DOMAIN ORIENTED SYNTAX BASED ASPECT DETECION FOR STUDENT FEEDBACK SYSTEM OF UCS Taungoo

NILAR SOE

M.C.Sc. **JUNE 2022**

DOMAIN ORIENTED SYNTAX BASED ASPECT DETECION FOR STUDENT FEEDBACK SYSTEM OF UCSTaungoo

By

Nilar Soe
B.C.Sc.

A dissertation submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science (M.C.Sc.)

University of Computer Studies, Yangon
June 2022

ACKNOWLEDGEMENTS

First and Foremost, I would like to express my gratitude and my sincere thanks to **Prof. Dr. Mie Mie Khin,** Rector, the University of Computer Studies, Yangon and **Prof. Dr. Ei Ei Hlaing**, Rector, the University of Computer Studies (Taungoo), for allowing me to develop this thesis and giving me general guidance and workable environment during the period of study.

My sincere thanks and regards go to Dr. Si Si Mar Win and Dr. Tinzar Thaw, Professor, Faculty of Computer Science, University of Computer Studies, Yangon, for their kind management throughout the completion of this thesis.

My heartfelt thanks and respect go to my supervisor, **Dr. Paing Thwe Soe,** Professor, Head of Department of Information Technology Supporting and Maintenance, the University of Computer Studies (Taungoo), for her invaluable recommendations regarding the thesis topic, giving me detailed guidance throughout the work of this thesis and invaluable guidance and support throughout the development of the thesis.

I would like to express my respectful gratitude to **Daw Aye Aye Khine** , Associate Professor and Head of the Department, the Department of English, University of Computer Studies, Yangon, and **Daw Chaw Ei Su**, Associate Professor and Head of the Department, the Department of Language, University of Computer Studies (Taungoo), for editing my thesis from the language point of view.

I would like to acknowledge my thanks to my teachers of the University of Computer Studies, Yangon and the University of Computer Studies (Taungoo), and all of my dear teachers from childhood to the present time.

In addition, I would like to thank the board of examiners for making precious comments and detailed corrections to my thesis and those who are pressing power to improve the end result. Last but not least, I especially thank my parents and my family for raising me and inspiring me all the time. Finally, I am grateful to my colleagues and all my friends for their cooperation and help.

STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

---------- ----------

Date Nilar Soe

ABSTRACT

Opinion Mining becomes popular in seeking the information on online review or feedback system. This technique can be usually used in recommender system that supposes the customers for making trust upon the products based on other user's opinion. Moreover, this technique can also help the development or maintenance of different kinds of products or activities by evaluating the users' opinion. In conventional opinion mining techniques, it can examine the people feeling from their reviews or comments such as positive or negative only. The process of examining such positive and negative score is also known as sentiment analysis. Sentiment analysis can be applied at different levels of scope such as *sentence level*, *document level* and *aspect level*. So, in the current trend, the goal of sentiment analysis is to dig the aspect word that is the fine grained sentiment information based on the reviews or comments of various domains. So, the proposed system aims to analyze the aspect level sentiment analysis on student feedback system.

The required feedback data are collected from the University of Computer Studies, Taungoo(UCST). First step of sentiment analysis is *part-of-speech* tagging (POS tagging) that can identify the form of each word in the sentence. For POS tagging, this system uses *OpenNLP* parser which parses the sentence as adjectives, verbs and nouns, respectively. For defining the sentiment score of each word, this system uses *sentiWordNet* lexical resources by applying the SWN3 algorithm which finds the score of each word in lexical resource and attaching with this word. In order to dig the aspect word for feedback statement, the Domain Specific Ontology relating to UCST is created in the preprocessing stage of this system which composed with the main aspect words of the domain. Finally, the proposed algorithm *Onto-to-List* can definitely find the matching aspect word from the feedback statement by confirming the domain specific ontology. This system is evaluated by using confusion matrix and the accuracy measurement based on the prediction of user's opinion. The accuracy of this system is 94% that is evaluated over 100 history records of this system. This system will assist the administrator of UCST to evaluate the performance of the University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1

INTRODUCTION

Most of the business problems change their environments to web application technologies. The customers or users also change their nature as sending the text to say about the products or somethings over the World Wide Web. For this reason, many of the information about the product or something already collected by web technologies of corresponding application. In many cases, which is needed to process to become the useful knowledge because the texts directly sent by the users are natural or unstructured language and they are not valued. To solve this problem, one of the data mining techniques, also called opinion mining is emerged to obtain the valuable information.

