Figure (5.18): Retrieving all blocked sender names from the knowledge base.....	73
Figure (5.20): Running the phishing semantic rule to classify a message.....	74
Figure (5.21): SPARQL query to return a messaged classified as spam .....	74
Figure (6.1): Accuracy comparison for three different expermintns .....	82
Figure (6.2): Recall rates comparison for three different expermintns.....	84
Figure (6.3): MCC comparison for three different expermintns .....	84

## List of Abbreviations

<b>API</b>	Application programming interface
<b>BSC</b>	Base Station Controller
<b>BSS</b>	Base Station System
<b>HLR</b>	Home Location Register
<b>IWMSC</b>	Inter-Working Mobile Switching Centre
<b>MCC</b>	Matthews Correlation Coefficient
<b>MS</b>	Mobile Station
<b>MSC</b>	Mobile Switching Center
<b>OWL</b>	Web Ontology Language
<b>RDF</b>	Resource Description Framework
<b>SME</b>	Short Message Entity
<b>SMPP</b>	Short Message Peer to Peer
<b>SMS</b>	Short Message Service
<b>SMSC</b>	Short Message Service Centre
<b>SPARQL</b>	SPARQL Protocol and RDF Query Language
<b>SS7</b>	Signaling System Number 7
<b>SWRL</b>	Semantic Web Rule Language
<b>VAS</b>	Value-Added Service
<b>VLR</b>	Visitor Location Register

# **Chapter 1**

## **Introduction**

# Chapter 1

## Introduction

### 1.1 Background and Context

Recently, Short Message Service (SMS) has evolved into one of the most communication used due to the rapid growth in the number of mobiles worldwide. According to the Global System Mobile Association (GSMA), Palestine has two mobile operators, Jawwal and Wataniya with 3.3 million mobile connections (GSMA, 2015).

This increase has enticed spammers and caused SMS spam problem such as e-mail spam. Recent reports indicate that the volume of SMS spam messages was increasing every year (GSMA, 2015).

SMS spam problem is more critical problem than email spam due to mobile phones are very personal devices. Users may have multiple email accounts, but usually have one mobile phone.

Spammers are using targets' mobile phones to break into accounts and steal personal information, in so-called 'smishing'. Some SMS has links provided in the message which links can install malware on mobile and to spoof sites that look real but whose purpose is to steal personal information. spam SMS often uses the promise of free gifts, like computers or gift cards, or product offers, like cheap mortgages, credit cards, or debt relief services to get you to reveal personal information, like how much money you make, how much you owe, or your bank account information, credit card number, or Social Security number. Clicking on a link in the message can install malware that collects information from your phone. Once the spammer has your information, it is sold to marketers or, worse, identity thieves (FTC, 2013).

SMS spam differs from email spam in other attributes. Email spam is identifiable by its structure. SMS spam detection and filtering is a relatively new task, which we can inherit the SMS spam issues and solutions from email spam detection and filtering.



Normally, mobile phone operators provide SMPP connection to SMS providers through an internet service point and providers in turn provide their customers with access Value-Added Service (VAS), e.g. SMS messaging. SMS providers sell BulkSMS for end users, spammers can buy BulkSMS from providers and send SMS to thousands of users. Therefore, providers should have an efficient anti-spam tools to catch and detect any unsolicited SMS to prevent it from operators.

Currently there is much work on SMS spam filtering using techniques such as Black and White List, Text Classification (Taufiq, Abdullah, Kang, & Choi, 2010), Boyer and Moore (BM) (Liu, Ke, & Zhang, 2010), Naïve Bayesian classifiers (Zhang & Wang, 2009) (Deng & Peng, 2006), neural networks (Anchal 2014), and frame model of ontology-based detection (Balubaid & Manzoor, 2015) to name a few.

Spammers always try to find method to bypass current filters, the new filters need developing for more effective spam filtering. Ontologies can be a basis for such sharing SMS providers and developers where they allow for machine to understand the semantics of data.

Ontology is a semantic web concept that can be used for decision support and information retrieval systems (Kalfoglou, 2007). Based on these concepts, ontology can also be helpful in SMS spam detection (Balubaid & Manzoor, 2015) (Noy & McGuinness, 2001) .

Ontology is defined as "a formal explicit description of concepts in a domain of discourse. Properties of each concept describe various features and attributes of the concept, and restrictions on slots. Ontologies together with a set of individual instances of classes constitute a knowledge base" (Noy & McGuinness, 2001). There are multiple languages such as RDF, RDF-schema, and OWL that can be used to represent and build ontology.

In this research, we propose to use ontology to help in detecting spam in Arabic SMS. We design an Arabic SMS spam ontology as a set of classes, properties, and relationships. The Arabic SMS spam ontology is the main core of the approach to

detect SMS spam. We collected SMS dataset from local SMS providers, and adds it to the ontology to establish an SMS spam knowledge base.

Subsequently we build an SMS spam detection approach that consists of several modules such as query module, reasoning module, synonym module, SMS module and finally classifier module. These modules are connected and are dependable on the knowledge base (ontology and instances of SMS spam). We develop a prototype of the approach to test its ability to detect SMS spam with high accuracy.

Next, we state the research problem and derive its objectives followed by the significance of the research, the scope and limitation of the research. We then put up the methodology to be followed to achieve the research objectives. Finally, we give an overview of the rest of the thesis.

## **1.2 Statement of The Problem**

SMS spam is one of the critical malicious activities worldwide such as SMS spoofing, scam, virus links, waste of time, where spammers are constantly developing more sophisticated tactics that makes current SMS spam detection approaches and methods no longer effective.

The SMS spam detection based on the semantic web remains largely unexplored especially at the server side where the ontology based processing and reasoning can be used in detecting SMS spam and classify it as spam or legitimate.

The problem of this research is how to develop an efficient ontology-based approach for detecting spam in Arabic content of SMS and classify it as spam or legitimate.

## **1.3 Objectives**

### **1.3.1 Main Objective**

The main objective of this research is to develop an efficient approach based on ontology to detect Arabic SMS spam with high accuracy in spam message filtering.

### **1.3.2 Specific Objectives**

**The specific objectives of the research are:**

- To collect SMS corpus containing spam from SMS providers.
- To build ontology and the knowledge base of Arabic SMS spam based on the collected SMS corpus.
- To develop the approach for detecting spam in Arabic SMS based on the built ontology.
- To use semantic relations and add rules to the ontology to detect SMS spam.
- To conduct various experiments to evaluate on the proposed approach based on ontology for efficient detection of spam. Efficiency measure is based on the accuracy of classifying SMS message as spam or legitimate.

### **1.4 Significance of The Research**

- This work is important to explore the use of ontology in detecting Arabic content of SMS spams.
- The design of the ontology would allow to add any new instances of SMS spam domain leading to a knowledge base that can be used in other type of SMS spam and for other purposes.
- It encourages using knowledge of Arabic SMS spam from other SMS providers to protect their systems from spams and integrate it in BulkSMS systems.
- Improve the efficiency of ontology for classification SMS.
- Protect subscribers of mobile operators from SMS spam as well as protect SMS providers from sending SMS spam to operator leading to more trust and confidence.

### **1.5 Scope and Limitations**

- The approach serves only Arabic language content of SMS and not work for trickery of special characters.

- The approach would be used at BulkSMS provider's server side before SMS can reach the operators who in turn send these spammed messages to the end user mobile handsets.
- HTTP protocol would be used for sending SMS between clients and providers.
- The system would be at the server side of the SMS providers while end users do not need to make any installation of the spam detection system on their devices.
- The efficiency of approach will focus on the accuracy of the classification but not on the speed of the classification because in BulkSMS and advertisements SMS the important factor is the delivery of the messages, rather the time of delivery.
- The ontology will not cover all Arabic SMS spam, it will cover selected domain corpus of SMS. This is to insure the correct function of the system and ensures the ability to test and evaluate the results based on the domain. Doing the same for other domains can follow based on the results of the approach and based on the selected domain.

## 1.6 Research Methodology

To accomplish the objectives of this research, the following methodology will be followed:

1. **Research and survey:** Review of recent related works to the research problem especially in the SMS spam filters. Upon analyzing the existing solutions, which can support us to formulate a solution to the problem.
2. **Data collection:** We will collect a corpus of SMS spams from SMS providers in Palestine. In this phase, we will select the appropriate provider and the nature and size of the SMS messages.
3. **Data processing:** Some preprocessing in Arabic SMS corpus is performed. It includes applying stop words removal, tokenizing strings to words and applying suitable term stemming. This process is necessary for maintaining the knowledge base which consists of the ontology and the RDF store. The ontology refers to xml and is often stored in a file.

4. **Word extraction and matching relation:** After data processing, we need to extract spam words from WordNet to enrich ontology with vocabulary of SMS spam, then we make relation between them by make object and data properties between these words in the ontology and set probability weights of spam words.
5. **Building the SMS spam knowledge base:** The SMS spam knowledge base consists of two parts namely, the SMS spam ontology and the instances (individuals) which enrich the ontology and enable the approach to detect the spam messages. To build the ontology and the knowledge base, we follow the ontology building process (Boyce & Pahl, 2007; Noy & McGuinness, 2001) using some tools such as OWL language, SPARQL Query, and Protégé (Protégé, 2016) that includes the following tasks:
  - A. Determine domain of the ontology and scope.
  - B. List the important terms in the ontology.
  - C. Define classes.
  - D. Define properties.
  - E. Define facets of the slots.
  - F. Create instances of SMS spams.
  - G. Apply reasoner to get new facts for SMS spams.
  - H. Execute some queries on the ontology to ensure the correct building of ontology and correct retrieval of information and checks whether it returns what we expect.
6. **Creating semantic rules to our domain ontology:** after build ontology and make relation between instances, we need to create rules that achieve detection and classification of SMS spams.
7. **Applying reasoner and querying:** we need to apply reasoner to get new facts for SMS spams, and use SPARQL query to show new facts after applying reasoner.
8. **Developing a prototype for the proposed approach:** we develop system for The approach which contain the interface and using programming language such as JAVA and related APIs and tools such as JENA. Then combine and concept the knowledge based with the ontology.

- 9. Evaluate the Approach:** we evaluate the proposed approach. We will analyze the obtained results and evaluate the accuracy of SMS spam detection. Specific techniques to perform this evaluation related specifically to SMS spam can be selected and used based on the experiments.

## 1.7 Overview of The Research

The research consists of seven chapters organized as follows:

- **Chapter 1 (Introduction):** An introduction stating the problem of the research objectives, scope, significance, limitation and the research methodology.
- **Chapter 2 (Theoretical and technical foundation):** Describe a List of the theoretical and technical foundation needed for the research work such as Short Message Services (SMS), SMS spam, semantic web, ontology concepts, ontology building, ontology tools, and evaluation techniques.
- **Chapter 3 (Related Works):** Reviews several approaches and works of SMS spam filtering using techniques such as Black and White List, Text Classification, Boyer and Moore (BM), Naïve Bayesian classifiers, neural networks.
- **Chapter 4 (Arabic SMS Spam Ontology):** Presents the steps to develop the SMS spam domain ontology and then it presents the evaluation of the ontology.
- **Chapter 5 (Arabic SMS Spam Detection):** Presents and discuss the steps of analyzing, designing and developing the prototype of the approach. It presents the structure of the proposed approach, collection of SMS data, creating semantic rules, developing the parts of the prototype and finally the system functions.

- **Chapter 6 (Results and Discussion):** Presents the experiments of proposed ontology and prototype and discuss results of experiments.
- **Chapter 7 (Conclusion and The Future Works):** Presents the research conclusions and the future works.

# **Chapter 2**

## **Theoretical and Technical Foundation**



## **Chapter 2**

### **Theoretical and Technical Foundation**

In this chapter, we present the theoretical and technical foundation of the proposed approach. We talk firstly about the Short Message Services and SMS spam, then we present short description of the Semantic web and ontology, enumerate the tools that we used to develop the proposed approach, and finally we talk about evaluation method.

#### **2.1 Short Message Services**

The short messaging service (SMS) is a bi-directional service to send text over wireless communication systems. It consists of a message that can be up to 160 alphanumeric characters in text and 70 alphanumeric in Unicode characters (Harrington, 2008).

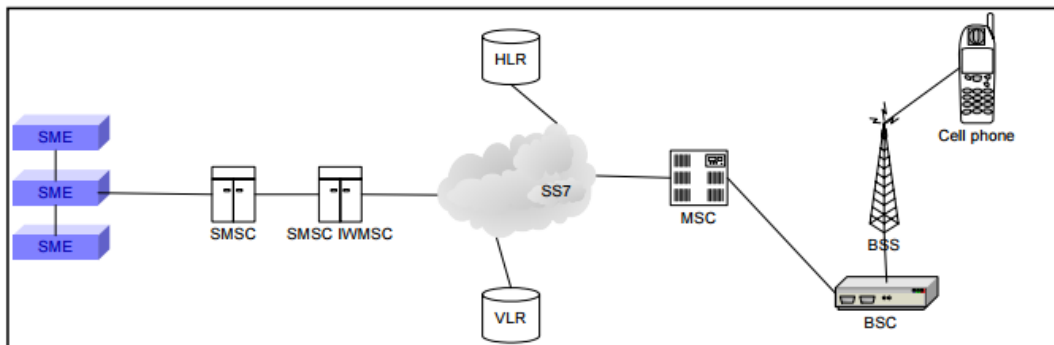
SMS has been existence from the second generation (2G) until present of fourth generation (4G) GSM mobile (Pereira & Sousa, 2004). This GSM data service has established the simplest one-to-one communication by exchanging short text messages. Now SMS has been the most popular messaging service due to the low cost of SMS, network reliability has made sending SMS messages an economic option for GSM subscribers (Yoon, Kim, & Huh, 2010).

In a GSM Network, Short Messages are sent over SS7 (Signaling System Number 7) network.

The network elements that are directly related with SMS service as shown in Figure (2.1) are as follows (Ortiz & Prieto, 2004)

- Short Message Entity (SME): is any entity which is capable to send and/or receive short messages.
- Short Message Service Centre (SMSC): element that sends or store and forward messages from a SME to a Mobile Station (MS)
- Inter-Working Mobile Switching Centre (IWMSC): Gateway node for short messages originated with a mobile on that network.

- Home Location Register (HLR): is a database used for storage and management of subscriptions, which informs the SMSC of initiated unsuccessful short message delivery attempts to a specific mobile station.
- Visitor Location Register (VLR): is a database that contains temporary information about subscribers. This information is needed by the MSC in order to service visiting subscribers.
- Signalling System No. 7 (SS7): is telephony signalling protocols used to set up and tear down most of the world's public switched telephone network (PSTN) telephone calls. It also performs number translation, local number portability, prepaid billing, Short Message Service (SMS), and other mass market services.
- Mobile Switching Centre (MSC): make the switching functions of the system and control the calls to and from other phone and data systems.
- Base Station Controller (BSC): manage the radio resources and controls items such as handover within the group of Base Transceiver Station (BTS).
- Base Station System (BSS): responsibility is to transmit voice and data traffic between the mobile stations.



**Figure (2.1):** SMS network basic scheme (Ortiz & Prieto, 2004)

The SMS sending process can be summarized as follows (Ortiz & Prieto, 2004):

- The SMS is sent from SME to SMSC.
- The SMSC communicates with the HLR and retrieves the necessary routing information to get through to the receiver.
- The SMSC/IWMSC sends the SMS to the MSC.
- The MSC extracts the receiver information from the VLR.
- The MSC transfers the SMS to the receiver.

- The MSC returns the results of the transmission operation to the SMSC.
- If the SME asks for a confirmation, the SMSC will send back a message with the transmission operation result.

The SMS receiving process can be summarized as follows (Ortiz & Prieto, 2004):

- The mobile phone transfers the SMS to the MSC.
- The MSC asks the VLR to verify if any network restriction is being overridden.
- The SMSC sends the SMS to the mobile phone.
- The SMSC acknowledges a successful transmission.
- The MSC sends the result of the operation to the mobile phone.
- All GSM network elements have to be taken as black boxes, with location, routing, forwarding etc., capabilities.

SMS Components are (Harrington, 2008):

- Length of SMS.
- Service Centre Timestamp.
- Originator address: the phone number of the sender.
- Protocol identifier.
- Data coding scheme.
- User Data Length: tells how long the message is
- User Data: the message itself.

## **2.1 Spam in Different Media**

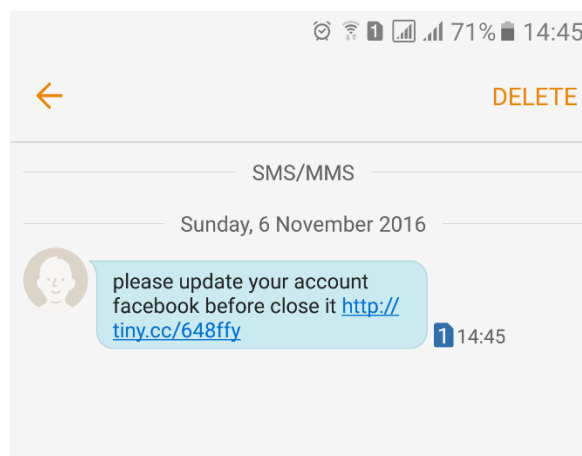
Spam exists in different media such as SMS, email, instant message, use net newsgroup, social media, search engines and internet telephony. The technical differences between all these media makes spam in general too complex for one overview (Blanzieri & Bryl, 2008). Next, we elaborate in SMS spam since it is the focus of this research.

## **2.3 SMS Spam**

Mobile SMS spam, also known as SMS spam is any unsolicited, unwelcome text message sent to a mobile device. These messages often promote unwanted products

and services, or try to trick recipients into providing personal information. SMS Spam include things such as “win free stuff scam”, “payday loan scam”, “debit relief scam”, “adult content”, “political or religious incitement” (GSMA, 2013). Lately the industry has seen an increase in fraudulent spam attempting to spread mobile botnets and steal money or identity from mobile subscribers (GSMA, 2016) .

SMS spam is classified as 32.3% irritating, 24.8% squandering of time and 21.3% violating personal privacy (Taufiq Nuruzzaman, Lee, Abdullah, & Choi, 2012). For example, in countries such as India estimates of over 100 million SMS spam is received per day (Shahi & Yadav, 2013). SMS spam not just irritating but also incurring significant cost on both the Mobile Network Operators and the customers as well (Skudlark, 2015). SMS spammers can reach their victims by generating phone numbers unlike the email, where the number of possible email addresses is unlimited which makes users to fall victims of fraudulent activities such as phishing identity theft and fraud as shown in Figure (2.2) message have untrusted url which ask user to login and update his account information.



**Figure (2.2):** SMS spam asking to update Facebook account through fishing URL

## 2.4 SMS Spam Filtering Methods

There is much work on SMS spam filtering using techniques such as Black and White List, Text Classification (Taufiq et al., 2010), Boyer and Moore (BM) (Liu et al., 2010), Naïve Bayesian classifiers (Zhang & Wang, 2009) (Deng & Peng, 2006), neural networks (Anchal 2014), and frame model of ontology-based detection

(Balubaid & Manzoor, 2015) to name a few. More details in SMS filtering methods are found in Chapter 3.

## **2.5 Semantic Web**

The Semantic Web which is an extension, not a replacement of the current Web. The Semantic Web provides a common framework that allows data to be shared and reused across systems and applications.

In the semantic web, the applications understood by machine, with the help of meaning associated with each component stored on the web. A component representation scheme called ontology. Ontology allow computer understandings and interpretations of symbols, ontology allows semantic annotation of resources for information retrieval with inference (Sugumaran & Gulla, 2011). Next, we describe the ontology which is the essence of the semantic web.

## **2.6 Ontology**

Ontology is defined as “Is a formal, explicit specification of a shared conceptualization “. (Sugumaran & Gulla, 2011) .In other words, an ontology describes the concepts in the domain and the relationships that hold between these concepts. It is a shared vocabulary that can be used to domain (Taylor & Pohl, 2009).

