A proposed Framework for
Aspect-Based Sentiment Analysis for Arabic Content using Semantic-Based Approach

By
Ahmed Yousri Salama

A Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Information Technology
College of Computer Science and Information Technology
Sudan University of Science and Technology

Supervisor
Dr. Nisreen Beshir Osman
August 2014
ACKNOWLEDGMENT

In the name of Allah the most gracious and most merciful

I wish to thank my fellow graduate students at SUSTECH University; it was a friendly and stimulating environment where I had the chance to learn about language and Semantic. Special thanks to my friends Abdallah Emad and Nizar Almoonah, for always having their doors open. In terms of the content and quality of this work, the biggest thanks must go to my supervisor, Dr. Nisreen Bashir for gently guiding me each step of the way. Finally, I want to thank my family, especially my parents.
ABSTRACT

This research introduces a combined approach and framework for Arabic contents to analyze the Arabic opinions called sentiment analysis using semantic approach. This thesis focusing on the semantic model, which is centered on deriving the semantic orientation of individual words and expressions, and machine learning classifiers, using ontology engineering, text tagging and part of speech tagging. To show the potential long-term advantages of this approach, the creation of the framework process had been described, designed and developed from five components, first one is Arabic Part Of Speech “POS” tagger to assign the correct tag for every word in the opinion, second components is Arabic Ontology Classifier to query RDF Ontology using ontology engineering SPARQL language to extract the main concepts, the third component is Arabic Sentiment Lexicon which act like Arabic dictionary and the final component is the Counter and Report component which do the calculation and final results display, All previous components worked together as one solid framework called SUSTASA.

SUSTASA framework had been designed and developed using advance integration technologies such as Stanford POS tagger and Stanford Ontology protégé. SPARQL, RDF had been used to design the Ontology. The framework tested against pilot movie Arabic comments and Arabic movie ontology, and the results from framework was able to detect and classify a pilot comments written in Arabic language and measure them against the rating five star and .

The results from framework analysis on movie reviews which had been chosen for test, the results also agree well with the calculated stars which entered by the opinion holders comments values.
المستخلص

تحليل المشاعر هو عملية يتم فيها استخدام علم اللغة الحاسوبي والتحليل النصي من أجل تحديد نوع النص والكشف عما يحمله من مشاعر سواء إيجابية أو سلبية أو محايدة تجاه موضوع النص. تكمن أهمية تحليل المشاعر من خلال استنباط التوجه العام للكاتب.

في هذا البحث تم إفتتاح منهجية جديدة لقياس الآراء وتعابيرها ومحاولة استنباط نوع التعبير هل هو إيجابي أم سلبي أم محايد، و تم استخدام عدة تقنيات ومناهج لتكوين إطار عمل لتحليل النص العربي إلى تحليل فني عن طريق دمج تقنية تحليل النص العربي وانتلوجيا الويب الدالالي للغة العربية.

تم اختيار مجال الافلام كمثال تطبيقي، وتم عمل انتطاعيات الافلام ومن ثم تجزئة النص العربي واستخرج الصفات والموصوفين عن طريق منهجية مفصلة وفي النهاية يظهر التحليل ما إذا كان التعليق إيجابياً أو سلبياً أو محايداً.

تم قياس عدة تعلقات لأفلام معروضة وجاءت النتائج إيجابية ومتوافقه مع التقييم المصاحب لكل تعليق من خمسة نجوم.
List of Content

Abstract .......................................................................................................................... I
Acknowledgment ........................................................................................................ III
List of Figure ................................................................................................................ IV
List of Tables .................................................................................................................. VI
List of Abbreviation ...................................................................................................... VII
1. Introduction
   1.1 Background ........................................................................................................... 1
   1.2 Problem Statement .............................................................................................. 2
   1.3 Aims and Objectives ............................................................................................ 2
   1.4 Organization of the Thesis ................................................................................... 2
2. Literature Review
   2.1 Background .......................................................................................................... 3
   2.2 Machine Learning based approach ....................................................................... 4
   2.3 Lexically based approach ..................................................................................... 4
3. Proposed Approach and Framework
   3.1 SUSTASA framework ......................................................................................... 7
   3.2 Data Collection ..................................................................................................... 7
   3.2.1 POS Arabic tagger Component ......................................................................... 8
   3.2.2 Ontology Classifier Component ...................................................................... 9
   3.3.3 Arabic Sentiment Lexicon Component ............................................................ 10
   3.4.4 Counter Component ........................................................................................ 11
   3.5.5 Report Component .......................................................................................... 11
4. Results and Discussions
   4.1 Arabic Nouns and Adjectives Extraction ............................................................... 12
   4.2 Arabic Movie Ontology ....................................................................................... 12
   4.3 Classification ......................................................................................................... 13
   4.3 Conclusion ............................................................................................................ 14
5. Conclusion and Future Works
List of Figures