Opinion mining also known as Sentiment analysis is widely used in natural language processing and text analysis to identify the opinion or extract the specific things from the information source [1]. This type of analysis is applied to reviews on comments of social media or a variety of applications, such as marketing, customer service and product evaluation. The goal of this analysis is to determine the emotional state of the users or customers with respect to some statement or overall of a document. This emotion state becomes his or her judgment or evaluation of their intended products or something. The opinion mining can be used at the following scopes [2]:

- Document Level: It defines the scope for all statements of the whole document.
- Sentence-Level: It defines the scope for a single statement.
- Sub-sentence Level: It defines the scope for each sub-expression of a statement.

Upon each of these scopes, it needs to apply the sentiment analysis to dig the emotion of the users about the products or things of business.

1.1 Introduction to Aspect detection

The normal sentiment analysis determines the users' opinion by reviewing the overall statement or comment. For that, the admin cannot easily see that the users give the good or bad opinion upon what kind of products or things. To overcome this problem, the application developer creates the list of products or things to show the users and then the

user's focus or select their desired products or things to give the feedback statements about these products or things. When there are so many kinds of products or things which are needed to know the user's or customer's responses, it is not possible to show the icon lists of all products or things. Here, if the analysis can be made on the users' comments or feedback statements to dig the important items of the domain business from these statements, it does not need to show all products' icons list. So, sentiment analysis could be categorized as aspect based sentiment analysis to fulfill this requirement.

The main goal of sentiment analysis is to predict the opinion of users about the products or things such as their good or bad mood upon these products. Sentiment analysis can be done by the operations such as extracting opinion data, finding the polarity and defining for what subject matter and correlating with opinion holder [3]. There are four main types of sentiment analysis:

- **Fine-grained sentiment analysis**
    - This analysis firstly parses the incoming sentence into phrases and clause. And then it finds that who talk about what exact things of product or somethings in users' feedback comments. This analysis takes more cost-intensive compared with other types.
- **Emotion detection**
    - This analysis aims to detect and recognize types of feelings through the expression of texts, such as anger, sadness, surprise and so on. Emotion detection may have useful applications like that Gauging how happy our citizens are, Pervasive computing, to serve the individual better and Understanding the consumer.
- **Aspect-based sentiment analysis**
    - This analysis helps in the improvement of the business by knowing the specific features in their product which they need to improve according to customer's feedback to make their product a best seller.
- **Intent analysis**
    - The analysis helps you identify the intent of the consumer – whether the customer intends to purchase or is just browsing around. If the customer is

willing to purchase, you can track them and target them with advertisements.

(In order to improve) the performance of university by analyzing all features in the campus, aspect- based sentiment analysis is the most suitable.

1.2 Related Works

Aspect -based sentiment analysis has two main processing phases: aspect detection and opinion mining. There are many research relating with aspect detection process and opinion mining techniques. The researchers use their own strategies to make the aspect-based sentiment analysis. The student feedbacks system gathers the useful methods from the following related works and can make new strategy to dig aspect word from the sentence and to guess opinion of user who submitted this statement to the system.

K.M.Sam and C.R.Chatwin, explained about the lexical based sentiment analysis in the paper "Ontology-Based Sentiment Analysis Model of Customer Reviews for Electronic Products". In this paper, the products ontology and emotions ontology are created as part of sub_categories to maintain the customers' behavior [4]. All emotional features words cannot be grasped sufficiently in emotions ontology. So, the system in this paper applied lexical based approach by using sentiWordNet lexical resource to define the sentiment scores of each emotional word in given feedback. The product ontology expressed in this paper was specifically used for collecting the targeted item for each product.

S.d.Kok and L.Punt, the author of "Review-Level Aspect-Based Sentiment Analysis Using Ontology" defined the aspect level sentiment analysis [5]. This paper showsed how to parse the sentence to extract aspect related keywords and how to classify the aspect words based on the knowledge base using ontology. So, this paper share the knowledge for extracting the aspect related feature words from the input feedback by using the OpenNLP parser and to get the opinion scores of the words that are remaining after selecting the aspect words by referring the sentiWordNet lexical resources.