There are many roles and tasks of ontology which are summarized as follows (Mizoguchi, Vanwelkenhuysen, & Ikeda, 1995) :

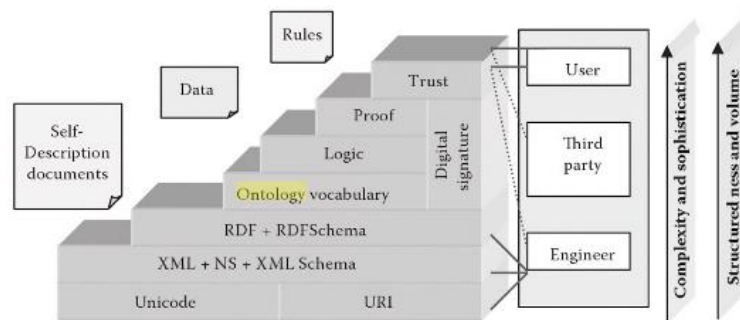
- Extract and organize the vocabulary for specific domain.
- Identify knowledge for problem solving.
- Provide domain experts with human-readable conceptual primitives in terms of which they can express their way of problem solving.
- Enable translation of the knowledge-level description of the problem-solving process in to tributes to clarify domain knowledge.

There are two approaches to design any domain ontology. First, top-down and second, bottom-up. In the top-down approach the experts determine the concepts and the relationships based on domain knowledge. In the bottom-up approach the experts

select the important concepts by analysis of data coverage and patterns related to them. Both top-down and bottom up approaches need participation of human. Also some automatic tools can reduce manual efforts. (J. Kim, Dou, Liu, & Kwak, 2007).

Based on the notion of semantic web and its important part, the ontology, there is an agreed upon architecture of semantic web see Figure (2.3) where a group of ontology-based layers define the structure of the semantic web. Next, we describe the most ontology related languages and representation.

- Resource Description Framework (RDF) is a framework for create statements in a form called triples. It allows to represent information about resources in the form of graph.
- RDF Schema (RDFS) have basic vocabulary of RDF. Using RDFS allow to create hierarchies of classes and properties.
- Web Ontology Language (OWL) is extends of RDFS by adding more advanced constructs to describe semantics of RDF statements. It allows more additional constraints, such as cardinality, restrictions of values, or characteristics of properties such as transitivity. It is based on description logic and it gives reasoning power to the semantic web.
- SPARQL is a RDF query language, it can be used to query any RDF-based data which querying language can retrieve information for semantic web applications.
- Semantic Web Rule Language (SWRL): is a rule format and important to allow describing relations that cannot be directly described using description logic used in OWL.



**Figure (2.3):** Layers of languages used for the semantic web (Sugumaran & Gulla, 2011)

## **2.7 Ontology Building**

Ontology building needs some experience in the domain which we want to build the ontology and it needs time and effort. (Noy & McGuinness, 2001) list the standard steps involved in developing ontology, included the following:

- 1- Determine the domain of ontology and scope.
- 2- Reuse existing ontology.
- 3- List important terms in the ontology.
- 4- Define classes and subclasses.
- 5- Define properties.
- 6- Define facets of the slots.
- 7- Create instances.

We explained briefly these steps in Chapter 4 and we employ them to build our SMS spam ontology.

### **2.7.1 Determine The domain of Ontology and Scope**

This step defines the purpose and boundaries of the ontology. There are several questions to be answered:

- What is the domain that the ontology will cover?
- For what we are going to use the ontology?
- For what types of questions the information in the ontology should provide answers?
- Who will use and maintain the ontology?

### **2.7.2 Reuse Existing Ontologies**

This step checks if there an ontology has been developed before in the same subject area. If such ontology exists, it is easier to use and modify the existing ontology more than to create a new ontology.

### **2.7.3 List Important Terms in The Ontology**

In this is step, we try to create a list of an expected concept terms that we can used on the ontology development.

### **2.7.4 Define The Classes and Subclasses**

There are several approaches for developing a class hierarchy (Uschold & Gruninger, 1996):

- A top-down approach starts with the definition of the most general concepts in the domain and subsequent of the concepts.
- A bottom-up approach starts with the definition of the most specific classes, the leaves of the hierarchy, with subsequent grouping of these classes into more general concepts.
- Combined approaches are a combination of the top-down and bottom up approaches.

### **2.7.5 Define The Properties**

The classes alone do not provide enough information to answer the questions from Step 1. We have already selected classes from the list of terms in Step 3. Most of the remaining terms most probably to be properties of these classes.

### **2.7.6 Define The Facets of The Slots**

Slots can have different facets such as allowed values, value type, the number of the values (cardinality), and other features.

### **2.7.7 Create Instances**

Finally, we need to create individual as instances of classes. Defining an individual instance of a class needs firstly choosing a class, then create an individual instance of this class, and finally set it in the slot values.



## 2.8 Resource Description Language (RDF)

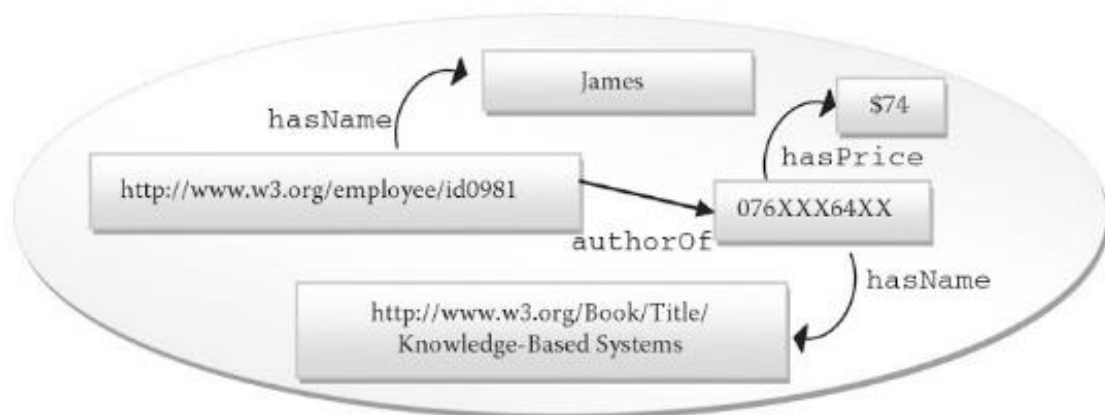
The RDF is a framework written in XML for describing resources on the web that facilitates automatic content understanding. It was developed to annotate web pages with machine-processable meta-data. It can be used to express Knowledge (Sugumaran & Gulla, 2011).

The design of RDF meet the following goals:

- Simple data model.
- Formal semantics and provable inference.
- Extensible URI-based vocabulary.
- XML-based syntax.
- Supporting XML schema data types.
- Allowing anyone to make statements about any resource.

RDF identifies things using web identifiers (URIs), and describes resources with properties and property values known as RDF triple which:

- A resource is anything that can have a URI, such as "<http://www.w3.org/rdf>".
- A property is a resource that has a name, such as "authorOf" or "hasname".
- A property value is the value of a property, such as "James" or "<http://www.w3.org/employee/id0981>" The following Figure (2.6) describe the resource "<http://www.w3.org/employee/id0981>".



**Figure (2.6):** RDF triples showing relationships between an employee, book, and price (Sajja & Akerkar, 2012).

RDF allows to link resources together but it cannot classify objects to classes for example that we cannot do that person is a subclass of human. For more powerful description language there is extends of RDF, RDF schema (RDFS) and web ontology language (OWL).

- RDF schema, it allows a number of constraints on the individuals and relationships in RDF triplets. It allows declaring objects and subjects as instances of certain classes, inclusion statements between classes and properties make it possible to express semantic relations between classes and between properties. It is also possible to semantically relate the “domain” and the “range” of a property to some classes.
- OWL allows to add more restrictions to knowledge representation. It categories properties (relationships) into object properties and data properties and allows to add restrictions on these properties.

## 2.9 SPARQL

SPARQL is the standardized query language for RDF, it able to manipulate and retrieve the data stored in RDF format, the same way of standardized query language (SQL) for relational databases. There are some similarities keywords such as SELECT, WHERE. It also has new keywords which have not seen in a SQL such as FILTER, OPTIONAL. (Sajja & Akerkar, 2012). The basic structure of a SPARQL query:

- PREFIX: the SPARQL equivalent of declaring an XML namespace.
- SELECT: like its twin in an SQL query, it is used to define the data that will be returned by the query.
- FROM: identifies the data against which the query will be run, can be given in runtime as well.
- WHERE: defines the part of RDF graph we are interested in.
- Variables: are prefixed with either "?" or "\$".

An example of SPARQL query is:

```
PREFIX plants: <http://www.linkeddatatools.com/plants>  
SELECT * WHERE
```

```
{  
?name plants:family ?family  
}
```

In this example SPARQL retrieve all triples with a matching all the plant URIs (subjects) and plant family names (literal-type objects) from the data.

## 2.10 DL Query

The DL Query provides a powerful feature for searching in ontology. The query language expression is based on the Manchester 25 OWL syntax, it is user friendly syntax for OWL DL that is fundamentally based on collecting all information about a particular class, property, or individual into a single construct, called a frame (protegewiki, 2016). An example of DL Query is:

hasGivenName value "Ayman"

in the above query suppose that we have several hundred instances of class Person in the ontology, to find an individual named "Ayman".

## 2.11 Protégé

Protégé is an ontology editor, and knowledgebase framework which is developed by Stanford University and Manchester University. Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development. It is a desirable tool for editing and browsing ontologies and for performing, some reasoning operations such as incoherence detection of the ontology. Protégé has recently embedded HermiT, Pellet and FaCT++ reasoners to makes reasoning more convenient (Protégé, 2016).

## 2.12 Reasoning

One of the important tools of an ontology is the reasoned. Reasoning is the process of inferring new information from an ontology (Sugumaran & Gulla, 2011). There are many available reasoners today that exploit the capabilities of Description Logics. A reasoner provides the basic core usability of ontology by testing for concept

satisfiability, class subsumption by concept hierarchy, class consistency, and instance checking (Wang, Zhang, Gu, & Pung, 2004).

Many reasoners use first-order predicate logic to perform reasoning. Inference commonly proceeds by forward chaining and backward chaining. The first order logic reasoning in description logics is based on concepts, rules, and individuals. Concepts relate to classes in ontology language, rules are equivalent to relationships, and individuals found in both cases. As described, reasoners allow the information contained within an ontology to be utilized to its fullest potential to maintain and infer information.

We use some reasoners in our research such as:

- **HermiT:** HermiT is a free (under LGPL license) Java reasoner for OWL 2/SROIQ with OWL 2 datatype support and support for description graphs. It implements a hypertableau-based decision procedure, uses the OWL API 3.0, and is compatible with the OWLReasoner interface of the OWL API.
- **JENA Reasoning Agent:** JENA is a Java framework for building Semantic Web systems. It support programming for RDF, RDFs and OWL, and support using queries such as SPARQL and includes a rule-based inference engine. (Totewar & Chatur, 2011).

## 2.13 Semantic Rules

The area of semantic rules is perhaps the most important for the Semantic Web's core technology and standards, using rules for, or with, more expressive OWL ontologies. All rules are expressed in terms of OWL concepts (classes, properties, individuals), which rules saved as part of ontology.

The Semantic Web Rule Language (SWRL) is a proposed language for the Semantic Web that can be used to express rules as well as logic.

An example of human readable rule syntax:

$\text{hasParent}(?s1,?s2) \wedge \text{hasBrother}(?s2,?s3) \Rightarrow \text{hasUncle}(?s1,?s3)$

This rule fire by rule to infer new fact which mean ?s1 has parent property with ?s2 and ?s2 has brother with ?s3 then implies ?s1 has uncle ?s3.

## **2.14 WordNet**

WordNet is a huge lexical database contains nouns, verbs, adjectives and adverbs which are grouped into sets of synonyms called synsets. Synsets have conceptual semantic and lexical relations. WordNet is also freely and publicly available for use and download. WordNet's structure is a useful tool for computational linguistics and natural language processing (Elkateb et al., 2006).

Arabic WordNet is being constructed following methods developed for EuroWordNet (Vossen, 1998). EuroWordNet approach maximizes compatibility across wordnets and focuses on manual encoding of the most complicated and important concepts (Elkateb et al., 2006). Language-specific concepts and relations are encoded as needed or desired. This results in a so-called core WordNet for Arabic with the most important sets of synonym (synsets), embedded in a solid semantic framework.

In our research, we use Arabic WordNet to extract synonym of SMS spam words in the ontology.

## **2.15 Evaluation Method**

In order to classify SMS spam or legitimate, there are different measures that we can use to evaluate the classification approach, we choose confusion matrix to get main performance measures for evaluation, such as Accuracy, Error rate, F-Measure, Precision, Recall and Matthews Correlation Coefficient (MCC) (Karami & Zhou, 2014).

A confusion matrix as shown in Table (2.1) is a matrix that use to describe the performance of a classification approach or "classifier" by test data. Which contains the number of correct and incorrect predictions values and broken down by each class. The confusion matrix overcomes the limitation of using classification accuracy alone.

The steps for prepare a confusion Matrix:

1. Dataset for testing with expected outcome values.
2. Create a prediction for each row in the dataset.
3. From the expected outcomes and predictions count:
  - The correct predictions number for each class.
  - The incorrect predictions number for each class.

These numbers set into a matrix as each row of the matrix match to an actual class, and each column of the matrix match to a predicted class. The counts of correct and incorrect classification are set into the matrix.

**Table (2.1):** Confusion Matrix

		Predicted	
		Spam	Legitimate
Actual	Spam	a (TP)	b (FN)
	Legitimate	c (FP)	d (TN)

**True Positive (TP):** positive instances that are correctly classified.

**False Negative (FN):** positive instances incorrectly classified as negative.

**False Positive (FP):** negative instances incorrectly classified as positive.

**True Negative (TN):** negative instances that are correctly classified.

The evaluation metrics were defined based on the confusion matrix, as shown in equations (1) to (6).

### **Accuracy**

Accuracy (ACC) is the number of correct predictions divided by the total number of the dataset. The best accuracy value is 1 whereas the worst is 0. It can also be calculated by  $1 - \text{ERR}$ .

$$\text{Accuracy}(R) = \frac{a + d}{a + b + c + d} \quad (1)$$

### **Error rate**

Error rate (ERR) is the number of all incorrect predictions divided by the total number of the dataset. The best error rate is 0, whereas the worst is 1.

$$\text{ERR} = \frac{b + c}{a + b + c + d} \quad (2)$$

### **Precision (Positive predictive value)**

Precision is the number of correct positive predictions divided by the total number of positive predictions. It is also called positive predictive value (PPV). The best precision is 1, whereas the worst is 0.

$$\text{Precision}(P) = \frac{a}{a + c} \quad (3)$$

### **Recall (Sensitivity or True positive rate)**

Recall or Sensitivity is the number of correct positive predictions divided by the total number of positives. It is also called true positive rate (TPR). The best sensitivity is 1, whereas the worst is 0.

$$\text{Recall}(R) = \frac{a}{a + b} \quad (4)$$

## **F-measure**

F-measure is a harmonic mean of precision and recall.

$$F - measure = 2 * \frac{P * R}{P + R} \quad (5)$$

## **Matthews Correlation Coefficient (MCC)**

MCC is used to determine the quality of classification methods, ranging between -1 and +1 with +1 indicating the best performance.

$$MCC = \frac{(TP * TN) - (FN * FP)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (6)$$

## **2.16 Summary**

In this chapter, we have presented a foundation for this research. We presented SMS spam definition and types. Additionally, we defined the semantic web and ontology and explained the steps that must be followed to build it. We also defined and explained the terminology of RDF, SPARQL and other tools used in the implementation and programming our approach such as JENA, Protégé. And explained the semantic rules and reasoning and WordNet and finally evaluation method.



# **Chapter 3**

## **Related Works**

## **Chapter 3**

### **Related Works**

We study and investigate different works related to SMS spam detection and email spam detection mainly using ontologies. They are introduced and analyzed with respect to this problem to show how far these works address the research problem. Parts of the related works can be a basis for solving our SMS spam detection problem. We point these out during the presentation and discussion in the next sections.

#### **3.1 SMS Spam Detection Methods**

S.-E. Kim, Jo and Choi (2015) proposed a light and fast algorithm for SMS filtering which can be performed within mobile phones independently. It employs techniques for remove unneeded data. These techniques include data filtering, feature selection, data clustering, etc. They select important features using relative volume of feature values. (S.-E. Kim, Jo, & Choi, 2015)

They use WEKA tool, which is a machine learning tool to evaluate the performance of feature selection methods such as Naïve Bayes, J-48 Decision Trees, and Logistic. They compared the performance of this method with standard feature selection methods. The new FR (Frequency Ratio) attribute selection technique has an advantage that it has a simple calculation formula compared to other techniques.

In this study, we notice that the result of reducing data has an advantage in reducing the execution time but it decreases in accuracy, but our research focuses on how to achieve more accuracy in SMS filtering process.

Delany, Buckley and Greene (2012) presented a state of the art SMS spam detection and filtering techniques and they reviewed some of different approaches to the SMS spam, also they discussed important issues with data collection and availability for further research. They analyzed a large dataset of SMS spam. They collected instances of SMS dataset by collecting messages from two public consumer complaints websites: GrumbleText and WhoCallsMe, which have assembled a corpus of 1,353 unique SMS spam messages. (Delany, Buckley, & Greene, 2012)

In this study the authors identified a number of challenges and directions which are visible now such as multilingual environments, shared data, hybrid solutions, advanced address-based filtering, scalability and real-world deployment, industry collaboration. The results of the work indicate that there is as yet no consensus on what the best techniques are for SMS spam filtering.

Gómez Hidalgo, Bringas, Sáenz and García (2006) analyzed Bayesian filtering techniques used to catch email spam to be applied in to SMS spam detection problem. They built two SMS spam test collections of significant size, dataset contains English and Spanish languages. The English database consists of 1,119 legitimate messages, and 82 spam messages, and Spanish database consists of 199 (14.67%) spam messages, and 1,157 (85.32%) legitimate messages. They have tested on them a number of messages representation techniques and machine learning algorithms, in terms of effectiveness. They have used the following algorithms: Naïve Bayes (NB), C4.5, PART, and Support Vector Machines. To evaluate classifiers, they used the Receiver Operating Characteristics (ROC) method to make performance comparisons among classifiers. They have performed a series of experiments with different attribute definitions, using several learning algorithms to check if Bayesian filtering technique can be transferred to SMS spam filtering. The results show that Bayesian filtering technique can be effectively transferred from email spam to SMS spam. (Gómez Hidalgo, Bringas, Sáenz, & García, 2006)

In this study, we notice that we can extend the use of email filtering methods to SMS filtering too, and all supervised machine learning give positive results in SMS filtering specially Bayesian techniques. Naïve Bayes classifiers work by correlating the use of tokens typically words with spam and non-spam e-mails and then using Bayes' theorem to calculate a probability that an email is spam or not, in our research we can classify spam words using weights of spam words as instances ontology.

Khemapatapan (2010) proposed two filtering methods for Thai-English SMS spam message and then Applying Support Vector Machine (SVM) and Naïve Bayesian (NB) algorithms for filtering. The two filtering methods perform Thai word segmentation to classify the words in SMS message.

The first method modifies English-based spam message filtering perform Thai word segmentation. They apply text normalization to remove some symbols such as a set of number and special characters. They perform Thai word segmentation process to classify words in SMS message before the filtering process such as SVM and NB algorithms. Finally, they got new knowledge from filtering which it can used to update the database to modify weight of each spam word.

The database used to store Thai and English words from dictionary, which each row in the database contains 2 fields the word and its spamming weight.

The second method, in the Thai language may be can use some of vowel characters. There are words have similar spelling but different means by using different vowel characters. Sometimes users type Thai words using the wrong vowel. Thus, semantic analysis and Thai word pre-preprocessing phase will be applied in the filtering similar to the filtering method number 1. First of things they apply text normalization after the SMS message is obtained. Then removing a duplicated vowel is additionally performed in this process in order to reduce the number of irregular words because users can type duplicated or missing vowels in some words. In the next process, each word is separated from the SMS message after segmentation process. Then, semantic analysis and correcting processes are applied. This process first finds the words having no meaning by comparing against the database. Then, the process tries to correct them by removing and/or swapping vowel characters using simple semantic analysis. Thus, after this process all wrong or irregular words should be modified to be corrected words. (Khemapatapan, 2010)

They use two data sets are for this purpose training and testing, for each set, there is a total of 400 SMS messages in which there are about 120 SMS spam messages.