<table>
<thead>
<tr>
<th>Figure No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3-1</td>
<td>8</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3-1</td>
<td>7</td>
</tr>
<tr>
<td>Table 3-2</td>
<td>8</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>SPARQL</td>
<td>SPARQL Protocol and RDF Query Language</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
<tr>
<td>SUST</td>
<td>Sudan University of Science and Technology</td>
</tr>
<tr>
<td>SUSTASA</td>
<td>Sudan University of Science and Technology</td>
</tr>
<tr>
<td></td>
<td>Arabic Sentiment Analysis</td>
</tr>
<tr>
<td>SA</td>
<td>Sentiment Analysis</td>
</tr>
<tr>
<td>SO</td>
<td>Sentiment Orientation</td>
</tr>
<tr>
<td>JSf</td>
<td>Java Server Faces</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
</tbody>
</table>
Chapter One

Introduction
1.1 Background

An important part of our information-gathering behavior has always been to discover what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people can, and do, actively use information technologies to seek out and understand and analyze the opinions of others.

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer’s feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents [1]. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today [2]. The Web is a big store of structured and unstructured data and information. The analysis of this information to extract latent public opinion and sentiment is a challenging task.

Sentiment analysis starts with a question: “What other people think?”, and finally convert into a commercial business. After the great success of Web-2.0, sentiment analysis became a demanding and commercially supported research field.

The semantic field theory is based on an analytical approach, which considers the meaning of word within given views of the world [3]. Ontologies define the terms and relationships utilized to describe and specify an area of concern. Ontologies are used to classify the terms that can be employed in a particular application, defining possible relationships, and determine possible constraints on using those terms [4].

Although Arabic is considered as one of the top 10 languages mostly used on the Internet based on the ranking carried out by the Internet World State rank in 2010 and it is spoken by hundreds of millions of people [5], there exist limited annotated resources for sentiment analysis such as labeled corpora, and polarity lexica. This could be considered the main reason which has motivated the generation of an opinion corpus for Arabic in this work.
1.2 **Problem Statement**

Identify polarity in Arabic user generated opinions according to features described in specific domain ontology.

Given an Arabic movie review containing multiple features and varied opinions, the objective is to extract expressions of opinion describing a target feature and classify it as positive, negative or natural.

1.3 **Aims and Objectives**

The aim of this study is to develop a method to identify polarity of a movie features in Arabic contents using semantic Ontologies.

The objective of this study is to propose and design and develop a framework to identify polarity in Arabic user generated opinions according to features described in specific domain ontology on movie by use RDF to design the ontology and Sanford NLP POS tagger libraries.

1.4 **Thesis Contribution**

The designed framework is able to identify Arabic user generated opinions by designing for first time pilot Arabic movie ontology using SPARQL (pronounced "sparkle", a recursive acronym for SPARQL Protocol and RDF Query Language) is an RDF query language and using Arabic Part Of speech Tag.

The designed framework was used Stanford API for Part of Speech tagging for Arabic language in Arabic text analysis for first time which successfully was showed the ability to extract Arabic Nouns(NN) and Arabic Adjectives(JJ) for Arabic rich text corpus like movies review comments as example.

Although The domain of movies reviews had been chosen because of simplicity and availability of standard movie ontology in English and in future work the Arabic ontology could be enhanced with existing English version, but this framework can be applied for any domain like political election, crisis expectation, public opinion measurement depends on the designed ontology. Sample of Arabic Opinions Taxonomy had been designed which useful in determining the negative filtered users comments from positive reviews also for first time in Arabic content.
All pervious contributions was combined into one framework and successfully implemented and programmed successfully using Java and JSf, RDF, SPARQL and Stanford Tagger API. One of the main contribution could be considered in this thesis is that, it was came across three huge research area which are: Sentiment Analysis, Semantic Web, Part Of Speech Tagging and Text Analysis. This is multidisciplinary research.