Z.T.T.Myint and K.K.Win, the authors of "Triple Pattern Extraction for Accessing Data on Ontology" described the detail structures of building ontology and using SPARQL query to retrieve the data associating with ontological class, relational properties and data

values of domain ontology [6]. So, these works are useful for creating domain ontology of university campus. The purpose of ontology using in this proposed system is to define the level of aspects words.

Goutam Majumder, Partha Pakray, Alexander Gelbukh and David Pinto discussed about the methods of semantic similarity in the paper "Semantic Textual Similarity Methods, Tools, and Applications: A Survey" shows different kinds of methods used in semantic similarity base on the words/ terms, sentences, paragraph and document [7]. From the methods presented in this paper, the appropriate semantic similarity method could be chosen for the student feedback system.

P.K.Singh and M.S.Husain presented the paper "Methodological Study of Opinion Mining and Sentiment Analysis Techniques" in which machine learning techniques can be used in sentiment analysis and opinion mining [8]. They made the comparison between the results of the following approaches: Multilayer Perception, SVM, Naïve Bays, clustering and SVM. Here, the student feedbacks system works with the parser result and lexical resources to guess the opinion of users. So, it does not need to use any of these analysis techniques but it chooses the confusion matrix and accuracy measurement from this paper for evaluating the opinion result of this system.

1.3 Motivation of the System

In the current trend for reaching quality education in university campus, listen to the students' voice is very important manner to make improvement in some special things of the campus. There are many kinds of things in university campus such as classrooms, lab rooms, teachers, course, canteen, etc. It is possible to take manual action to hear all kinds of voice from students. It is also hard to recognize all of the voice and so, this work is naught activity. The alternative manual activity such as colleting the students' voice via paper work is able to solve the recognizing problem. However, this work still has a problem because it was erroneous when distinguishing the paper based on many kinds of things in the campus. On the other hand, it is hard to maintain the collected papers which are needed to compare the next time feedbacks to know the progressive activity on their focus things. For these reasons, these two actions follow the computerized system for more convenient

when taking the feedbacks gathering state and maintaining the collected feedbacks for further use.

When creating the computerized system, the system goals lead to mine the users' emotions upon the specific things from the feedback statements rather than collecting and maintaining the feedback statements. It is not possible to see here and there statements and determining the users' opinion on specific things. So, this work follows the analysis activity known as opinion mining or sentiment analysis which was popular in e-commerce site to get the information for enhancing their products. So, the common web application systems show the list of products or movies or specific things to offer the users to give the feedbacks response about their listed items. When considering only about the keyword "teachers" in students' feedback system, there are a number of teachers who are serving in one university. So, it is not possible to show all highlighted items in the campus of one university over the web page of computerized system.

To be convenient with all the requirements discussed in above two paragraphs, aspect based sentiment analysis is appropriate for students' feedback system. The aspect data from one statement means the keywords relating with the specific products or things including in the feedback response. So, the student feedback system organized the main items of the campus as the keywords which are composed in the ontology by using web ontology language OWL. To dig the aspect words from student's feedback, the system uses the algorithm "*parserAction*" which parse the feedback statement as nouns, verbs and adjectives. The nouns are assumed as the aspect related keywords and are matched with the aspect data from domain ontology with the help of proposed algorithm "*Onto_to_Lists*". The verbs and adjectives relating with one of the nouns are used to guess the users' opinion by utilizing the *sentiwordnet* lexical resource and "*SWN3*" algorithm. For the feedback statement which has more than one noun, the system solves this case together with the domain ontology based on the class hierarchy of the ontology. For evaluating the system's opinion result, it uses confusion matrix and accuracy measurement to proof the accuracy of system. After completing all of the processes discussed above, the system becomes useful for university administrators.

1.4 Objectives of the thesis

The main objectives of this thesis are as follows:

- To analyze the aspect level opinion mining on student's feedback comments
- To develop the domain oriented ontology by organizing the targeted terms of domain area
- To automatically extract the aspect terms from the unstructured sentence
- To assist the administrator who evaluates the performance of university.

1.5 Organization of the thesis

This book is organized with four chapters.

Chapter 1 presents an introduction of Opinion Mining. Then it follows the related works and motivations that have led to the need for further study of Aspect Detection in Sentiment Analysis of Opinion Mining. In addition, objectives and organization of thesis are also presented.