The proposed methods take more processing time than the previous SMS spam filtering. In the results, the filtering method number 2 using SVM-based filtering provide highest accuracy for Thai-English SMS messages but taking longer processing time. However, in practical cases, the filtering method number 2 using NB-based filtering is better due to its processing time is low and it gives an acceptable accuracy.

In this study, we note that should remove all stop words characters from text and make text normalization before applying classifiers.

Akbari and Sajedi (2015) proposed an algorithm called (GentleBoost algorithm) for SMS spam detection. They tried to reduce the number of word attributes significantly without reducing accuracy in comparison to other successful methods. They used content of messages and tried to extract the words which are more repeated in spam messages. They applied tokenization by removing stop words or symbols such as “the”. For classification, they applied GentleBoost algorithm and finally by optimizing the features and applying GentleBoost algorithm which combines features of AdaBoostM1 and LogitBoost algorithms. They obtained only 32 word attributes and 98.30% accuracy. (Akbari & Sajedi, 2015)

Boosting works by sequentially applying a classification algorithm to reweighted versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus produced. The algorithm performs very well for binary classification and unbalanced data. One of the most advantages of this method is in feature extraction. For feature extraction, they tried to reduce the number of word attributes as far as possible by removing unused word attributes and optimizing the features.

In this study, the authors use a new algorithm for detecting SMS spam and they get high accuracy, and algorithm focus in the weight of words in the dataset. To reduce the number of words, we will use only spam words in ontology.

Liu Jun, Ke Haifeng and Zhang Gaoyan (2010) proposed pattern-matching algorithm called BM algorithm. They evaluated the system and filtering algorithms by using the actual SMS data. The experimental data was 100,000 short messages randomly extracted from the actual system of operators, and tested the BM algorithm and proved that BM algorithm is suitable to run under the condition of high concurrency and real-time environment. (Liu et al., 2010)

Using pattern-matching algorithm (BM) before filtering system, finds keywords in text messages then sends them to the filtering system to reduce the number of test matches and enhances the overall efficiency of the system.

In this study, the authors use pattern-matching algorithm which will search in text for specific pattern, but in this case no semantic use and it will search for spam words without knowing about relations to other words.

Uysal, Gunal, Ergin and Sora Gunal (2012) investigated the impact of several feature extraction and feature selection approaches on filtering of SMS spam. This study extensively analyses the effects of several feature extraction and feature selection methods together on filtering SMS spam messages in two different languages, Turkish and English. The selected features are then combined with the structural features and fed into two distinct pattern classification algorithms, namely K-nearest neighbour and SVM to classify SMS messages as either spam or legitimate. The filtering framework is evaluated on two separate SMS message datasets consisting of Turkish and English messages. Experimental work indicated that the combinations of bag-of-words (BoW) and structural features, rather than BoW features alone, offer better classification performance most of the time. Efficacy of the utilized feature selection strategies was not significantly superior to each other for both languages.

Next, we present the use of ontology in email spam detection where is able to give better results than traditional approaches.. (Uysal, Gunal, Ergin, & Sora Gunal, 2012)

### **3.2 Using Ontology in Email Spam Detection**

Balakumar and Vaidehi (2008) proposed a method to create an email classification filter, It uses ontology for understanding the content of the email and using Bayesian approach for making the classification. (Balakumar & Vaidehi, 2008)

The term “categorization” is used to refer to the classification of email based on email content and classification indicates classifying mail into legitimate and spam. This is achieved by two different phases training phase and online Integration Phase.

In the training phase, they used database context to build ontology as tree structure with classes and other attributes as nodes and branches representing the relationship between the nodes. Ontology is planned to have whitelist, category and keywords as supper class, class and instances.

In online integration phase the ontology is created in training phase and is integrated to an email server or client. This phase tries to categorize the incoming email based on the trained keywords in ontology database. This phase includes checking the senders address, tokenizing the incoming mail, and verifying whether the email is spam, compare the tokens with that in ontology database. For each keyword obtaining the probability of dependency with respect to each category, then compute the overall probability for each category using Bayesian formula, save the email into a folder with the category name having an overall probability among all the other categories.

With a simple dataset, it is found that 98% of the email has been successfully classified as spam and legitimate, about 9500 has been categorized successfully. The ontology can be effectively used to learn an email and to classify the incoming emails into folders according to the content of the email.

The authors use ontology for email spam and they put content of emails in the ontology and they use machine learning classifier such as Bayesian, but in our research, we try to use semantic rules in ontology to classify contents using an external algorithms classifier.

Youn (2014) proposed two levels of ontology spam filters: a first level global ontology filter for each user to increase spam filtering accuracy and a second level user-customized ontology filter which is user-customized, scalable, and modularized. It can be embedded to many other systems for better performance.

A global ontology was created with a 2108 email dataset (42.82% are spam and 57.18% are legitimate email). The tfidf mechanism was used as a feature selection algorithm. In the Weka, the C4.5 decision tree algorithm was used for email classification.

Through Weka, apply the classifier and obtain the results, then the classified results are converted to RDF file. The RDF file is send into JENA which provides a programmatic environment for RDF, RDFS, OWL, and SPARQL and includes a rule-based inference engine.

The challenge they faced was mainly to make C4.5 classification outputs to RDF and to give it to JENA, i.e., interfacing two independent systems and creating a prototype that actually uses this information that flows from one system to another to get certain desired input. In this case, it was the classification of email.

The use of the global ontology filter showed about 91% of spam filtered, which is comparable with other methods. (Youn, 2014)

The authors depend on Weka tool to classify email then convert results to RDF file to send it to JENA, and then they can be using JENA feature to query and print results, but in our research the knowledge base SMS spam ready for use by JENA rule and classification depending on semantic rules.

Kiamarzpour, Dianat and Sadeghzadeh (2013) introduced a new method to classify the spam by combining the output of several decision trees and the concept of ontology. They have used the SpamBase Dataset, Weka and JENA to build the ontology. The database SpamBase contains 4601 emails of which 39.4% are spam and 60.6% are valid emails.

The first step is to make a smart decision tree, and then obtain the ontology based on the classification of trees j48. The second step is to map the decision tree to the ontology and then get a query from the obtained ontology and give it a test Email and determine whether it is a spam or not.

They have used 4101 emails in training session and have built the trees by the help of software Weka and converted them to the ontology format. They got the query from this ontology for the test stage and give them 500 test emails for classifying them into two groups of spam and valid emails.

They have compared the obtained results with the results of two methods SVM and Naïve Bayes which are the most common Email classification methods; thus They have found that the results obtained from voting the decision trees between two errors of considering the spam instead of valid email (FN) and a valid email instead of spam (FP) establish a reasonable balance.. (Kiamarzpour, Dianat, & Sadeghzadeh, 2013)



### 3.3 Using Ontology in SMS Spam Detection

Balubaid and Manzoor (2015) proposed ontology based SMS controller. The proposed system is Android-based application which analyze the text message and classify it using ontology as legitimate or spam. The proposed system use algorithm for SMS incoming to verify the spam messages, comprised of three steps Pre-processing, content analysis and spam classification. It study focused in content analysis by loading synonym and hypernym from ontology. Each concept is compared with the spam concepts one by one and matches are stored in a separate resultant set with labels O for Original, S for Synonym or H for Hypernym. The collective spam score of the resultant set is calculated by adding all individual concept scores.

This study is in line with our research to use ontology to detect SMS spam, but this study focused in building the solution in client side in mobile phone side which is not available for all people and not all people accept to install new application in their phones. (Balubaid & Manzoor, 2015)

Cao, Nie and Liu (2011) proposed a systematic frame model of ontology-based mobile phone spam messages detection system, which automatically detects and filters spam in real time, the frame model contains:

- Spam initial-detection module: where the information from the users in the white list may be sent directly and it is filtered directly if the sender is in the blacklist, etc.
- Mobile phone spam ontology model which stores SMS.
- Ontology mapping module which formalize the information from the external, and generate mapping rules in accordance with the structure of ontology model of mobile phone spam.
- Mobile phone spam detection module executes ontology reasoning and semantic similarity calculation for the newly acquired information and the samples in spam ontology samples database and determines the spam probability according to the calculation results.

- User credit calculation after identifying the spam the user credit will be calculated and it will be sent to administrator interface system to manage the blacklist automatically.
- Administrator interface is used to intercept messages or close the corresponding functions of the mobile phone number.

In this study we should divide our approach to different modules and the administrator of the system should accept filtered SMS or blocked it after system classified it. (Cao, Nie, & Liu, 2011)

### **3.4 Summary**

We presented a review of SMS spam detection works which we divided into three categories. In the first category, we studied different approaches of SMS spam detection such as SMS classifications using Naïve Bayesian, neural networks, string matching algorithm. In the second category, we focused on ontology-based approaches in spam detection for emails. In the third ontology-based approaches in spam detection for SMS. We conclude that SMS have common factors with email. The results of the reviewed works demonstrate that spam filtering and classification techniques can be effectively transferred from email to SMS. The use of modern techniques such as those from Semantic Web and ontology can effectively help to detect SMS spams with acceptable accuracy. In other side, the authors not used reasoner to classify SMS as spam or legitimate, which they used another algorithms as classifier, but in our research we try to use reasoner as classifier, and they not used synonyms of spam words and relation between this words.

**Chapter 4**  
**Arabic SMS Spam**  
**Ontology**

## **Chapter 4**

### **Arabic SMS Spam Ontology**

In this chapter, we present the steps to develop the Arabic SMS spam domain ontology to be used as a basis for detecting and classifying Arabic SMS messages. Additionally, we present the evaluation of Arabic SMS spam ontology.

The ontology content is related to spam words domain and is collected from a number of SMS spam corpus. The SMS spam ontology developed with the assistance of SMS provider in Palestine. The advantage of the ontology model is that it is easy to be extensible, the possibility to manage additional information that might be related to the detection results.

We use Protégé to build the ontology which is one of the most widely used ontology development that defines ontology concepts (classes), properties, taxonomies, various restrictions, class instances and rules. It also supports several ontology representation languages, including OWL (Jain & Prasad, 2014). Building the ontology consists of the following steps as present in section 2.7:

- Step 1: Determining the domain of ontology and scope.
- Step 2: Reuse existing ontologies.
- Step 3: Overview of the ontology.
- Step 4: List the important terms in SMS spam.
- Step 5: Define classes and subclasses of SMS spam.
- Step 6: Define the properties of classes.
- Step 7: Define the facets of the slots.
- Step 8: Create instances of spam words.

#### **4.1 Determining The domain of Ontology and Scope**

Developing ontology without any purpose is not a goal in itself. Ontology is a model reflecting a particular domain built for a particular use. It is an abstraction of a domain determined by its future usage and by future extensions that are already

anticipated. Defining the SMS spam ontology domain and scope requires answering some basic questions:

**1. What is the domain that the ontology will cover?**

The ontology covers and captures the structure of SMS message and spam classification of message, which is a specific and limited domain serving the purpose of using the ontology in detecting spam messages.

**2. What is the use of the ontology?**

The ontology is to provide a knowledge base of SMS spam. It will be used in a system to detect and classify spam messages in Semantic Empowered Web applications.

**3. What types of questions would be answered by the information contained in the ontology?**

The ontology would provide answers to questions related to SMS spam such as:

- What is the structure of SMS messages?  
The structure it means the parts of SMS which contains the sender name of SMS and mobile and text.
- Is the SMS messages spam or legitimate?  
The reasoner check and give result if SMS is spam or legitimate.
- What is the common spam sender names?  
The ontology retrieve common blocked sender name by SPARQL query.
- What is the type of spam?  
Such as (scam , commercial, political, phishing).
- What is the spam words and synonym and relations with other words?  
For example, "مسيره" (march) has synonym "مظاهرة".
- What is the spam words weights?  
"مسيره" (march) has weight 0.5.
- Why this SMS is spam?  
the reasoner can give reason why this SMS is spam depending on spam classification.

#### 4. Who will use the ontology?

The proposed ontology will be available to the web system, for any developer to reuse it, or for any BulkSMS providers that are interested to detect and classify SMS spam to protect customers of operator from SMS spams. It must be clear and scalable to add any possible developments for this domain.

#### 4.2 Reuse Existing Ontologies

This step is to ascertain if there exists ontology that is developed previously in the SMS spam area. If such ontology exists, it is easier to modify the existing ontology to suit ones needs than to create a new one. Existing ontology, such Youn (2014), do not separately cover the content of SMS and sender name structure. Therefore, after reviewing such ontologies, we decided to design our own ontology that both covers the SMS structures and message contents. the purpose of detecting and classifying SMS spam message within the proposed approach.

#### 4.3 Overview of The Ontology

We develop a specific ontology for Arabic SMS spam that consists of the structure of SMS classes, the main classes in the ontology shown in Figure (4.1).

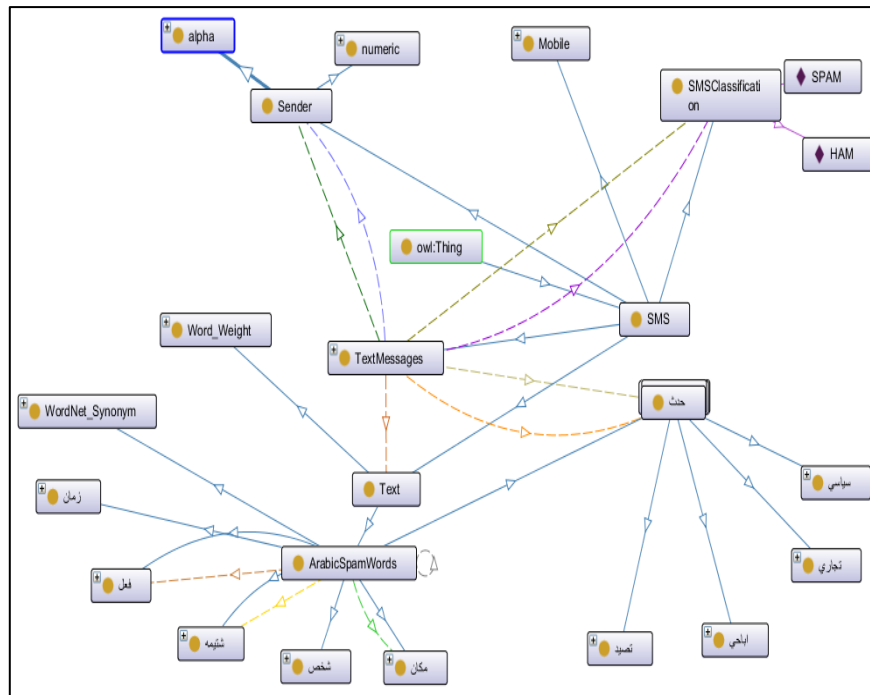


Figure (4.1): Main classes in the SMS spam ontology

We identify some Arabic spam words and match their synonyms from WordNet which are needed in the process of SMS spam detection in our approach. Table (4.1) show ontology metrics which includes a number of classes, object property, and data properties in SMS spam ontology.

**Table (4.1):** Ontology metrics

Domain and Scope of the Ontology	SMS spam
Axioms	501
Logical axioms	321
Declaration axioms	176
Classes	21
Object properties	12
Data properties	7
Individuals	135

#### 4.4 List The Important Terms in SMS Spam

This step can be viewed as a brainstorming activity in which we list the words that we want to use, to demonstrate the ontology terms, and the properties that may have. We also benefited from the collected SMS spam to get the knowledge about spam terms.

The following questions guide our brain storming activity to determine the terms:

1. What are the main terms that we want to talk about?

The main terms we need to talk about SMS message parts such as sender name, text, and spam words in text message.

2. What are the properties of these terms? What is needed to be said about those terms?

- “TextMessages” term has the following properties “اسم مرسل” (has\_sender), “تحتوي على” (contains\_to), SMSClassification, “سبب الحظر” (block\_reason), “تصنيف” (classification), “تحتوي على رابط” (contains\_url), “له رقم” (has\_number)
- “ArabicSpamwords” term has the following properties “يوجد فعل” (has\_verb), “مرتبط بشتيمة” (has\_revile), “له معنى” (has\_synonym), “ترتبط بمكان” (has\_place), “لها وزن” (has\_weight).
- “Sender” term has the following properties “عدد الرسائل المحظورة” (no\_blocked\_sms), “محظور” (blocked\_sender).

We can use these property terms to make it object properties and data properties in SMS spam ontology.

#### 4.5 Define Classes and Subclasses of SMS Spam

This step defines classes (concepts) used in the ontology domain. We define classes and sub-classes related to the ontology domain. Table (4.2) contains the ontology classes where ArabicSpamWords is the most general concept, the super classes are shown in bold which is the top of hierarchy of the structure of classes.

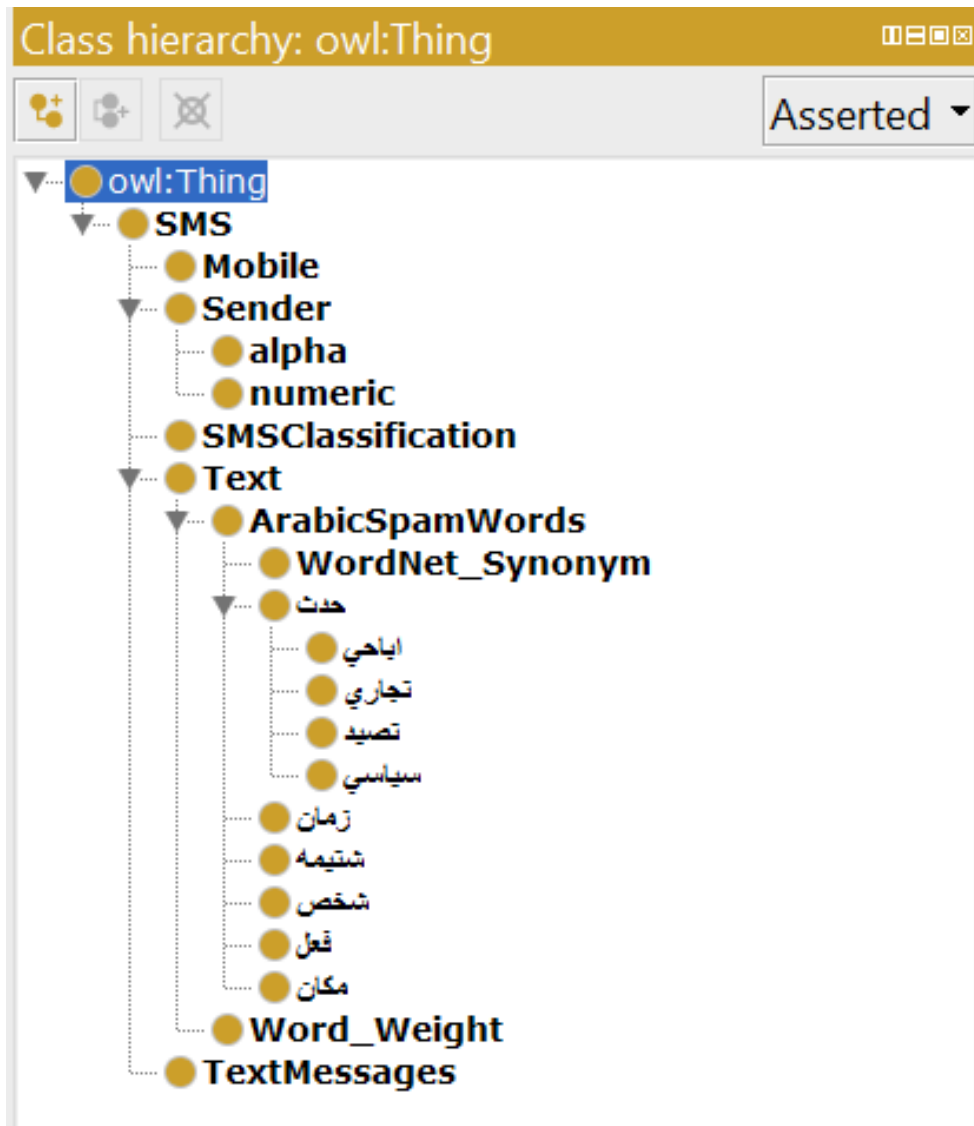
**Table (4.2):** The Arabic SMS ontology classes and subclasses

No	Class /Arabic	Class /English	Description
1	رسالة	SMS	Represents the main class of SMS spam
2	اسم المرسل	<b>Sender</b>	Represents the type of sender of SMS alphabetical or numeric
3	أبجدي	alpha	Represents type of sender using alphabetical characters
4	رقمي	numeric	Represents type of sender using numeric characters.
5	نص	<b>Text</b>	Represents all classes contain spam words
6	كلمات العربية المزعجة	<b>ArabicSpamWords</b>	Represents all classes contain Arabic spam words



No	Class /Arabic	Class /English	Description
7	حدث	Event	Represents the main Arabic spam words classes which include adult, commercial, political, and phishing.
8	اباحي	Adult	Represents the adult Arabic spam words
9	تجاري	Commercial	Represents the commercial Arabic spam words
10	سياسي	political	Represents the political Arabic spam words
11	تصيد	Phishing	Represents the phishing Arabic spam words
14	شخص	Person	Represents the important persons.
15	فعل	Verb	Represents the important spam verbs.
16	مكان	Place	Represents the important spam places
18	نص الرسالة	TextMessages	Represents the words of message sent.