Also this thesis contributed to Arabic Ontology international project un Birziet university which was started in 2005 and funded by UN to support Arabic Ontology. [6]
This thesis contributed to Nubian Language which is used in north Sudan with Arabic by developing and previous work done and published in international conference [7].

1.5 Organization of the Thesis
This thesis contains six chapters; chapter one is the introduction—Chapter two is the literature review and Chapter three describes the framework. Chapter four explains results and discussions. Chapter Five contains the conclusion and suggestions for future work.
Chapter Two
Literature Review
2.1 Background

Current textual information could be classified to two major categories: facts or opinions. Facts are defined as events or their properties that have actually occurred; they are objectives expressions that can be verified and proven. Opinions are subjective beliefs, and are the results of emotion or interpretations of facts. While an opinion may be supported by an argument, counter opinions can often be drawn from the same set of facts the concept of opinion is very broad. In this section we focus on previous research in the field of subjectivity and sentiment analysis and on semantic web especially when using Arabic contents. First we will make an introduction to content analysis and semantic web. Later, we discuss worked that has been done specifically pertaining to Arabic contents with a focus on three approaches: Machine Learning approach (ML) and semantic approach (SA) [8] [9] [10].

As an area of research sentiment analysis can be considered a part of computation linguistics. Natural Language processing NLP, and text mining.

Generally opinion analysis aims to show the author’s attitude towards specific subject or the overall contextual polarity of a text. This may be a judgment or an evaluation, an affective state or an intentional emotional communication [1].

The Semantic Web is the extension of the World Wide Web that enables people to share content beyond the boundaries of applications.
The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is a source to retrieve information from the web (using the web spiders from RDF files) and access the data through Semantic Web Agents or Semantic Web Services [11].

2.2 Machine Learning Approach

The problem is formulated as such: given an opinion-oriented document or text, whose overall opinion applies only to one item or idea, categorize its attitude by assigning it to one of the two contrasting sentiment polarities or locate it on a scale between these two polarities (Pang and lee, 2008) [1].

This classification task of tagging a text as containing either a positive or a negative view is called sentiment polarity classification of polarity classification.

Such documents have will-defined topics and in addition to the sentiment expressed in the language they also contain the author’s rating this label gives a quantitative indication of the author’s opinion.

Such documents are usually used as a best case example when measuring sentiment (Chevalier and Myzlin 2006) [12].

Pang et al. (2002) [1] employ three machine learning methods to classify movie reviews into two classes: positive and negative. Neutral reviews are omitted in the study, which made the problem less complicated.

### 2.3 Lexically based approaches

Significant success has been achieved in sentiment analysis using lexically based approaches, often targeting adjectives and adverbs.

Much work has been done to automatically acquire opinion word resources (Hatzivassiloglou and McKeown 1997 [14]; Baroni and Vegnaduzzo 2004; Turney and Littman, 2003 [15]). Turney and Littman’s point wise mutual information method has been applied successfully in a number of sentiment analysis applications, including opinion retrieval from corpora of email and blogs and affect classification in short story fiction.

Other lexically based approaches to sentiment analysis have used lexicons manually crafted specifically for sentiment or affect analysis (Durbin et al 2003; [10]). Cesario et al. (2006, 2007) [8] semi-automatically developed an opinion-expressing word bank by applying a flexible word scoring function based on document-level human annotations. The annotations included harshness scores that reflect the degree to which documents were considered positive or negative.

In their OASYS opinion analysis system, the word bank scores feed into several document scoring functions that both qualitatively assess opinion with respect to specific topics, as well as generate a quantitative measure of that opinion.
Unlike many other research systems, OASYS runs continuously and has been applied to a growing inventory of millions of documents. Subasic and Heuttner’s (2000) [16] affect analysis system relies on a fully manually developed word list and distinguishes affect as a set of emotional categories that does not address opinion types directly. Subasic and Huettner describe an affect word lexicon of about 4000 terms, of various parts of speech.