After determining and defining classes, we create the class hierarchy in protégé OWL as shown in Figure (4.2). The class hierarchy contains SMS attributes such as mobile numbers and sender name. The sender name, for instance, is divided into two subclasses; alphabetical sender name and numeric sender name. The class Text (message text) contains a number of spam sub classes such as “حدث” (event) which in turn contains subclasses “اباحي” (pornographic), “سياسي” (political), “تجاري” (commercial), “تصيد” (phishing). Other important classes may have relations with spam words such as “زمان” (time), “مكان” (place), “شخص” (person), WordNet\_synonym. Finally, there is the class TextMessages which is used to store SMS message id.



**Figure (4.2):** The class hierarchy of the Arabic SMS spam ontology

#### 4.6 Define The Properties of Classes

After define classes, they do not provide enough information about our questions in step 1 so we define object properties (relations) among classes as a requirement to come up with the ontology. Creating object property plays important role in connecting classes (concepts) of the ontology in our Arabic SMS spam ontology domain. We used 11 object properties that connect the important concepts which have relations with each other that are illustrated in Table (4.3).

**Table (4.3):** Object properties of the ontology classes

No	Object Properties	Object Properties	Domain	Range
	Arabic	English		
1	اسم_مرسل	has_sendername	TextMessages	Sender
2	تحتوي_على	contain_of	TextMessages	Text
3	ترتبط_بمكان	has_place	حدث	مكان
4	تصنيف	classification	TextMessages	SMSClassification
5	له_معنى	has_sym	ArabicSpamWords	WordNet_Synonym
6	يوجد_فعل	has_verb	حدث	فعل
7	يوجد_تناقض	has_discrepancy	ArabicSpamWords	ArabicSpamWords
8	رقم_المحمول	has_mobile	TextMessages	Mobile
9	لها_وزن	has_weight	ArabicSpamWords	Word_Weight
10	مرتبط_بشتيمه	has_insult	ArabicSpamWords	شتيمه
11	سبب_الحظر	blocked_reason	TextMessages	SMSClassification

One of the important object properties is “تصنيف” (classification) which is used to classify new message as spam or legitimate.

Then, we create data properties including their domains and ranges, we used 6 data properties that connect the important concepts which have relations with each other that are illustrated in Table (4.4).

**Table (4.4):** Data properties of the ontology classes

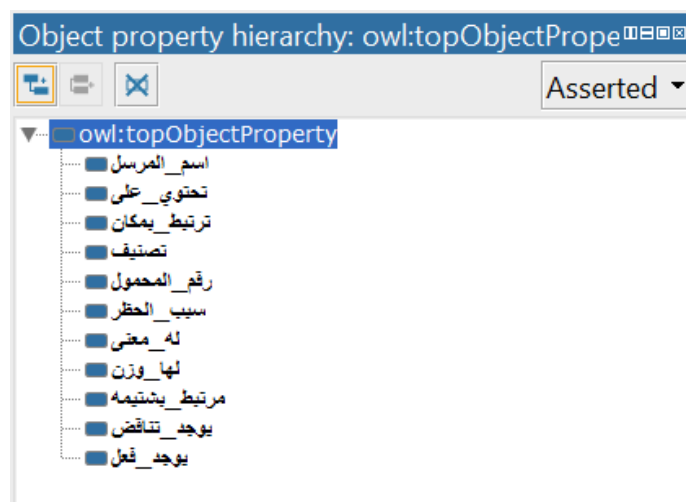
No	Data Properties/Arabic	Data Properties/English	Domain	Range
	1	تحتوي_على_رابط		
2	عدد_الرسائل_المحظوره	Contain_of	sender	Int
3	له_رابط	Has_place	تصيد	String
4	له_رقم	classification	TextMessages	Boolean
5	محظور	Has_sym	Sender	Boolean
6	وزن_الرسالة	Message_weight	TextMessages	Double

An example of a data property is “عدد\_الرسائل\_المحظوره” (no\_sms\_blocked) which relates instances of the domain SurpriseSMS to data value as shown in Figure (4.3).

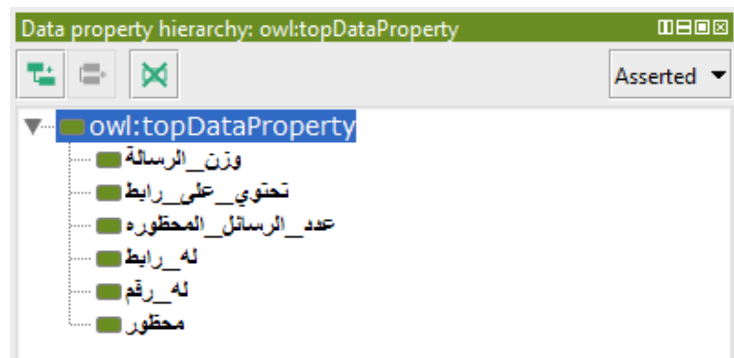


**Figure (4.3):** An Example of data property “عدد\_الرسائل\_المحظوره”

After determining object properties and data properties, we create them in protégé OWL. Figure (4.4) shows object properties in protégé and Figure (4.5) shows data properties in protégé.



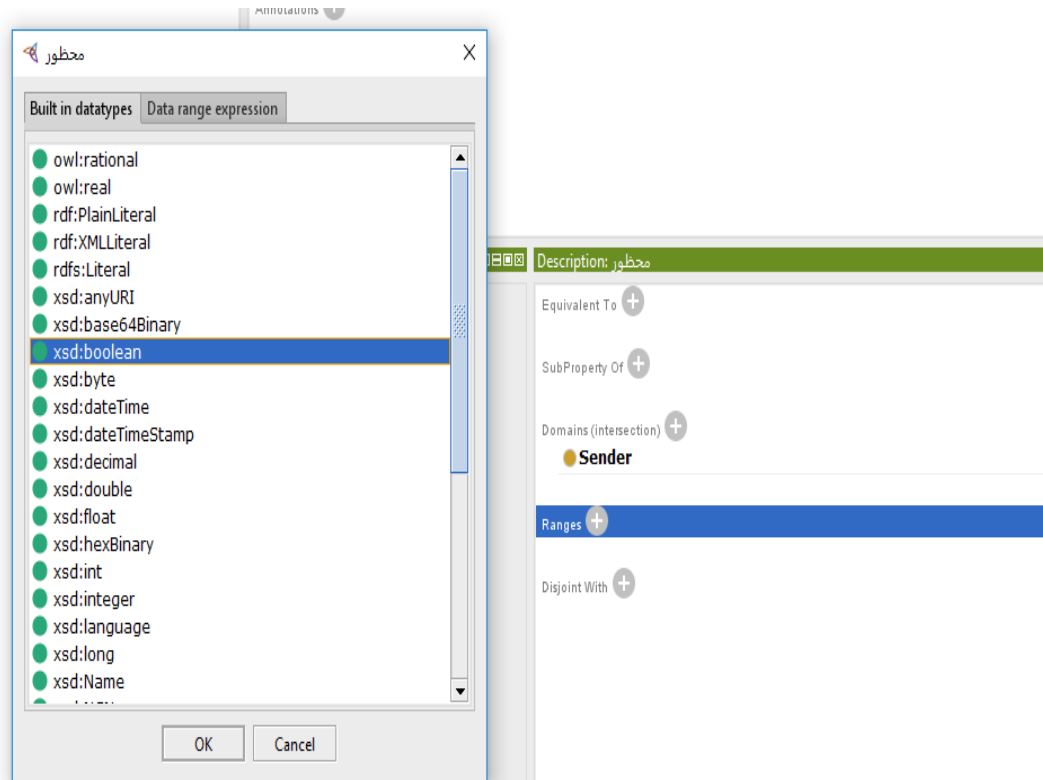
**Figure (4.4):** Object properties shown in Protégé



**Figure (4.5):** Data properties shown in Protégé

## 4.7 Define The Facets of The Slots

Slots (sometimes called roles or properties) have different facets (sometimes called role restrictions) that describe value types, allowed values, the number of the values (cardinality), and other features of the values the slots can take. For example, data property of “محظور” (blocked) is Boolean, and the “له رابط” (has\_url) has string value, “عدد الرسائل المحظورة” (no\_sms\_blocked) has int value. In Figure (4.6) show how to add data restriction by Protégé.

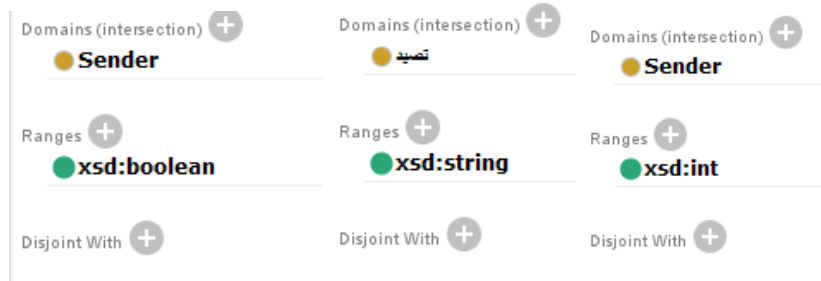


**Figure (4.6):** Creating data restriction

**Value type:** This describes the different types of values a property can take. For example:

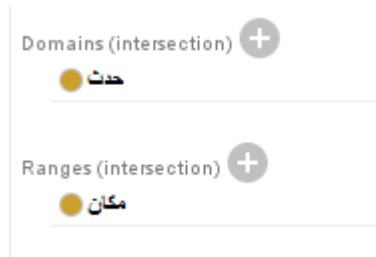
- 1- String:** The property “له رابط” (has\_url) has the value type string, which mean the domain of properties have range string value of blocked URL as shown in Figure (4.7).
- 2- Number:** The property “عدد الرسائل المحظورة” (no\_sms\_blocked) has the value type integer, which mean the domain of properties have range value number of blocked messages as integer value. as shown in Figure (4.7).

3- **Boolean:** The property “محظور” (blocked) has the value type Boolean, we used this for true–false flags that mean if sender name is blocked then value should be false. as shown in Figure (4.7).



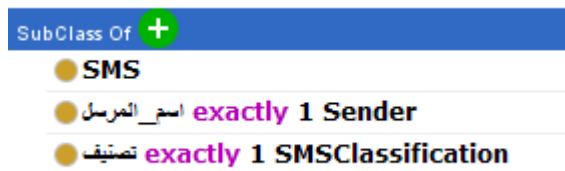
**Figure (4.7):** Example of different data types

**Allowed values:** This represents values allowed for different properties. The property “ترتبط\_بمكان” (has\_place) has allowed values are “حدث” (event) and “مكان” (place) as show in Figure (4.8).



**Figure (4.8):** Example of allowed values of slots

**Cardinality:** A property can have single value or multiple values. Cardinality defines how many values a property can have. For example, the property “اسم\_المرسل” (has\_sender) has exactly 1 Sender as show in Figure (4.9).



**Figure (4.9):** Example of cardinality

## 4.8 Create Instances of Spam Words

The last step is creating instances so we created the individual instances of all classes in the hierarchy of the ontology. Creating instances (individuals) is a very important step to enrich the ontology with direct relation with classes and sub-classes. For example, the class “تجاري” (commercial) have several instances which include “اربح” (win), “اشترك” (subscribe), “جائزه” (award) etc. Figure (4.10) depicts some of these instances.

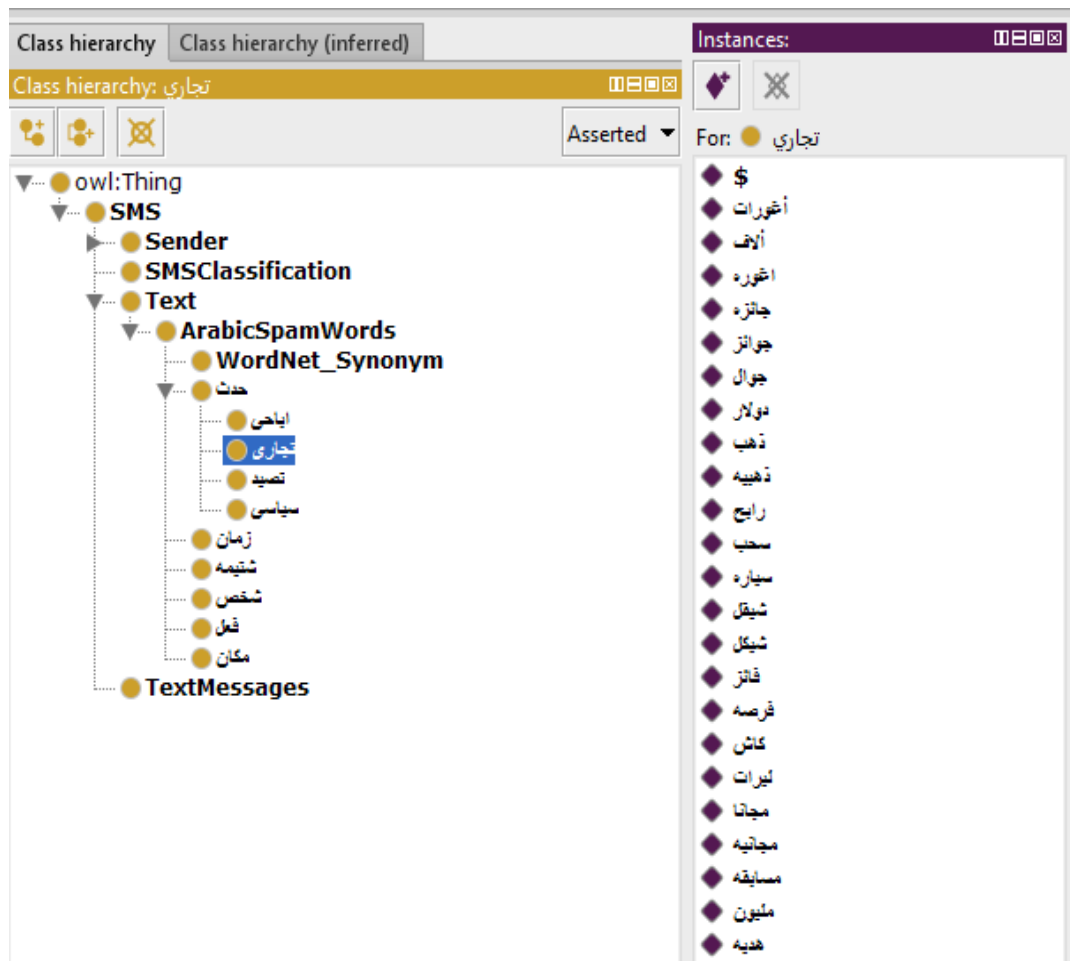


Figure (4.10): List of some ontology instances

## 4.9 Evaluate Ontology

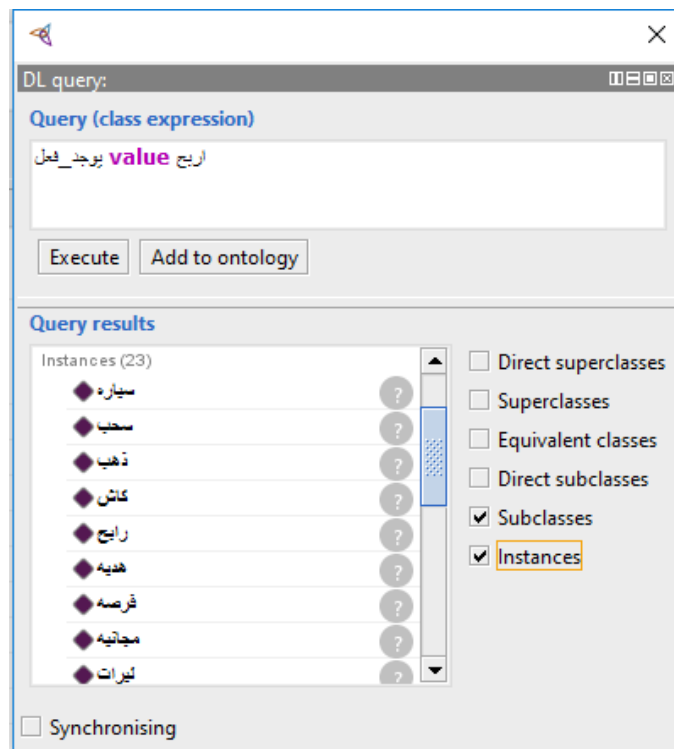
Before evaluating the ontology, we run HermiT reasoner to check whether or not the ontology is consistent, HermiT can handle OWL DL safe rules and the rules can directly be added to the ontology. Reasoning performed by testing the consistency of

a number of knowledge bases derived from the original ontology. We get new or hidden knowledge utilized in the ontology.

In order to evaluate the ontology, we use the Description Logic Query (DL-Query) that is a standard Protégé plugin and we can show explanation of reasoning for the result by DL-Query, and the SPARQL RDF Query Language (SPARQL).

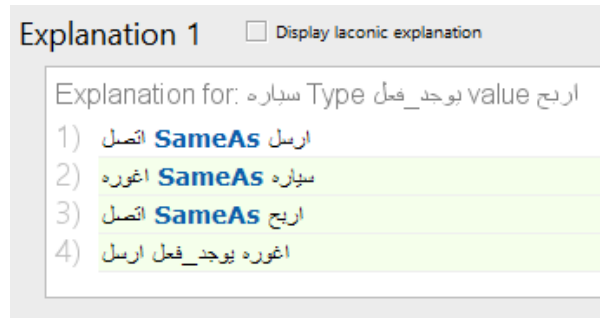
**Example 1:**

- The question: what are the commercial spam words related with verb “اربح” (win)?
- Reasoner: HermiT 1.3.8.413.
- Query type: DL-Query.
- The query: **value** اربح يوجد\_فعل
- The result of the query is shown in Figure (4.11) which returned the spam commercial words related with the verb “اربح”(win).



**Figure (4.11):** Query for all spam words related to the verb “اربح” (win)  
 The result show that all individuals have relation with verb “يوجد\_فعل” with “اربح” (win), Figure (4.12) show justification for the result “سيارة” (car). We use owl:sameAs and the reasoner inferred new facts.

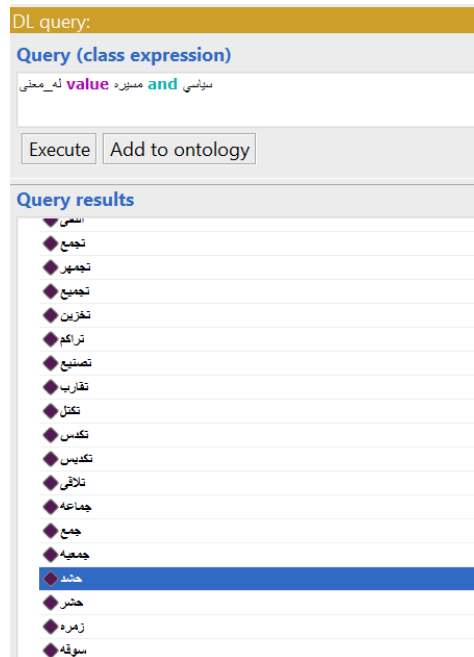




**Figure (4.12):** Justification result of query spam words related to the verb “اربح” (win)

**Example 2:**

- The question: what are the political synonym spam words of “مسيره” (march)?
- Reasoner: HerMiT 1.3.8.413.
- Query type: DL-Query.
- The query: سياسي and مسيره value له\_معنى
- The result of the query is shown in Figure (4.13) which returned political synonym spam words related to “مسيره” (march).



**Figure (4.13):** Query for all political synonym spam words of “مسيره” (march)

The result show that all the political individuals have relation with “له\_معنى” (has\_synonym) with “مسيره” (march), Figure (4.14) show justification for the result “حشد” (crowd), we use owl:Symmetric, owl:Transitive for object property “له\_معنى” (has\_synonym) and the reasoner match rule

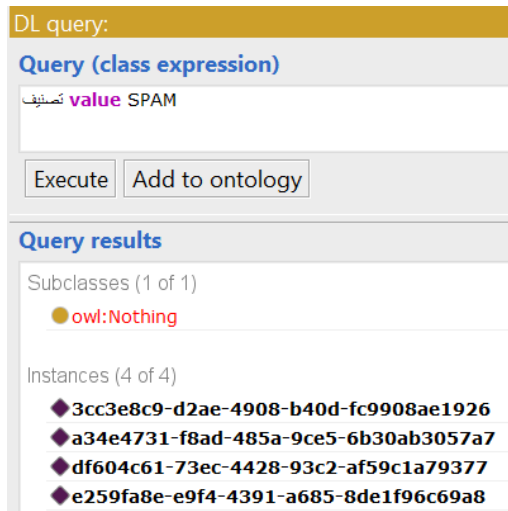
“(سياسي(?word) ^ له\_معنى (?word, ?m) -> سياسي(?m))” and inferred new facts.



**Figure (4.14):** Justification result of query political individual have relation with “له\_معنى” (has\_synonym) with “مسيره” (march)

**Example 3:**

- The question: what is the spam messages classified as spam?
- Reasoner: Hermit 1.3.8.413.
- Query type: DL-Query.
- The query: **value** تصنيف SPAM
- The result of the query is shown in Figure (4.15) which returned all messages classified as spam.



**Figure (4.15):** Query for all SMS classified as spam

The result show that all messages individuals which classified SMS spam, Figure (4.16) show justification for the result of message id “cc3e8c9-d2ae-4908-b40d-fc9908ae1926”, we use owl:Symmetric, owl:Transitive for object property has\_synonym “له\_معنى” and the reasoner match rule:

“TextMessages(?msg) ^ تحتوي\_على (?msg, ?word) ^ تحتوي\_على (?msg, ?x) ^  
 (-> يوجد\_فعل (?m, ?x) ^ سياسي (?m) ^ له\_معنى (?word, ?m) ^ تصنيف (?msg, SPAM) ^ سبب\_الحظر (?msg, SPAM) and inferred new facts.

Explanation 1  Display laconic explanation

Explanation for: 3cc3e8c9-d2ae-4908-b40d-fc9908ae1926 Type تصنيف value SPAM

1)	<b>Symmetric:</b> له_معنى	In ALL other justifications	?
2)	تحتوي_على مسيره 3cc3e8c9-d2ae-4908-b40d-fc9908ae1926	In ALL other justifications	?
3)	3cc3e8c9-d2ae-4908-b40d-fc9908ae1926 Type TextMessages	In NO other justifications	?
4)	تجمع له_معنى مسيره	In ALL other justifications	?
5)	تجمع يوجد_فعل نستفركم	In ALL other justifications	?
6)	سياسي Type تجمع	In ALL other justifications	?
7)	3cc3e8c9-d2ae-4908-b40d-fc9908ae1926 تحتوي_على نستفركم	In ALL other justifications	?
8)	TextMessages(?msg), تحتوي_على (?msg, ?word), تحتوي_على (?msg, ?x), له_معنى (?word, ?m), يوجد_فعل (?m, ?x) -> تصنيف (?msg, SPAM), سبب_الحظر (?msg, سياسي)	In ALL other justifications	?

Figure (4.16): Justification result of query classified as spam

#### Example 4:

- The question: What is the synonym of the “تجمع”?
- Query type: SPARQL.
- The result of the query is shown in Figure (4.17).

Snap SPARQL Query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX a: <http://www.semanticweb.org/ayman/ontologies/2016/7/sms#>
SELECT ?synonym
WHERE {
?s a: له_معنى ?synonym
FILTER(?s = ا:تجمع)
}

```

a:همج
a:اجتماع
a:جماعه
a:شمل
a:تلاقى
a:جمع
a:مجلس
a:تجمعهر
a:احتشد
a:تصنيع
a:تكتل
a:تكدير
a:البقاء_عند_نقطه
a:عوغاء

Figure (4.17): Query for all synonym spam words of “تجمع”

#### Example 5:

- The question: What is the sender name which is blocked?
- Query type: SPARQL.

- The result of the query is shown in Figure (4.18).

```

Snap SPARQL Query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX a: <http://www.semanticweb.org/ayman/ontologies/2016/7/sms#>
SELECT ?sender
WHERE {
?sender a:محظور true
}

```

a:Melody
a:SurpriseSMS
a:Guessing

**Figure (4.18):** Query for all blocked sender name

The SMS spam ontology is built to reflect the SMS spam domain. Since the domain is related to the Arabic language in terms of the content/text of the SMS message, it is difficult to cover the whole domain in the hierarchy and the relation of the ontology. Therefore, we need to resort to other means to enrich the ontology. One way is to use Arabic WordNet to support the ontology instances by word synonyms. This is important to achieve better results in classifying SMS messages as will be presented in Chapters 5 and 6. Additionally, we need to define a set of semantic rules based on the ontology as a necessary step to classify messages reflecting the manual process of filtering. This is presented in Chapter 5.

#### 4.10 Summary

In this chapter, we have described the development and evaluation of the Arabic SMS spam domain ontology. We followed an ontology development steps to build the ontology. At the beginning, we identified the domain and scope of the ontology. Then we defined the terms and the properties. We have used the ontology development protégé OWL to implement and realize the ontology. We have added individuals to

ontology (i.e. creating knowledge base) and explained some of the factors that are related to the values of some properties. Then we have presented an evaluation of the ontology and proved that the ontology has answered all questions and returns the correct results. In the next chapter, we build the approach that uses the Arabic SMS Spam ontology to detect and filter spam messages.

# **Chapter 5**

## **Arabic SMS Spam Detection**

## Chapter 5

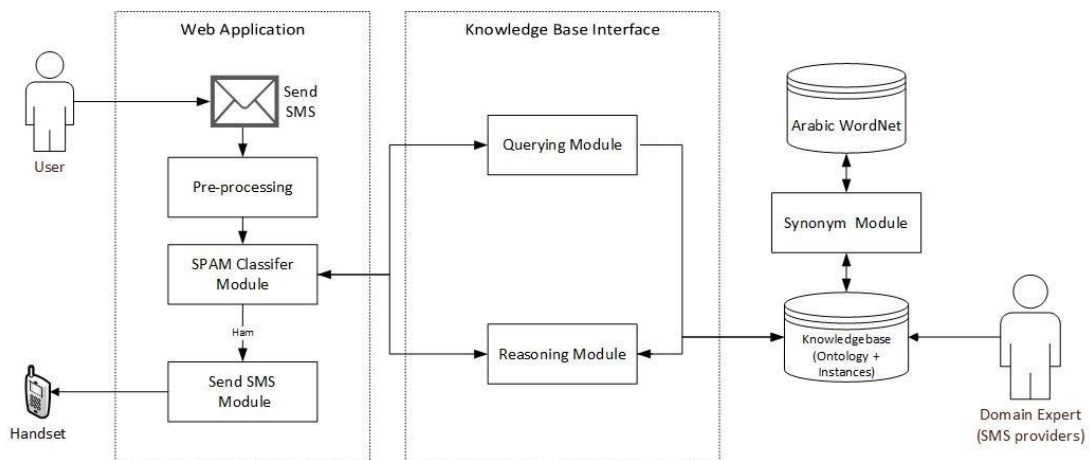
### Arabic SMS Spam Detection

This chapter discusses the design and development of an approach used for applying real time SMS spam detection, classification and hence filtering. The approach depends on the Arabic SMS spam ontology to check the text of the message and to make sure that a message is spam free.

We first describe the overall structure of approach, then describe the elements of the approach and the processing steps of the approach based on the functionality of these elements.

#### 5.1 Overall Structure of The Approach

The SMS spam detection and classification approach consists of the following modules as shown in Figure (5.1).



**Figure (5.1):** Structure of the SMS spam detection and classification approach

**SMS Spam Knowledge Base:** The most important part of our approach is the knowledge base which consists of two parts. The first part is the SMS spam ontology. The second part is the SMS message instances that are inserted and linked to the ontology and contains vocabulary of SMS spam words and semantic relations between these words together with weights of these words. The knowledge base is the ultimate target for classifying and detecting if an SMS message is spam or legitimate.

**Synonym Module:** Through this module, ontology terms and instance can be related to the synonyms from Arabic WordNet to enrich the knowledge based with new vocabulary. This enrichment helps to keep the knowledge based updated, hence, improves the SMS message classification and spam detection.

**Querying Module:** Through using this module, it can answer very specific quires with reasoning that would be difficult to looking it at ontology directly. We can use SPARQL queries to extract, filtering, classification with your data, and to summarize knowledge from the proposed ontology.

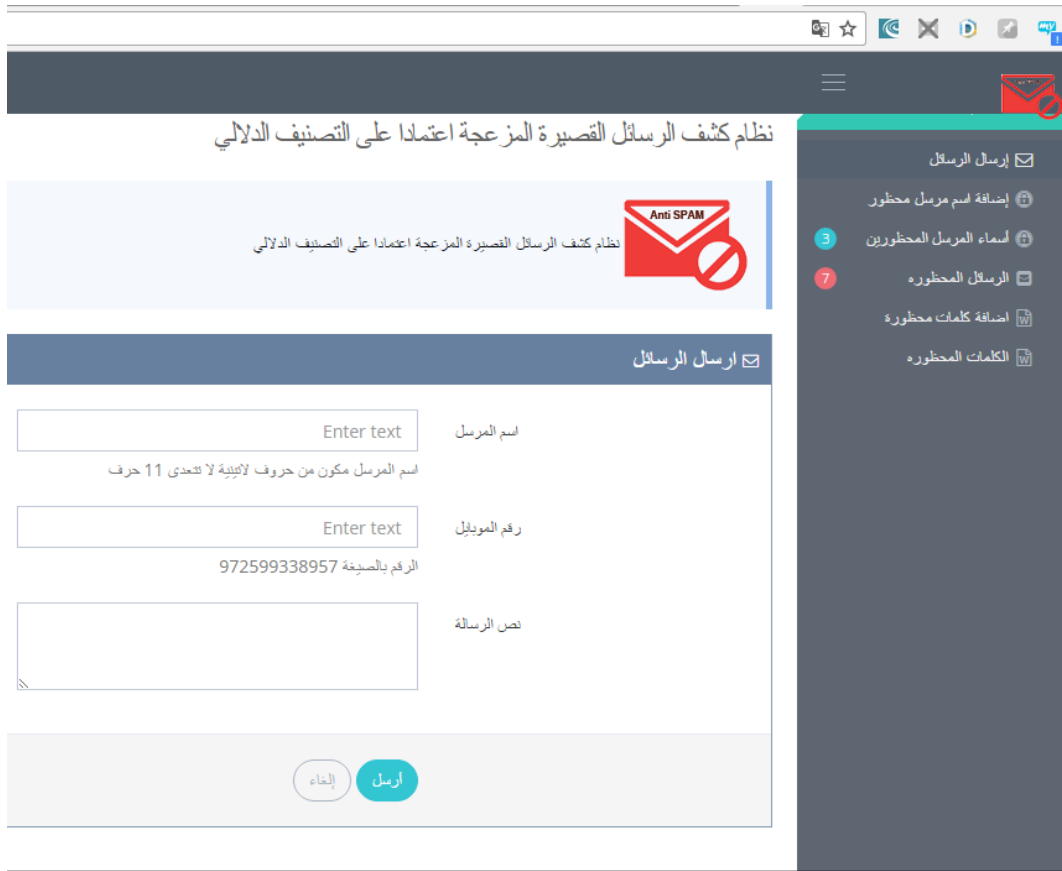
**Reasoning Module:** This module includes an OWL inference engine (i.e. JENA Reasoner). All inferred information is stored as new triples in dictionary thus exposing them to the queries. This enables the declaration of derived classes or the declaration of further property characteristics (e.g. transitivity and symmetry of properties) and the semantic rules. The Reasoning Module was implemented in Java by using the JENA API. It utilizes the rules to get best classification of SMS.

**Spam Detection (Classifier) Module:** This module will receive SMS from the user through the web application then decides if the SMS is spam or not by sending some specific classification rules to the reasoning module to applying and running it on the ontology, then using the querying module to get the results from the ontology, finally will send these results to the send SMS module to send SMS if not spam.

**SMS message Sending Module:** This module is responsible for sending SMS to mobile operators to deliver it to handset users.

**User Web Interface:** We implemented a prototype for the proposed SMS spam detection approach using Java and JENA library to perform reasoning and querying in the knowledge base, HTML web page for user interface which is used by clients to send SMS requests as shown in Figure (5.2) and is used by the administrator to add new spam words to the SMS spam knowledge base.





**Figure (5.2):** User interface for sending SMS messages

## 5.2 Functionality of The SMS Spam Detection Approach

The overall functionality of the approach is illustrated in Figure (5.3) and is explained in the following steps:

**Step 1:** The user sends an SMS message using the web interface shown in Figure (5.2).

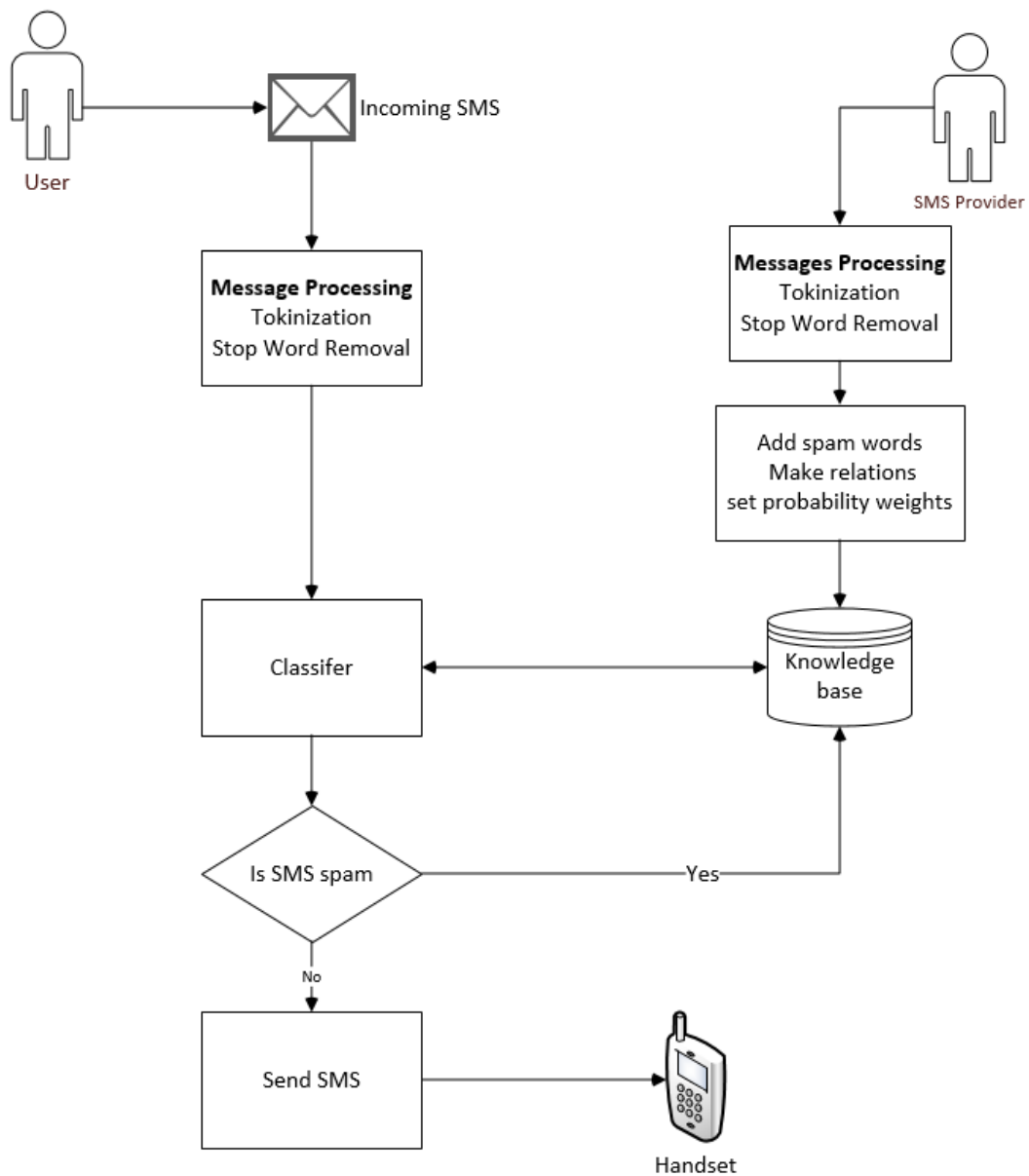
**Step 2:** The SMS message is processed by performing tokenization, stop words removal and tagging.

**Step 3:** The processed message is transferred to the classifier to decide if it is spam or legitimate. To perform this classification, the classifier depends on the knowledge base (the ontology and stored SMS spam instances), a set of SMS spam detection semantic rules, reasoning to infer spam filtering and classification and a SPARQL queries to return the classification result.

**Step 4:** If the classification results in the SMS message as legitimate, the message is sent to the handset of the user(s).

**Step 5:** If the classification results in the SMS message as spam, the message is added to the spam knowledge base and the spam words weights and sender name rank are updated.

At any time and irrespective of the above steps, the SMS provider can enrich the knowledge base manually with new SMS spam words, after making necessary preprocessing on manually classified messages, setting new weights probabilities for these spam words.



**Figure (5.3):** SMS spam detection

On the following sections, we describe and discuss the detailed steps and elements of the approach. We start with the data collection step.

### 5.3 Data Collection

Preparing and collecting the corpus is one of the most important stages in the research project. The corpus is a collection of SMS messages. we collected about nearly 1409 SMS messages. They were collected manually from local SMS providers.

The collected messages are chosen from classified area like “سياسي” (Politics), “إباحي” (pornographic), “تحايل تجاري” (commercial Scam) and “تصيد” (Phishing). In Figure (5.3) show commercial SMS spam dataset collected.

no	Sender	text	Class	type
1	Turkish	آخر أحداث وأخبار الدراما التركية مع خدمة "تركي" بس ب 20 أغورة! للاشتراك أرسل حرف "ت" للرقم 37189	spam	تجاري
2	TASHBIK.COM	أحصل على 500 رسالة SMS بـ 10 نيكيل فقط. أرسل كلمة سجل للرقم 37775	spam	تجاري
3	Tadabor	تحلّي بأخلاق القرآن الكريم مع خدمة "كنس القرآن". للاشتراك أرسل "ت" للرقم 37929. للشروط والنفاسيل اتصل مجانياً على 15147	spam	تجاري
4	SurpriseSMS	عزيزي المشترك، نقدر أن نلتفت العالية بنا وضمن بقية مفاجآت الربيع من جوال يسرنا أن نقدم لك 10 دقائق اتصال مجانية باتجاه شبكة جوال اليوم الجمعة	spam	تجاري
5	SurpriseSMS	عزيزي المشترك، نمتع بحزم الانترنت اسبوعية ويومية لا محدودة بسعر مخفض جداً. للنفاسيل والشروط اتصل مجانياً على 15165	spam	تجاري
6	SurpriseSms	نمتع ب 35 دقيقة اتصال يومياً باتجاه جوال ب 1 نيكيل فقط. للاشتراك أرسل رساله تحتوي كلمة SUB للرقم 37015 بسعر 19 اغورة غير شاملة ض.	spam	تجاري
7	SpicyNews	فقط ب 20 اغورة يومياً يتوصلك معلومات طريفة من أنحاء العالم، للاشتراك أرسل "س" للرقم 37918	spam	تجاري
8	Sandouk	اربع سيارة هونداي فولستر 2015 من برنامج "المنطوق". للاشتراك أرسل رسالة فارغة للرقم 37513. للشروط والنفاسيل اتصل مجانياً على 15103	spam	تجاري
9	Rotaniat	لمشاهدة أحدث الفيديو كليبك من روتانا على جوالك، للاشتراك أرسل "ف" للرقم 37845. للشروط والنفاسيل اتصل مجانياً على 15173	spam	تجاري
10	Personality	انترك بخدمة "اعرف شخصيتك" واحصل على اسبوع مجانياً على الخدمة للاشتراك أرسل تاريخ ميلادك مثل: 2/12/1990 للرقم 37194. للشروط والن	spam	تجاري
11	Personality	لتعرف صفات شخصيتك أو أي شخصية يتخيلك. أرسل تاريخ ميلادك المكون من اليوم ثم فراغ ثم الشهر للرقم 37795 مثال "14 فراغ 7"، للشروط و	spam	تجاري
12	NabedWatan	أجمل الكلمات والأشعار الوطنية على جوالك مباشرة! للاشتراك أرسل "و" للرقم 37903	spam	تجاري
13	Monajah	إبدأ صباحك بأجمل الأدعية من اذاعة القرآن الكريم مع خدمة "مناجاة". للاشتراك أرسل "مناجاة" للرقم 37792. للشروط والنفاسيل اتصل مجانياً على 0	spam	تجاري
14	Melody	تابع أخبار الفنانين من ميلودي بـ 20 اغورة يومياً! للاشتراك أرسل حرف "ب" للرقم 37929	spam	تجاري
15	MBC	بنك تحقّق حلمك؟ أكثر من مليون دولار ل 7 راجين! أرسل "حلم" للرقم 37890 وانترك لتكون انت الراجين □□□	spam	تجاري
16	Love	انترك بخدمة "حب" واحصل على 3 أيام مجانياً على الخدمة! للاشتراك أرسل "ح" للرقم 37186، للشروط والنفاسيل اتصل مجانياً على الرقم 15143	spam	تجاري
17	Kooora	نمتع بتنظية كامله لأخر البطولات والأحداث الرياضية من الموقع الرياضي العربي الأول "كooورة" فقط ب 20 اغورة للرسالة. أرسل "ك" للرقم 7936	spam	تجاري
18	Khusomat	خصومات وعروض حصورية من محلات طوكريم وجين وكثلية مباشرة على جوالك! أرسل "ن" للرقم 37706	spam	تجاري
19	Jobs	بنك وظيفة؟ استقبل الوظائف الشاغرة على جوالك في الضفة وغزة من موقع Jobs.ps للاشتراك أرسل "ض" أو "ع" للرقم 37953	spam	تجاري
20	phishing	مباراة شيفة ومجموع جوائزنا \$7000! توقع الفائز واربح الآن! أرسل 55 لثاناً أو 66 للبريد للرقم 37877. للشروط والنفاسيل اتصل مجانياً على 122	spam	تجاري

Figure (5.3): Part of the SMS dataset

### 5.4 Data Preprocessing

The step of preprocessing includes a tokenization stage, stop word removal and POS tagging stage.

**Tokenization:** Tokenization is the process of breaking a stream of text up into words, phrases, or other meaningful elements called tokens. Tokens are separated by whitespace characters, line breaks. All terms and text in ontological dictionary are normalized for pre-processing to map them with words in the message. This involves the following steps (Dilekh & Behloul, 2012) :

- Remove punctuation
- Remove diacritics (primarily weak vowels)
- Remove non-letters
- Replace the ة or the أ initial by bare Alif ا
- Replace the إ by the ا
- Replace the ء of order by the ئ
- Replace the ى final by the ي
- Replace the ة final by the ه

These steps are encoded programmatically in Java as shown in Figure (5.4).