Each word is manually annotated with an affect category and a centrality and intensity value with respect to that category. The lexicon was shown to be useful for determining the overall affect content of documents, though the affect categories are numerous and sometimes quite subtle in their distinctions (e.g. repulsion, aversion). Their affect measurement system provides only gross overall scores for each affect category, and does not provide information about specific entities or events to which the affect is attributed, in contrast to the OASYS system, which assesses opinion directly with respect to specific topics.
Chapter Three
Research Framework
3.1 The proposed framework

The design framework, framework consists of five components (Table 3-1), three of them are core activities and two are programming (code) helpers.

The core components of the SUST Arabic Semantic Analysis (SUSTASA) framework are:

1. “Part of Speech “POS extractor
2. Ontology classifier
3. Lexicon classifier

Table 3-1: SUSTASA Framework Components Summery

<table>
<thead>
<tr>
<th>Component</th>
<th>Task</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Extractor</td>
<td>Extract Features(Nouns + Adjectives)</td>
<td>Row Data - Corpora</td>
<td>Nouns(Features) + Adjectives</td>
</tr>
<tr>
<td>2 Ontology Classifier</td>
<td>Classify if Feature is a main Concept in the Domain of Sentiment</td>
<td>Nouns(Features)</td>
<td>Nouns or Null (not exist as main concept in the domain)</td>
</tr>
<tr>
<td>3 Lexicon Classifier</td>
<td>Classify if the feature Negative or Positive or Natural using Manual Lexicon contain 6 words</td>
<td>Adjective</td>
<td>Positive – Negative – Natural</td>
</tr>
</tbody>
</table>
3.2 Data Collection

The current investigation involved sampling and analyzing ten comments to measure their sentiment orientation using the designed Arabic Ontology and Arabic Lexicon.

A total of 10 comments was collected from array of objects from MySQL comments database and then analyzed for the designed framework.

Samples were entered and rated to evaluate the SUSTASA framework against entered data by going through framework activities or components as described in next section.

3.2.1 POS extractor

The first component is Part Of Speech POS extractor which has important role in cleaning the data set by separating the nouns (opinion holder) and the Adjectives (the sentiment) from given sentence, using Part Of Speech tagging approach.

Part-of-Speech tagging is assigning the correct part of speech (noun, verb, etc.) to words [17]. Stanford NLP SW [18] was used to extract the Arabic nouns and Adjectives, the nouns take the tag (NN) and the Adjectives take the tag (JJ) see Table 3-2.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Coordinating conjunction</td>
</tr>
<tr>
<td>CD</td>
<td>Cardinal number</td>
</tr>
<tr>
<td>DT</td>
<td>Determiner</td>
</tr>
<tr>
<td>EX</td>
<td>Existential <em>there</em></td>
</tr>
<tr>
<td>FW</td>
<td>Foreign word</td>
</tr>
<tr>
<td>IN</td>
<td>Preposition or subordinating conjunction</td>
</tr>
<tr>
<td>JJ</td>
<td>Adjective</td>
</tr>
<tr>
<td>JJR</td>
<td>Adjective, comparative</td>
</tr>
<tr>
<td>JJS</td>
<td>Adjective, superlative</td>
</tr>
<tr>
<td>LS</td>
<td>List item marker</td>
</tr>
<tr>
<td>MD</td>
<td>Modal</td>
</tr>
<tr>
<td>NN</td>
<td>Noun, singular or mass</td>
</tr>
</tbody>
</table>

Table 3-2: Alphabetical list of part-of-speech tags
3.2.2 Ontology Classifier

Ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them [3] [9]. The reasons to use Ontology are:

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

Sharing common understanding of the structure of information among people or software agents is one of the more common goals in developing ontologism [9] [3]. For example, suppose several different web sites contain medical information or provide medical e-commerce services. If these Web sites publish the same ontology of the terms they all use, then computer agents can extract and aggregate information from these different sites. The agents can use this aggregated information to answer user queries or as input data to other applications.

In SUSTASA framework, for example: the Actor, “الممثل” , “السيناريو” is the main concepts or words that could be included and which expected to appear in almost all the texts talked about movies. Stanford protégé [4] was used to design the Ontology see screenshot (Figure 3-1)
The Ontology Classifier will separate the unrelated nouns, concepts from the important concepts using SPARQL [19] querying the nouns. The extracted adjectives will be tested against lexicon of sentiments to return the polarity of the adjective: Positive or Negative or Natural.

Figure 3-1: Movie Ontology Entities

3.2.3 Arabic Sentiment Lexicon

This component developed to query specific Arabic adjective and then get the polarity (Negative VS Positive).

There is well known Sentiment Lexicon for English language like the one done by Bing Liu [1] who maintains and freely distributes a sentiment lexicon consisting of lists of strings.
- Positive words: 2006 [1]
- Negative words: 4783 [1]

Useful properties: includes misspellings, morphological variants, slang, and social-media mark-up [1].

Arabic is much hard to crack than other languages. The biggest problem that one faces when dealing with Arabic text is the various forms a root word can take on given context, tense, and surrounding words. In the linguistics world, this is referred to as morphology.

Quoting the University of Columbia’s Owen Rambow, “Arabic morphology is usually considered more complex than English morphology because it involves not only prefixes and suffixes (such as -ed or -ing in English), but also changes to the lexical root itself (perhaps somewhat like English sing, sang, sung, song, but more common).”

The final output from these lexicon components was the polarity of the adjectives. For example: “good” adjective is normally has positive polarity in the lexicon which in Arabic: ﺗﺠﯿّﺪ

Final result of polarity identification from the lexicon component will be passed to the Counter component.

3.2.4 Counter Component

This component was designed to do the final sentiment calculations.

3.2.5 Report and Integration Component

This component was designed to do the final reporting and integration task for future needs calculations.
Chapter Four

Results and Discussion
4-1 Results

The current investigation involved sampling and analyzing ten comments to measure their sentiment orientation using the designed Arabic Ontology and Arabic Lexicon.

A total of 10 comments was collected from array of objects from MySQL comments database and then analyzed for the designed framework.

Samples were entered and rated to evaluate the SUSTASA framework against entered data by going through framework activities or components as described in next section.

The results from framework analysis agree well with the calculated stars which entered by the opinion holders comments values.

4-1-1 Arabic Nouns and Adjectives Extraction

According to Figure 4-1 POS extractor which was developed using Stanford NLP POS tagger [1] for Arabic comments, However Stanford NLP POS tagger had been used to extract the Nouns and Adjectives, which is shorthanded as: NN and JJ respectively.

The designed component extracted movie comments from array and successfully converted it using Stanford NLP algorithm and Arabic-training tagger to get and the Arabic Adjectives (the sentiment)and Arabic nouns (opinion holder).

4-1-2 Arabic Movie Ontology

Data obtained in previous component ‘Arabic Nouns and Adjectives Extraction’ passed to the ontology classifier. The ontology had been developed to classify and determine if the feature is a main concept in the domain of sentiment, using Apache Jena and Stanford Protégé.

Jena is an open source Semantic Web framework for Java. It provides an API to extract data from and write to RDF graphs. The graphs are represented as an abstract "model". A model can be sourced with data from files, databases, URLs or a combination of these. A Model can also be queried through SPARQL [20].

16
SPARQL is a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format. It was made a standard by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium, and is recognized as one of the key technologies of the semantic web. On 15 January 2008, SPARQL 1.0 became an official W3C Recommendation and SPARQL 1.1 in March, 2013 [21] [19].

This component takes the Nouns NN from the (POS extractor) to check if the specific Noun is in the main concepts map within the ontology or not.

The final goal of this ontology classifier is to determine if the specific comment should be considered to next steps or not.

4-1-3 Classification

The Third component is Lexicon classifier which had been developed to classify if the feature sentiment is Positive, Negative or Natural, by using Arabic Dictionary based approach

One of the simple techniques in this approach is Dictionary based which based technique using a small set of seed opinion words and an online dictionary, e.g., WordNet [22]. The strategy is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet for their synonyms and antonyms.

Opinion words are also known as polar words, opinion-bearing words [23] and sentiment words. Positive opinion words are used to express desired states while negative opinion words are used to express undesired states. Examples of positive opinion words are: beautiful, wonderful, good, and amazing (جيد، رائع ممتاز، جميل) . Examples of negative opinion words are bad, poor, and terrible (سيء، ضعيف، قبيح). Apart from individual words, there are also opinion phrases and idioms [1], e.g., cost someone an arm and a leg. Collectively, they are called the opinion lexicon. We used Dictionary based approach to generate Arabic Lexicon.