```

for (String value : Words) {
    value = value.replace(" ", "_");
    value = value.replace("ة", "ه");
    value = value.replace("إ", "ا");
    value = value.replace("أ", "ا");
    value = value.replace("ء", "ئ");
    value = value.replace("ى", "ي");
    value = value.replace("ال", "");
    Individual wn = ontclass.createIndividual(ns + value);

    wordind.addProperty(property, wn);
}

```

**Figure (5.4):** Replacing some characters in the process of tokenization

**Stop Word Removal:** Stop words are those words which rarely contribute to useful information in terms of a document relevance and appear frequently in text but provide less meaning in identifying the important content of the document. Those words include prepositions, conjunctions and other high frequency words. Figure (5.5) shows some of these words.

ان بعد ضد يلي الى في من حتى وهو يكون به وليس أحد على وكان تلك كذلك التي وبين فيها عليها  
 إن وعلى لكن عن مساء ليس منذ الذي أما حين ومن لا ليسب وكانت أي ما عنه حول دون مع لكنه  
 ولكن له هذا والتي فقط ثم هذه أنه تكون قد بين جدا لن نحو كان لهم لأن اليوم لم هؤلاء فإن فيه ذلك لو  
 عند اللذين كل بد لدى وثي أن ومع فقد بل هو عنها منه بها وفي فهو تحت لها أو إذ علي عليه كما كيف  
 هنا وقد كانت لذلك أمام هناك قبل معه يوم منها إلى إذا هل حيث هي إذا او و ما لا الي إلي مازال لازال  
 لايزال مايزال اصبح أصبح أمسى أمسى أضحى أضحى ظل مابرح مافتئ مافتئ بات صار ليس إن  
 كأن ليت لعل لاسيما ولايزال الحالي ضمن اول وله ذات اي بدلا اليها انه الذين فانه وان والذي وهذا  
 لهذا الا فكان ستكون مما أبو بن الذي اليه يمكن بهذا لدي وأن وهي وأبو آل الذي هن الذي

**Figure (5.5):** Examples of stop words

**Tagging:** we use a tagger in order to determine the type of a word as verb or noun. This will be used later to classify spam verbs in “فعل” class.

An example of a data processing:

Message:

"إربح سيارة هونداي فولستر 2015 من برنامج "الصندوق" للإشتراك أرسل رسالة فارغة للرقم 37513، للشروط والتفاصيل اتصل مجاناً على 15103".

1-Tokenization

"إربح سيارة هونداي فولستر 2015 من برنامج صندوق اشتراك أرسل فارغه رقم 37513 شروط و تفاصيل اتصل مجانا على 15103".

2-Stop word removal

"إربح سيارة هونداي فولستر 2015 برنامج صندوق اشتراك أرسل فارغه رقم 37513 تفاصيل اتصل مجاناً 15103".

3-Tagging

Verbs: "إربح أرسل اتصل".

Noun: "سيارة هونداي فولستر برنامج صندوق فارغه تفاصيل مجاناً".

Numbers: "2015 37513".

## 5.5 Word Extraction, Matching with WordNet

After data processing is performed, we need to extract words from the messages to enrich the ontology with SMS spam vocabulary. Then match the synonyms of these words (vocabulary) from Arabic WordNet dataset.

Then we estimating probabilities of spam words using Bayes conditional probability theorem according to which the probability of a word given that the message is spam can be estimated (Abdoh, Musa, & Salman, 2009) as follows:

$$P_s = \frac{\frac{F_s}{C_s}}{\frac{F_{ns}}{C_{ns}} + \frac{F_s}{C_s}}$$

Where:

$P_s$  is the probability of a word given the SMS is spam.

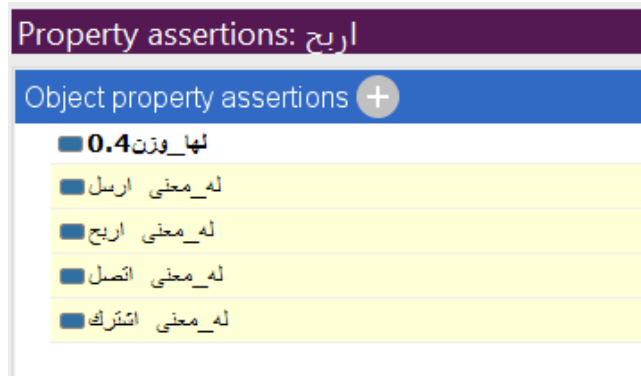
$F_s$  is the frequency of word in the all SMS spam dataset.

$F_{ns}$  is frequency of words in the all SMS legitimate dataset.

Cs is the count of SMS spam dataset.

Cns is the count of SMS legitimate dataset.

After estimate probability weights of spam words we set it as instances with relation of object property has\_weight “لها\_وزن”. As shown in Figure (5.6).



**Figure (5.6):** Setting the probability weight for spam words

In Figure (5.7) and Figure (5.8) the code to get synonym of word from Arabic WordNet.

```
41
42 public List<String> WN_synonym(String word) {
43
44     Awn awn = new Awn("c:\\upc_db_dic.xml", false);
45
46     List<String> synonyms = new ArrayList<>();
47     List<String> listItemID = new ArrayList<>();
48     List<String> listWordID = awn.Get_List_Word_Id_From_Value(word);
49
50     if (listWordID.size() != 0) {
51         for (int i = 0; i < listWordID.size(); i++) {
52             listItemID.add(awn.Get_Synset_ID_From_Word_Id(listWordID.get(i)));
53         }
54     } else {
55         listItemID = awn.Get_Item_Id_From_Name(word);
56     }
57
58     for (int i = 0; i < listItemID.size(); i++) {
59         List<String> synsetList = awn.Get_List_Word_Id_From_Synset_ID(listItemID.get(i));
60         for (int j = 0; j < synsetList.size(); j++) {
61             String synset = awn.Get_Word_Value_From_Word_Id(synsetList.get(j));
62             if (synset.equals(word) == false) {
63                 synonyms.add(synset);
64             }
65         }
66     }
67
68     return synonyms;
69 }
```

**Figure (5.7):** Getting synonym of words from Arabic WordNet to enrich the SMS spam Knowledge base

As shown in Figure (5.7) the WN\_synonym function has a parameter for word as string and it returns a list of strings for synonym words. Line 44 in Figure (5.7) uses Arabic WordNet class (AWN) API to parse XML file. After we get the synonyms of words we need to make object property relations such as “له معنى” (has\_synonym) and add it in ontology as instances as shown in Figure (5.8).

```
159  
160         OntClass ontclass = model.getOntClass(ns + "WordNet_Synonym");  
161         Individual wordind = model.getIndividual(ns + word);  
162         Property property = model.getProperty(ns + "له معنى");  
163         Individual wn = ontclass.createIndividual(ns + value);  
164         wordind.addProperty(property, wn);  
165
```

**Figure (5.8):** Adding synonym of a word from Arabic WordNet to the ontology

## 5.6 Building The Ontology

The ontology play a major role in the process of SMS spam classification approach. The ontology together with the various spam words and spam messages form the knowledge base of the SMS spam. The details of the ontology building process together with the knowledge base is covered in Chapter 4. The knowledge base can be enriched with new spam vocabularies through manually adding new spam words and messages and additionally through the Arabic WordNet synonyms (see Section 5.9). The system refers to the knowledge base to perform querying as well as reasoning needed in the decision as whether a given message is spam or legitimate. This is aided by a set of SWRL rules as explained next.

## 5.7 Create Semantic Rules

In this section, we describe the definition of a set of rules which provide significant help to obtain satisfying results from the knowledge base. Before applying the reasoner, we need to define important rules to make necessary to refer to the ontology and classify SMS messages as spam or legitimate.

### Example 1:

The first set of semantic rules allow to check the sender name of SMS, which is a numeric or alphabet characters. The numeric sender name will be blocked because it is not allowed to be used in BulkSMS service. The alphabet sender name is checked using another rule. If this sender name is already blocked for several reasons such as legal issues according to the operator's requests. An example of this set of rules is:

**Rule 1:** TextMessages(?msg) ^ اسم\_المرسل(?msg, ?sndr) ^ numeric(?sndr) -> محظور(?sndr, true)

This rule state that: if numeric "اسم المرسل" (sender name) the message then it will be "محظور" (blocked). The result of this rule will be used in the next rule to detect the SMS spam messages.

**Rule 2:** TextMessages(?msg) ^ اسم\_المرسل(?msg, ?sndr) ^ محظور(?sndr, true) -> تصنيف(?msg, SPAM)

This rule states that: if message has a "اسم المرسل" (sender name), and this sender name is "محظور"(blocked) previously for a reason, as we mentioned before, are "تصنيف" (classified) as a spam message.

### Example 2:

The next set of semantic rules allow to check the entent of every word in the text of the message, which is explicit or implicit meaning. We can detect the explicit meaning by checking the name of the class that contains the desired word. If the word is classified as spam such as "إباحي" (Pornographic), the message is classified as spam. An example of this set of rules is:

**Rule 3:** TextMessages(?msg) ^ تحتوي\_على(?msg, ?word) ^ اباحي(?word) -> تصنيف(?msg, SPAM) ^ سبب\_الحظر(?msg, اباحي)

This rule state that: if the messages has a word that included in the spam class "إباحي" (Pornographic), then it classified in class "تصنيف"(classify) spam.



We can also detect the implicit meaning of a word by weight calculation of the message by summation of every weight of the included words by summation function. That can not make it in rule due to JENA Rules and SWRL Rules is monotonic rules so the counting, modification not supported.

**Rule 4:** TextMessages(?msg) ^ وزن\_الرسالة (?msg,?msgWeight) ^  
 swrlb:greaterThan(?1, ?msgWeight) ->  
 تصنيف(?msg, SUSPICIOUS)

This rule state that: if a text messages has a message "وزن\_الرسالة" (message\_weight) and if the weight of the message is greater than "1", then the message will be considered suspicious message, to enter again to another checking phase. The result of previous rule will be used in the next rule to detect the implicit spam meaning of SMS messages by checking the relations of words to other words that included in the message. An example of this set of rules is:

**Rule 5:** TextMessages(?msg) ^ تحتوي\_على (?msg, ?word1) ^  
 تحتوي\_على (?msg, ?word2) ^ differentFrom(?word1, ?word2) ^  
 تجاري(?word1) ^ فعل(?word2) ^ يوجد\_فعل (?word1, ?word2) ^  
 تصنيف(?msg, SUSPICIOUS) ->  
 تجاري(?msg, SPAM) ^ سبب\_الحظر (?msg, SPAM)

This rule state that: if a text messages contains a two different words "word1" and "word2", and the class name of the first word "word1" is classified as a spam such as "تجاري" (commercial), and the second word "word2" is a "فعل"(verb), and there is relations "يوجد\_فعل" (has\_verb) for "word1" and "word2" and the classification of the text message is "SUSPICIOUS" based on the Rule 4, then the message is considered a spam message for "تجاري" (commercial) reason.

### Example 3:

The next set of semantic rules allow to check the synonym of every word in the message weather, it has a spam meaning or not and this is by adding the new synonym to the class that contains spam words. Some examples of these set of rules, the first one is:

**Rule 6:** TextMessages(?msg) ^ تحتوي\_على (?msg, ?word) ^  
 -> (?synonym) تجاري (?synonym) له\_معنى  
 (?word) تجاري

This rule state that: if a text messages contains a word "word", and this word has a synonym word "synonym", and this synonym is included in class that contains spam word such as "تجاري" (commercial), then the word "word" will be added to the class "تجاري" (commercial).

The second example is:

**Rule 7:** TextMessages(?msg) ^ تحتوي\_على (?msg, ?verb) ^  
 -> (?synonym) فعل (?synonym) له\_معنى  
 (?verb) فعل

This rule state that: if a text messages contains a word "verb", and this word has a synonym word "synonym", and this synonym is included in class that contains "فعل" (verbs), then the word "verb" will be added to the class "فعل" (verb), then both above Rule 6 and Rule 7 are used as part of Rule 5 to detect SMS spam.

Examples of some of the previous rules and how they look in Protégé OWL are shown in Figure (5.9). The implementation of these rules in JENA are shown in Figure (5.10).

Rule
TextMessages(?msg) ^ له_رقم (?msg, true) ^ تحتوي_على (?msg, ?word) ^ تصيد (?word) -> تصنيف (?msg, SPAM)
سياسي (?word) له_معنى (?word, ?m) -> سياسي (?m)
TextMessages(?msg) ^ مرتبط_بشئيه (?word, ?x) ^ شئيه (?word) ^ سياسي (?word) ^ تحتوي_على (?msg, ?word) ^ تحتوي_على (?msg, ?word) ^ تصنيف (?msg, SPAM)
TextMessages(?msg) ^ مرتبط_بشئيه (?word, ?x) ^ شئيه (?word) ^ شخص (?word) ^ تحتوي_على (?msg, ?word) ^ تحتوي_على (?msg, ?word) ^ تصنيف (?msg, SPAM)
TextMessages(?msg) ^ محظور (?sndr, true) -> تصنيف (?msg, SPAM)
TextMessages(?msg) ^ اسم_المرسل (?msg, ?sndr) ^ numeric (?sndr) -> تصنيف (?msg, SPAM)
TextMessages(?msg) ^ فعل (?word3) ^ شئيه (?word2) ^ مكان (?word) ^ مكالمة (?word3) ^ تحتوي_على (?msg, ?word3) ^ تحتوي_على (?msg, ?word2) ^ تحتوي_على (?msg, ?word) ^ تصنيف (?msg, SPAM)
TextMessages(?msg) ^ اباحي (?word) -> تصنيف (?msg, SPAM)
TextMessages(?msg) ^ يوجد_فعل (?word, ?x) ^ فعل (?x) ^ سياسي (?word) ^ تحتوي_على (?msg, ?word) ^ تحتوي_على (?msg, ?word) ^ تصنيف (?msg, SPAM)

**Figure (5.9):** Some rules to classify messages using ontology terms and reasoning

```

public List<String> spamSMS() throws IOException {
    List<String> list = new ArrayList<String>();
    OntModel model = ModelFactory.createOntologyModel();
    FileManager.get().readModel(model, "http://localhost:8080/AntiSPAM/assets/SMS.owl");

    String rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#";
    String abc = "http://www.semanticweb.org/ayman/ontologies/2016/7/sms#";