Using all previous components and approach from many research field in Arabic has result in promising pilot system that can analyze Arabic comments, texts and get it’s polarity or sentiments using previous mentioned tools such as Eclipse IDE, Stanford Arabic Tagger JAR library, Apache Jena JAR libraries and Stanford Protégé IDE for designing the ontology.
The result was promising because with comparing it with Review Rating system it showed the developed SUSTASA framework can give result almost identical to the one given by Rating System see Figure 4-3.

Furthermore the pilot system can exclude the Arabic “dump” comments from an array of comments and remove them from calculation.

This filtering done using the Ontology which it’s role is to determine the and filter the nouns if the nouns is not with the ontology domain or concepts it will be filtered out and will not be considered. See Figure 4-4.

For instance the review with user “Ahmed” said that: "أفضل فلم" will go through many steps as described below:

- **Separation Step**
  
The comment will be entered and extracted from array of opinions and using Java Tokenizer library [18] will cut the whole comment to tokens or words see Figures(4-1,4-2,4-3,4-5,4-6)

- **The POS tagger**
  
  After separation process, The POS tagger will take the comment tokens and by use trained Stanford library it will determine if this word (token) is Noun or Adjective and separate it. See Figure (4-6).

- **Ontology classifier**
  
  By using Ontology the system will exclude any token which is not in the main concepts or domain See Figure (4-7).

- **Lexicon classifier**
  
  The Arabic Lexicon classifier will classify if the extracted adjective “أفضل” is “Positive”, “Negative” or “Natural” word based on Lexicon table

- **Finally calculation**
  
  Finally calculation was done here to collect all “Positive”, “Negative” and “Natural” result from the array of opinions see Figures (4-7,4-8,4-9).
4-2 Discussions

It could be observed that Part-of-speech tagging was able to assign the correct label (noun, verb, etc.) to words using the POS approach see Figure 4-2.

It is apparent that using trusted training could increase the precession of the POS Arabic tagger.
Figure 4-1: Dump Objects to Test the Framework

```java
Openion[] ops = new Openion[3];
Openion op = new Openion();
op.setComment("test1");
op.setId(1);
op.setOwner("ahmed");
op.setStar_numbers(5);
op.setTagg(""); 

Openion op1 = new Openion();
op1.setComment("test2");
op1.setId(1);
op1.setOwner("ahmed");
op1.setStar_numbers(2);
op1.setTagg(""); 

Openion rejected = new Openion();
rejected.setComment("test3");
rejected.setId(1);
rejected.setOwner("ahmed");
rejected.setStar_numbers(0);
rejected.setTagg(""); 

//2 second extract main noun and the main adjective by query ontology
ops[0] = op;
ops[1] = op1;
ops[2] = rejected;
//get all openion form DB
return ops;
```
falseNot in Doamin even one word
number_of_positive_comments: 1
number_of_negative_comments: 1
number_of_positive_stars: 1
number_of_negative_stars: 1
number_of_Naturla_comment: 1

Figure 4-2: Command Line Results
Figure 4-3: One Movie Show
Figure 4-4: One Movie Rating
Figure 4-5: Save Button
Figure 4-6: Many Movie Revies and Rating
Figure 4-7: Sample 1 Bar Chart For Report Of Movie Review

Figure 4-8: Sample 2 Bar Chart For Report Of Movie Review
Figure 4-9: Sample 3 Pie Chart For Report Of Movie Review
Chapter Five
Conclusion and Future Work
5-1 Conclusion

Sentiment analysis using semantic web ontology is able to measure the opinion orientation, the thesis provides a framework build on well known previous research on part of speech tagging and ontology engineering to classify user generated comment for specific domain.

SUSTASA framework had been developed starting from specifying the domain needed to be mapped into ontology RDF and query the RDF using SPARQL, and then by extract nouns and adjectives the framework can output the right label or tag for specific sentence as positive or negative or natural the designed framework successfully done following tasks:

- Arabic Ontology for Arabic Opinions had been designed using protégé.
- Arabic Features and opinion has been extracted using Stanford NLP group library.
- Extracted Arabic opinion features classified as positive or negative or natural depends on Arabic Opinions Ontology.

5-2 Future Work

Further works could be done as extension to this work, such as:

- Using Arabic ontology for more domain in politics, e-commerce and marketing
- Using this framework with other local dialect like Sudanese, Nubian and others, it’s important to consider applying this framework for Nubian language since there is previous work on this language published.
6. References


J. A. a. M. D. Chevalier, "The Effect of Word of Mouth on Sales: Online Book