    String rule = "[ (?msg " + rdf + "type " + abc + "TextMessages), "
        + "(?msg " + abc + "على_تحتوي?word), "
        + "(?word " + rdf + "type " + abc + "اباحي) "
        + "->"
        + "(?msg " + abc + "تصنيف " + abc + "SPAM) ]";
}

```

**Figure (5.10):** Writing a rule in JENA

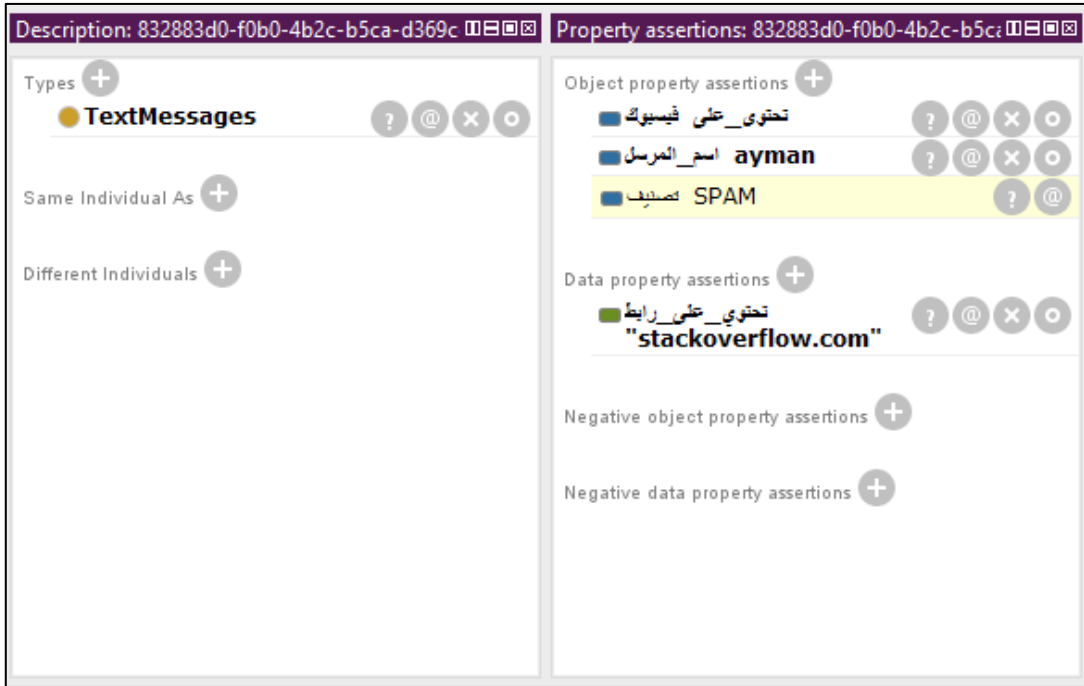
## 5.8 Apply Ontology Reasoner

After creating instances, we apply an ontology reasoner e.g. Hermit reasoner on the ontology by protégé and JENA reasoner on the ontology and RDF data. This is necessary to identify new relations from existing ones. The reasoner is able to identify the different types of ontological relations such as transitive, symmetric, inverse and functional properties and use them to add new facts. So, when we run a reasoner and perform reasoning on the ontology, we get new or hidden knowledge utilized in the ontology. This reasoning process are aided by the semantic rules defined before.

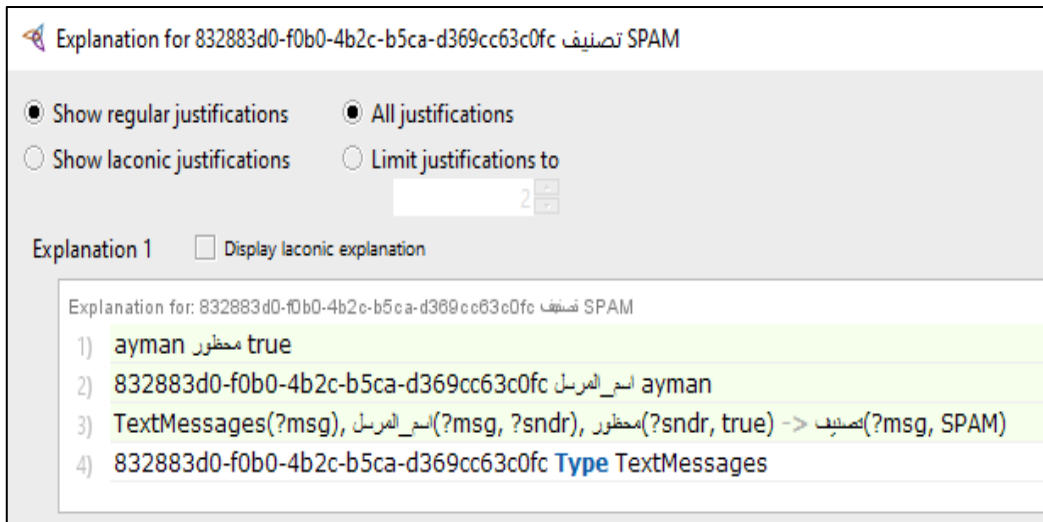
Semantic rules added to ontology can be used by the reasoner to give new facts. For example, if the sender name of a message is blocked, then the reasoner classifies it as SMS spam based on this rule as shown in Figure (5.11) and based on the following rule:

**TextMessages(?msg) ^ اسم\_المرسل(?msg, ?sndr) ^ محظور(?sndr, true) -> تصنيف(?msg, SPAM)**

The rule together with the explanation of its results are shown in Figure (5.12). The explanation says that the rule resulted in the message as spam because it is matched, i.e., the sender name is blocked using the “محظور” (blocked) property.



**Figure (5.11):** Results of reasoner use semantic rules



**Figure (5.12):** Explanation of the inferred message classification with the used semantic rule

Based on the basic functionality of the approach presented on section 5.3, we present in some details these functionalities as they are related to the sender of the message and the provider of the spam filtering system. These functionalities include adding new relations to the ontology, querying and classification.

## 5.9 Adding New Relations to Ontology

In our approach, the spam filtering system can automatically or SMS providers can manually add new relations to the ontology to enrich knowledge base. We cite an example for each kind.

### Example 1: Adding instances by the system

- 1- System can add rank for sender name if the message is classified as spam after the sender name exceeds 10 times of SMS spam as shown in Figure (5.13). We update the number of spams send by this sender using data property “عدد\_الرسائل\_المحظورة” (no\_blocked\_sms). Figure (5.14) shows the result of executing such a code where it returned 5 as the number of spams send by this sender.

```
749 Individual senderIndv = senderClass.createIndividual(ns + sender);
750 Property property = model.getProperty(ns + "عدد_الرسائل_المحظورة");
751 model.addLiteral(senderIndv, property, i);
752 senderIndv.addProperty(property, senderIndv);
```

**Figure (5.13):** Adding rank (data property) for sender name



**Figure (5.14):** Data property “عدد\_الرسائل\_المحظورة”

- 2- SMS provider can add new instances and relations to knowledge base as shown in Figure (5.15). He can add new spam words and make relations among other spam words using object and data properties which are defined in the ontology. The figure shows the interface to enter the word in the ontology including its

classification, its synonyms from the Arabic WordNet, its relations to specific spam words such place, time, and revile and whether it includes a URL.



**Figure (5.15):** User interface for SMS provider to adding new words to the SMS ontology

Figure (5.16) shows part of the code to add a word and make a relation with other words using object and data properties.

```

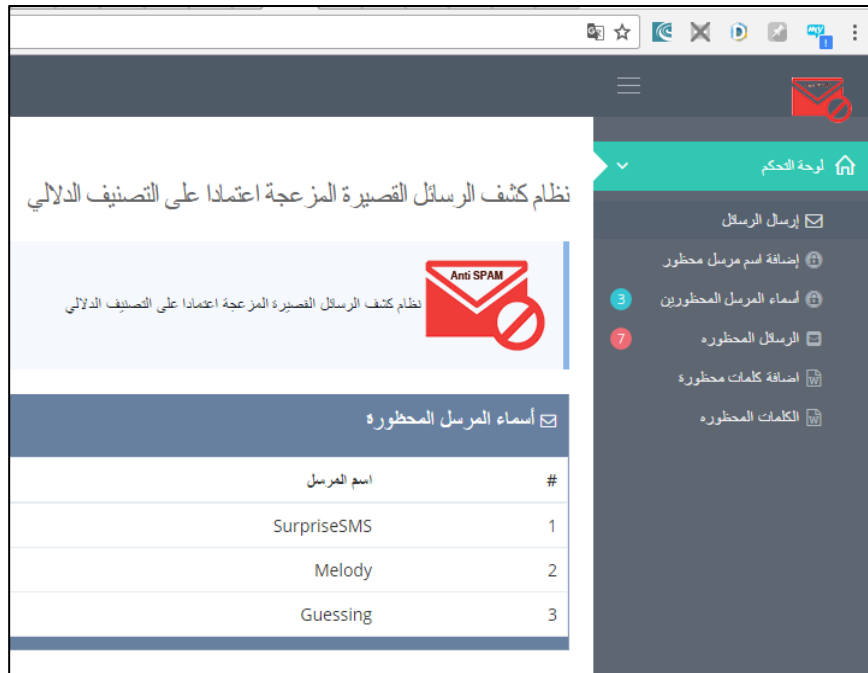
298
299         if(!verb.equals("")){
300             Individual wordind = model.getIndividual(ns + verb);
301             Property property = model.getProperty(ns + "فعل_يوجد");
302             in1.addProperty(property, wordind);
303         }
304         if(!url.equals("")){
305             Property property = model.getProperty(ns + "رابط_له");
306             model.addLiteral(in1, property, url);
307         }
308

```

**Figure (5.16):** Adding new words and relations to the ontology using object and data properties

## 5.10 Querying

The semantic querying allows to perform query statements which are written in SPARQL Query or DL-Query. This semantic queries enable the system to retrieve both explicitly and implicitly derived information. For example, we can retrieve blocked sender name from the ontology as shown in Figure (5.17).



**Figure (5.17):** Interface to show all blocked sender names in the knowledge base

Figure (5.18) shows the code to use SPARQL query to retrieve all blocked sender names from the knowledge base.

```
555 OntModel model = ModelFactory.createOntologyModel();
556 FileManager.get().readModel(model, "http://localhost:8080/AntiSPAM/assets/SMS.owl");
557 Query query = QueryFactory.create(
558     "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>\n" +
559     "PREFIX owl: <http://www.w3.org/2002/07/owl#>\n" +
560     "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>\n" +
561     "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>\n" +
562     "PREFIX a: <http://www.semanticweb.org/ayman/ontologies/2016/7/sms#>\n" +
563     "\n" +
564     "SELECT DISTINCT ?subject \n" +
565     "WHERE { \n" +
566     "\n" +
567     "?subject rdf:type ?object . \n" +
568     "?object rdfs:subClassOf a:Sender. \n" +
569     "?subject a: محظور?x.\n" +
570     "\n" +
571     "}"
572 );
573 QueryExecution qe = QueryExecutionFactory.create(query, model);
```

**Figure (5.18):** Retrieving all blocked sender names from the knowledge base

## 5.11 Classification

Classification of SMS are the core of the approach as we explain in Section (5.1). The knowledge base plays an important role in the system where it stores the knowledge about blocked sender names, forbidden words, and relations of these words and their synonyms.

Based on the step of creating rules and the step of reasoning, we can classify new SMS messages sent to the system to detect if they spam or legitimate.

An example of a semantic rule (phishing rule) that is used for classifying SMS spam is shown in Figure (5.20).

```
OntModel model = ModelFactory.createOntologyModel();
FileManager.get().readModel(model, "http://localhost:24840/AntiSPAM/assets/SMS.owl");
String rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#";
String abc = "http://www.semanticweb.org/ayman/ontologies/2016/7/sms#";
String xsd = "http://www.w3.org/2001/XMLSchema#";
String rule = "[ (?msg " + rdf + " type " + abc + " TextMessages), "
    + " (?msg " + abc + " على_تحتوي?word) "
    + " (?msg " + abc + " رابط_على_تحتوي?link1) "
    + " (?word " + rdf + " type " + abc + " نصيد) "
    + " (?word " + abc + " رابط_له?link2) "
    + " notEqual(?link1,?link2) "
    + "->"
    + " (?msg " + abc + " تصنيف " + abc + " SPAM) ]";
Reasoner reasoner = new GenericRuleReasoner(Rule.parseRules(rule));
InfModel infModel = ModelFactory.createInfModel(reasoner, model);
```

**Figure (5.20):** Running the phishing semantic rule to classify a message

After classifying a message, we need to make SPARQL query to get the message if it was classified as spam. This is shown in Figure (5.21).

```
443
444
445     Query query = QueryFactory.create(
446         "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>\n"
447         + "PREFIX owl: <http://www.w3.org/2002/07/owl#>\n"
448         + "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>\n"
449         + "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>\n"
450         + "PREFIX a: <http://www.semanticweb.org/ayman/ontologies/2016/7/sms#>\n"
451         + "SELECT ?subject\n"
452         + "    WHERE { ?subject a: تصنيف:SPAM }"
453     );
454     QueryExecution qe = QueryExecutionFactory.create(query, infModel2);
455     ResultSet rs = qe.execSelect();
```

**Figure (5.21):** SPARQL query to return a messaged classified as spam



## **5.12 Summary**

In this chapter, we have described the Arabic SMS spam filtering approach. We have presented the phases of building the approach which includes collection of data, building the knowledge base (ontology and instances), creating the semantic rules and reasoning. Finally, we commented on the implementation of the approach including some usage examples through the user interface.

# **Chapter 6**

## **Results and Discussion**

## Chapter 6

### Results and Discussion

This chapter presents the experiments performed for the evaluation of the SMS spam filtering approach. We discuss the results of the experiments and evaluate the approach through a number of measures including accuracy, precision, recall and f-measure. Finally, we give a short comparison with the results of classification using Naïve Bayes method.

#### 6.1 Experiments

We performed a number of experiments to demonstrate the ability of our approach to classify SMS messages based on the constructed Arabic SMS spam ontology and knowledge base.

In the first stage SMS messages are collected from the local BulkSMS providers. It contains 1409 manually labeled, which are divided into two groups as shown in Table (6.1). which are Legitimate (Non-Spam) with a total of 1161 messages. Spam with a total of 248 SMS messages.

**Table (6.1):** SMS data set

	<b>Amount</b>	<b>Percentage</b>
<b>Legitimate (Non-Spam)</b>	1161	82.40%
<b>Spam</b>	248	17.60%
<b>Total</b>	1409	100%

The SMS spam contains four types of SMS messages as political messages, commercial and promotional messages, pornographic and adult messages, scam and phishing messages.

In the second stage for the experimentation, three different types of experiments are constructed. First, we built the ontology concepts using 100 SMS spam messages without WordNet semantic rules. Second, we built the ontology concepts using 200 SMS spam messages without WordNet semantic rules. Third, we built the ontology concepts using 300 SMS spam messages with WordNet semantic rules. Afterward we

tested the proposed approach by sending a serial of SMS messages to the classifier from dataset.

Next, we explain and discuss the results of these experiments based on the evaluation metrics.

## 6.2 Evaluation Metrics

A classification task involves assigning which out of a set of categories or labels should be assigned to some data according to some properties of the data. The spam filtering in our approach assigns a spam or no spam status to every SMS message. Therefore, it is binary classification which accuracy in this case can hide the detail we needed to check the performance of approach due to the limited two categories in our case.

A confusion matrix is a summary of results on a classification case. The number of correct and incorrect predictions are summarized with count values and broken down by each class and are presented as a confusion matrix. For evaluating the performance of spam detection, basic measures that we can use are Accuracy (ACC), Error rate (ERR), Precision, Recall, F-measure and Matthews Correlation Coefficient (MCC). The evaluation metrics were defined based on the confusion matrix, as shown in equations (1) to (6) in section 2.15.

## 6.3 Evaluation Results

The results of the 3 experiments based on the confusion matrix are shown in Table (6.2).

**Table (6.2):** Confusion Matrix results

		Predicted					
		Exp 1		Exp 2		Exp 3	
		Spam	Legitimate	Spam	Legitimate	Spam	Legitimate
n = 1409							
Actual	Spam	209	39	217	31	233	15
	Legitimate	51	1110	39	1122	33	1128

Based on these 3 confusion matrices, we compute the Accuracy (ACC), Error rate (ERR), Precision, Recall, F-measure and Matthews Correlation Coefficient (MCC) for the 3 experiments.

### **Accuracy (Acc)**

The Accuracy is a measure of the overall correctness of the approach, it's the number of SMS that are correctly classified divided by sum of the total SMS.

$$\text{Accuracy(Acc)} = \frac{209 + 1110}{209 + 39 + 51 + 1110} = 0.936 \quad (1.1)$$

$$\text{Accuracy(Acc)} = \frac{217 + 1122}{217 + 31 + 39 + 1122} = 0.950 \quad (1.2)$$

$$\text{Accuracy(Acc)} = \frac{233 + 1128}{233 + 15 + 33 + 1128} = 0.965 \quad (1.3)$$

To see overall performance of approach we will calculate other metrics as explain next.

### **Error rate (ERR)**

The Error rate (ERR) is a prediction error metric for a binary classification problem. It is calculated as the number of all incorrect predictions divided by the total number of SMS. The best error rate is 0.0, whereas the worst is 1.0.

$$\text{ERR} = \frac{39 + 51}{209 + 39 + 51 + 1110} = 0.063 \quad (2.1)$$

$$\text{ERR} = \frac{31 + 39}{217 + 31 + 39 + 1122} = 0.049 \quad (2.2)$$

$$\text{ERR} = \frac{15 + 33}{233 + 15 + 33 + 1128} = 0.034 \quad (2.3)$$

The calculated error rate values indicate the small percentage of misclassification of proposed approach classifier which are acceptable since they are so small and far from reach 1.0.

### **Precision**

Precision is calculated by dividing the number of true positives (TP) on the total number of total true positives plus false positives (TP + FP).

$$\text{Precision(P)} = \frac{209}{209 + 51} = 0.803 \quad (3.1)$$

$$\text{Precision(P)} = \frac{217}{217 + 39} = 0.847 \quad (3.2)$$

$$\text{Precision(P)} = \frac{233}{233 + 33} = 0.875 \quad (3.3)$$

The results of the three experiments indicate a high precision especially when the number of true positives increases and the number of FP decreases. We could have an achieved a higher precision if the number of spam words and spam messages in the knowledge base is bigger.

### **Recall**

Recall also known as sensitivity is calculated by dividing the number of true positives (TP) by the total number of true positives plus false negative (TP + FN).

$$\text{Recall(R)} = \frac{209}{209 + 39} = 0.842 \quad (4.1)$$

$$\text{Recall(R)} = \frac{217}{217 + 31} = 0.875 \quad (4.2)$$

$$\text{Recall(R)} = \frac{233}{233 + 15} = 0.939 \quad (4.3)$$

The results of the three experiments indicate a high recall especially when the number of true positives increases and the number of FN decreases. We looking to decrease FN due to It is important measure for classify SMS spam as legitimate. We could have

an achieved a higher recall if the number of spam words and spam messages in the knowledge base is bigger.

### **F-measure**

$$F - \text{measure (F)} = 2 * \frac{0.803 * 0.842}{0.803 + 0.842} = 0.822 \quad (5.1)$$

$$F - \text{measure (F)} = 2 * \frac{0.847 * 0.875}{0.847 + 0.875} = 0.860 \quad (5.2)$$

$$F - \text{measure (F)} = 2 * \frac{0.875 * 0.939}{0.875 + 0.939} = 0.905 \quad (5.3)$$

F-measure the harmonic average of precision and recall, indicates how accurate a classifier is after calculate precision and recall, F-measure is favored over accuracy when we have an unbalanced dataset.

### **Matthews Correlation Coefficient (MCC)**

to determine the quality of binary (two-class) classification methods we calculate the MCC. MCC utilizing true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) values as given in section 2.15.

it is return a value between  $-1$  and  $+1$ . A coefficient of  $+1$  represents a perfect classification,  $0$  no better than random classification and  $-1$  indicates completely wrong binary classifier.

$$MCC = \frac{(209 * 1110) - (39 * 51)}{\sqrt{(209 + 51) * (209 + 39) * (1110 + 51) * (1110 + 39)}} = 0.784 \quad (6.1)$$

$$MCC = \frac{(217 * 1122) - (31 * 39)}{\sqrt{(217 + 39) * (217 + 31) * (1122 + 39) * (1122 + 31)}} = 0.831 \quad (6.2)$$

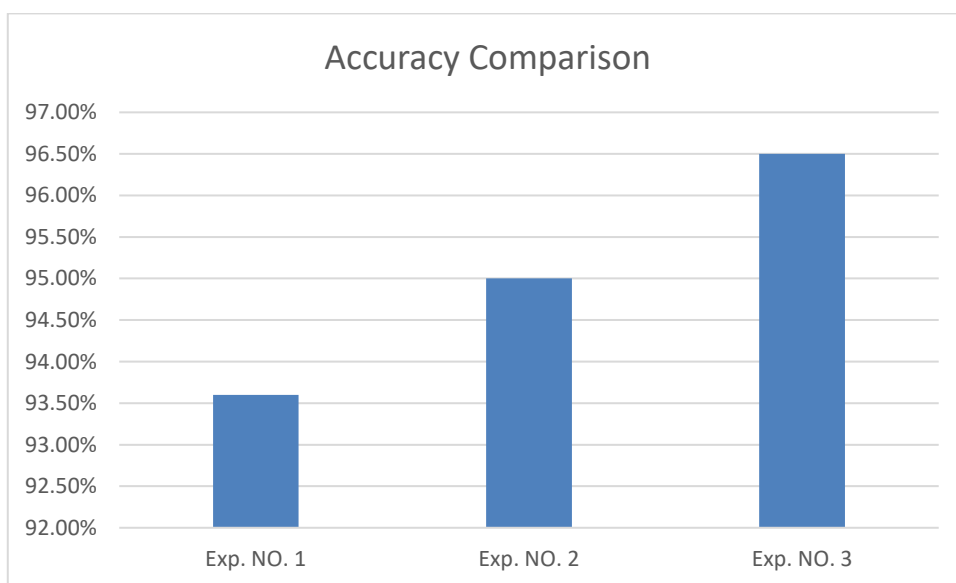
$$MCC = \frac{(233 * 1128) - (15 * 33)}{\sqrt{(233 + 33) * (233 + 15) * (1128 + 33) * (1128 + 15)}} = 0.886 \quad (6.3)$$

It is useful for unbalanced datasets to see overall performance of classifier, and we note that 0.886 which are acceptable since they are so far from reach -1.

After calculate all used measures we summarized the results in Table (6.3) for three experiments and Figure (6.1) shows the accuracy average of all the experiments.

**Table (6.3):** Experimental Results for different measures

Description	Exp. NO. 1	Exp. NO. 2	Exp. NO. 3
Accuracy	93.6%	95.0%	96.5%
Error rate	6.4%	4.9%	3.5%
Precision	80.3%	84.7%	87.5%
Recall	84.2%	87.5%	93.9%
F-measure	82.2%	86.0%	90.5%
MCC	78.4%	83.1%	88.6%



**Figure (6.1):** Accuracy comparison for three different expermint

The table shows differences in the results of evaluation measures of the classification in the three experiments. This is due to the following reasons:

- The increasing number of SMS spam messages in to ontology by extending and enriching the ontology with more spam words which can be used in the process of SMS classification.



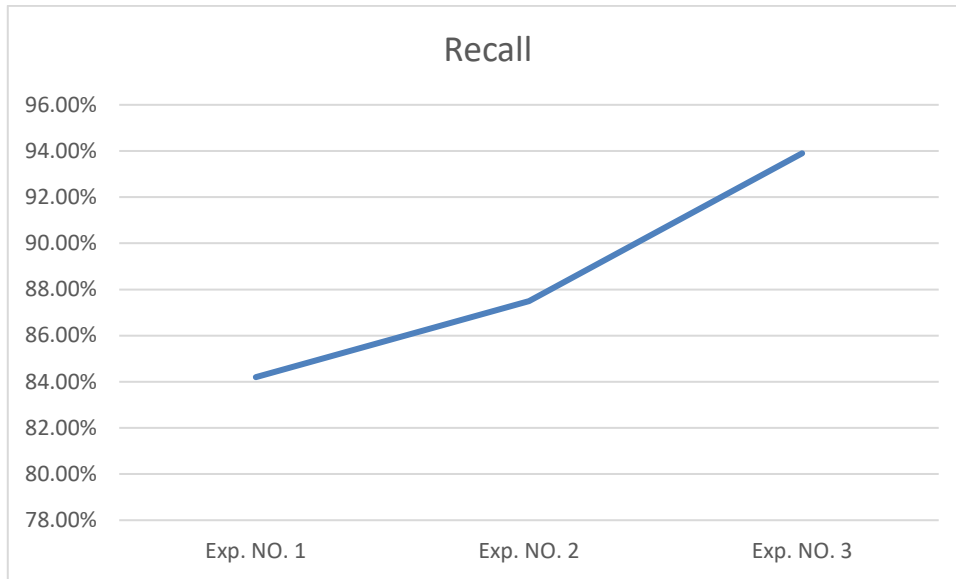
- The synonyms and meaning in WordNet may have contributed to the differences noticed. To be honest in some cases it may fail in providing equivalents for some domain specific words. For example, the word "مسيره" (march) has no equivalents in Arabic WordNet, so if the spammer writes another synonym word like "مظاهرة" (march) the system wouldn't be able to recognize that these words carry the same meaning as the word "equivalence" and therefore will not classify it as spam. To resolve this, we manually add new synonym for such words to the ontology.
- The 300 SMS spam messages for training maybe insufficient and therefore increasing the number may positively affect the results.
- Some words have high weight so the total weight of a message due to the weights of its constituent words will cause it to be classified as spam while, in fact, the message is legitimate leading to negatively affect the results. To resolve this, we make another phase of semantic rules which is depending on relation between this words as explained in section (5.8).

We can note the great difference in the results improvement as shown in Table (6.3) by adding new SMS spam to the ontology when and after applying our approach. For example, in the 100 spam messages case, the Accuracy is 93.6% and the F-measure is 82.2%, while in the 300 spam messages case and with WordNet semantic rules, the Accuracy is 96.5% and the F-measure is 90.5%. We can summarize accuracy results for all experiments with the highest accuracy result of 96.5%.

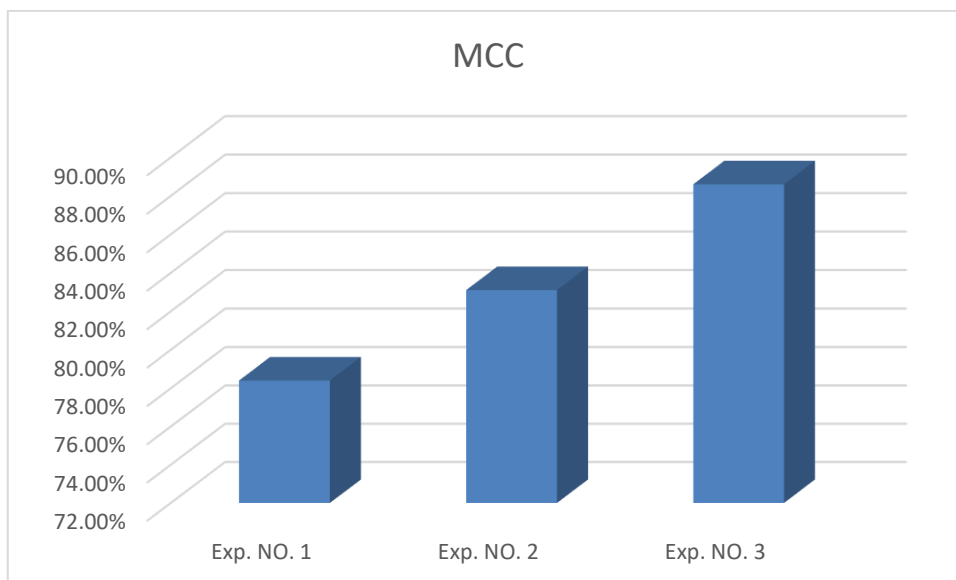
Another important measure is Recall (sensitivity) which is increased for all experiments which means that the approach decreased false negative (FN) result in the confusion matrix, this means decrease in spam messages classified as legitimate Figure (6.2).

In Precision, we get different results that mean for false positive (FP), which less important than FN because when our approach detects legitimate as spam, SMS provider can manually review them and pass them as legitimate.

The quality of classification detected by Matthews Correlation Coefficient (MCC) was achieved 88.6%, which is a high performance in indication Figure (6.3).



**Figure (6.2):** Recall rates comparison for three different experiments



**Figure (6.3):** MCC comparison for three different experiments

#### 6.4 Comparison with Other Works

We compared our results with results obtained from similar works in terms of the classification only. That is, we compare our ontology-based classification with traditional classification approaches. For this task we chose to compare with Naïve Bayes classifier (Shahi & Yadav, 2013). The Table (6.4) contains comparison of our results with Naïve Bayes with approximately the same number of spam and legitimate

SMS that have been used in training and testing stages. For that we choose RapidMiner tool.

We use the same dataset used for Arabic SMS spam ontology approach, then we used the same preprocessing such as tokenize and stop words removal.

**Table (6.4):** Comparison with the Naïve Bayes classifier

<b>Spam Type</b>	<b>ACC</b>	<b>Err</b>	<b>Precision</b>	<b>Recall</b>	<b>F-measure</b>
Naïve Bayes	96.4%	3.6%	90.0%	88.7%	89.42%
Proposed approach	96.5%	3.5%	87.5%	93.9%	90.5%

By comparing the results of our approach with Naïve Bayes classifier mentioned in Table (6.4), our approach gives better performance over Naïve Bayes classifier in terms of spam Recall and F-measure. This indicates that the use of ontology contributes effectively in the process of Arabic SMS spam classification. But in precision the Naïve Bayes better result which that due to the Naïve Bayes classifier training the legitimate SMS dataset also, so it should decrease false positive value (FP).

## **6.5 Summary**

In this chapter, we presented the experimental results of the SMS spam detection and classification approach. We evaluated the approach based on the different measures such as Accuracy, Error Rate Precision, Recall, F-measure and Matthews Correlation Coefficient (MCC). The classification was performed using Naïve Bayes method and the results were compared to those of our ontology based approach. They indicate that our approach outcomes the one based on Naïve Bayes method.

**Chapter 7**  
**Conclusions and Future**  
**Work**

## Chapter 7

### Conclusions and Future Work

In this thesis, we developed an ontology based approach for classify Arabic SMS messages as spam or legitimate. The main contribution of this thesis is the ontology and the related knowledge base together with the set of the semantic rules which can support SMS providers to detect SMS spam with higher rate of accuracy.

We have built a domain ontology for Arabic SMS spam consisting of spam vocabulary (words) and spam messages and is collected from a number of SMS spam corpus. The ontology developed with the assistance of SMS provider in Palestine as a domain expert. The ontology model has an important advantage of being extensible, i.e., open the possibility to adding future terms and relations related to SMS spam. Protégé OWL is used to build the ontology including its concepts (classes), properties, taxonomies, various restrictions, class instances and semantic rules. The ontology together with the various spam words and spam messages (instances) form the knowledge base of the SMS spam.

Since the domain of the ontology is related to the Arabic language in terms of the common spam words and the text of the SMS messages which includes spam and non-spam word, it is difficult to cover the whole domain in the hierarchy and the relations in the ontology. Therefore, we need to resort to other means to enrich the ontology. We used Arabic WordNet to enrich the ontology instances using word synonyms. This has contributed to achieve better results in classifying SMS messages. Additionally, we supported the process of message classification by a set of semantic rules based on the ontology reflecting the manual process of filtering as a necessary step to classify messages.

The system refers to the knowledge base to perform querying as well as reasoning needed in the decision as whether a given message is spam or legitimate. This is aided by a set of SWRL rules as well as spam word weights and relations between these words.

The overall approach, respectively a system prototype realizing the approach, consisted of several modules including SMS spam knowledge base consisting of the ontology and the spam instances, the synonym module used to relate ontology terms and instances their respective synonyms from Arabic WordNet, querying module used

to answer very specific queries (SPARQL) with the help of the reasoning module that would be difficult to look for directly in the knowledge base, spam detection (classifier) module used to receive SMS from the user through the user interface and decide if the SMS is spam or legitimate with the aid of the reasoning module, SMS message sending module used to send messages to mobile operators to deliver them in turn to the user handsets, user interface module used by users to send SMS requests to their clients and by administrators to manually add new spam words to the SMS spam knowledge base.

We performed a set of experiments to evaluate the proposed approach. Experimental results show an overall accuracy of 96.5% of the classification and an F-measure of 90.5%. Comparing these results to those of Naïve Bayes classifier indicates a better performance of the proposed approach over Naïve Bayes classifier. This indicates that the use of semantic-based classification contributes effectively in the process of Arabic SMS spam classification.

We tested the approach on a limited number of messages and that is why we got high accuracy. Increasing the number of messages may affect the accuracy and this needs further investigation.

An important factor affecting the performance of the approach is enriching the SMS spam knowledge base with synonyms. The Arabic WordNet used for this purpose is weak in terms of synonyms numbers and in terms of tagging facility. Therefore, enhancing it or replacing it with a better Arabic lexicon would give better spam filtering results.

Furthermore, we look forward to study the performance of the approach in terms of response time.

Finally, since only a prototype of the proposed approach is implemented, it is recommended to implement a complete system and API including other languages.

Improving the approach along the above aspects, encourages us to look forward to spread it as a tool for SMS spam detection and filtering available to local SMS providers.

# References

## References

- Abdoh, M., Musa, M., & Salman, N. (2009). Detecting Spam by Weighting Message Words. *Journal of Arts and Sciences*, 1.
- Akbari, F., & Sajedi, H. (2015). *SMS spam detection using selected text features and Boosting Classifiers*. In Information and Knowledge Technology (IKT), 2015 7th Conference on (pp. 1-5). IEEE.
- Anchal , A. S. (2014). SMS Spam Detection Using Neural Network Classifier. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(6).
- Balakumar, M., & Vaidehi, V. (2008). *Ontology based classification and categorization of email*. In Signal Processing, Communications and Networking, 2008. ICSCN'08. International Conference on (pp. 199-202). IEEE.
- Balubaid, M. A., & Manzoor, U. (2015). Ontology Based SMS Controller for Smart Phones. (*IJACSA International Journal of Advanced Computer Science and Applications*, 6(1).
- Blanzieri, E., & Bryl, A. (2008). A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, 29(1), 63-92.
- Boyce, S., & Pahl, C. (2007). Developing domain ontologies for course content. *Journal of Educational Technology & Society*, 10(3), 275-288.
- Cao, L., Nie, G., & Liu, P. (2011). *Ontology-based spam detection filtering system*. In 2011 International Conference on Business Management and Electronic Information.
- Delany, S. J., Buckley, M., & Greene, D. (2012). SMS spam filtering: methods and data. *Expert Systems with Applications*, 39(10), 9899-9908.
- Deng, W.-W., & Peng, H. (2006). *Research on a naive bayesian based short message filtering system*. In Machine learning and cybernetics, 2006 international conference on (pp. 1233-1237). IEEE.
- Dilekh, T., & Behloul, A. (2012). Implementation of a new hybrid method for stemming of Arabic text. *analysis*, 3(4), 5.
- Elkateb, S., Black, W., Vossen, P., Farwell, D., Rodríguez, H., Pease, A., & Alkhalifa, M. (2006). *Arabic WordNet and the challenges of Arabic*. In Proceedings of Arabic NLP/MT Conference, London, UKCiteseer.
- Federal Trade Commission. (2013). Text Message Spam. Retrieved from <https://www.consumer.ftc.gov/articles/0350-text-message-spam>
- Gómez Hidalgo, J. M., Bringas, G. C., Sáenz, E. P., & García, F. C. (2006). *Content based SMS spam filtering*. In Proceedings of the 2006 ACM symposium on Document engineering (pp. 107-114). ACM.
- GSMA. (2013). cloudmark-annual-threat-report-2013. Retrieved from <http://www.gsma.com/managementservices/cloudmark-annual-threat-report-2013/>
- GSMA. (2015). GSMA Intelligence market overview. Retrieved from <https://gsmaintelligence.com/markets/2797/dashboard/>
- GSMA. (2016). What is SMS spam? Retrieved from <http://www.gsma.com/managementservices/faq/what-is-sms-spam/>
- Harrington, M. (2008). Understanding SMS: Practitioner's Basics. *CFCE, EnCE*.



- Jain, V., & Prasad, S. (2014). Ontology based information retrieval model in semantic web: a review. *International Journal*, 4(8).
- Kalfoglou, Y. (2007). Using ontologies to support and critique decisions. *Engineering Intelligent Systems*, 15(3), 33-40.
- Karami, A., & Zhou, L. (2014). Improving static SMS spam detection by using new content-based features.
- Khemapatapan, C. (2010). *Thai-English spam SMS filtering*. In Communications (APCC), 2010 16th Asia-Pacific Conference on (pp. 226-230). IEEE.
- Kiamarzpour, F., Dianat, R., & Sadeghzadeh, M. (2013). Improving the methods of email classification based on words ontology. *International Journal of Computer Science Issues (IJCSI)*, 10(4), 262.
- Kim, J., Dou, D., Liu, H., & Kwak, D. (2007). Constructing a user preference ontology for anti-spam mail systems *Advances in Artificial Intelligence* (pp. 272-283): Springer.
- Kim, S.-E., Jo, J.-T., & Choi, S.-H. (2015). SMS Spam Filtering Using Keyword Frequency Ratio. *International Journal of Security and Its Applications*, 9(1), 329-336.
- Liu, J., Ke, H., & Zhang, G. (2010). *Real-time sms filtering system based on bm algorithm*. In Management and Service Science (MASS), 2010 International Conference on (pp. 1-3). IEEE.
- Mizoguchi, R., Vanwelkenhuysen, J., & Ikeda, M. (1995). Task ontology for reuse of problem solving knowledge. *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing*, 46-59.
- Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA.
- Ortiz, A., & Prieto, A. (2004). *SMS transmission using PDU mode and 7-bit coding scheme*. In ICWI (pp. 1147-1152).
- Pereira, V., & Sousa, T. (2004). Evolution of Mobile Communications: from 1G to 4G. *Department of Informatics Engineering of the University of Coimbra, Portugal*.
- Protégé. (2016). Protégé. Retrieved from <http://protege.stanford.edu/>
- stanford. (2016, 2016). DL Query. Retrieved from <http://protegewiki.stanford.edu/wiki/DLQueryTab>
- Sajja, P. S., & Akerkar, R. (2012). *Intelligent technologies for Web applications*: CRC Press.
- Shahi, T. B., & Yadav, A. (2013). Mobile SMS spam filtering for Nepali text using naïve bayesian and support vector machine. *International Journal of Intelligence Science*, 4(01), 24.
- Skudlark, A. (2015). *Characterizing SMS spam in a large cellular network via mining victim spam reports*. In 26th European Regional ITS Conference, Madrid 2015 International Telecommunications Society (ITS).
- Sugumaran, V., & Gulla, J. A. (2011). *Applied semantic web technologies*: CRC Press.
- Taufiq, M., Abdullah, M., Kang, K., & Choi, D. (2010). *A survey of preventing, blocking and filtering Short Message Services (SMS) spam*. In Proc. of

- International Conference on Computer and Electrical Engineering. IACSIT (pp. 462-466).
- Taufiq Nuruzzaman, M., Lee, C., Abdullah, M., & Choi, D. (2012). Simple SMS spam filtering on independent mobile phone. *Security and Communication Networks*, 5(10), 1209-1220.
- Taylor, D., & Pohl, J. G. (2009). Increasing the expressiveness of owl through procedural attachments. *Proceedings of InterSymp-2009: Baden-Baden, Germany*.
- Totewar, A. A., & Chatur, D. P. (2011). Enabling Semantics on Location Based Services Using Ontology Based Reasoning. *Journal of Computer Technology & Applications*, 2(2), 31-36.
- Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, 11(02), 93-136.
- Uysal, A. K., Gunal, S., Ergin, S., & Sora Gunal, E. (2012). The Impact of Feature Extraction and Selection on SMS Spam Filtering. *Elektronika ir Elektrotechnika*, 19(5), 67-72.
- Wang, X. H., Zhang, D. Q., Gu, T., & Pung, H. K. (2004, 14-17 March 2004). *Ontology based context modeling and reasoning using OWL*. In Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on (pp. 18-22).
- Yoon, J. W., Kim, H., & Huh, J. H. (2010). Hybrid spam filtering for mobile communication. *computers & security*, 29(4), 446-459.
- Youn, S. (2014). SPONGY (SPam ONtology): Email Classification Using Two-Level Dynamic Ontology. *The Scientific World Journal*, 2014.
- Zhang, H.-y., & Wang, W. (2009). *Application of Bayesian Method to Spam SMS Filtering*. In 2009 International Conference on Information Engineering and Computer Science (pp. 1-3).