Chapter 2 provides background theory for Opinion Mining and Aspect Detection of Sentiment Analysis. This section also describes how to organize the aspect related data in domain specific ontology and the implementation of OpenNLP and SentiWordNet will be presented. The Semantic Similarity Method and Evaluation method used in this system are also explained in this section.

Chapter 3 discusses the architecture of the Domain Oriented Syntax -based Aspect Detection system. In this section, system design diagram and detailed description of step by step process are described (completely) Moreover, it is followed by the explanation of the proposed algorithms included in the study.

Chapter 4 describes comprehensive case study, merits of implementation with Java language. Moreover, the results of this study taken out in the thesis are also included. In this section, system implementations are presented with corresponding figures.

(Last, but not least,) chapter 5, concludes the research paper with limitations and future studies that will challenge for further research issues.

CHAPTER 2
BACKGROUND THEORY

This chapter intends to point the background theory applied in this research work. It explains about the theory of opinion mining, tools used in sentiment analysis, ontology based sentiment analysis, lexical resource used for sentiment analysis, semantic similarity methods and evaluation methods used in this work. The organization of the chapter is as follows:

In the first section, opinion mining, levels of sentiment analysis and different kinds of approach that can be applied in sentiment analysis are described. Next, tools used in sentiment analysis with some kinds of parser are explained. And then ontology- based sentiment analysis and sentiment analysis are defined. After that, the semantic similarity methods and the methods used for evaluating the opinion result are presented as (the final part).

2.1. Opinion Mining

Opinion Mining, also known as Sentiment Analysis is an extraction of useful knowledge from the people's opinion about some things or topics. The major work of opinion mining is to find a set of results for a given product or things that is producing a list of features of products and their related opinion. Its main goal is to make computer to be able to recognize and generate emotions as human.

2.1.1 Sentiment Analysis

Nowadays, (a number of people) use Internet to learn or view the quality of products or other things and so, textual data, which is growing every day. For these reasons, there is an increase in accessibility of opinion resources such as movie reviews, product reviews, blog reviews, etc. and the challenge is to mine a large amount of texts using appropriate algorithms which understands the opinion from the texts. This process is called sentiment analysis or opinion mining and the resulted sentiments from this analysis help the organization or company for making decisions upon their products or some other things. In addition, these sentiment result results are also helpful for users as well because

reviews (of) products or other things can help the users in deciding whether it is good or not.

Sentiments are also known as moods of users. It can be neutral, excellent, good or bad. The analysis on such users' emotions is called sentiment analysis. This task is a kind of natural language processing and it uses computing approaches to extract the opinion related word and define it as neutral, positive or negative. The web page composed with the structured and unstructured textual information that includes opinion or sentiments of the user. The sentiment analysis defines the emotion of users and status of emotion [9]. The classifications of sentiment analysis are as follows:

- **Document level** –This analysis classifies the entire document opinion into different sentiment such as positive, negative or neutral for a specific thing.
- **Sentence level** –This analysis determines whether each sentence written by users expresses a neutral, positive or negative opinion, for a specific thing.
- **Entity and Aspect level** –Aspect level is the sentiment analysis based on feature and define sentiment on desired aspect/feature in a review.

2.1.1.1 Document Level Sentiment Analysis

This kind of analysis intends to classify a sentiment result for the entire document as shown a positive, neutral or negative opinion. So, it is called as the document level sentiment analysis because it holds the entire document as a basic information unit. For example, when considering the statements in one document like that "I bought a Hand Phone a few months ago. This is a pretty one. Its touch screen is cool. The voice quality of this phone is also clear. So, I love this phone!", all of these statements are determined as positive sentiment. Document level classification can do the best when the document expresses an opinion or sentiment on a same kind of entity.

When an opinion document was written by composing with more than one entity and then the sentiments or opinion sense on each entity can also be different. That means that the sentiment result may be positive for some entities and negative for some others entities. Therefore, it cannot be assigned with one sentiment result for the entire document. If each sentence in the document usually focuses on a single product or service and that is review by a single user, the sentiment result for the whole document can be defined based

on average score. However, when the reviewers express opinions on multiple entities and compare these entities within the sentences, it cannot be define the sentiment result for this document. Document-level sentiment classification does not perform fine grained tasks for each entity because it can only determine the sentiment for all the sentences.

2.1.1.2 Sentence Level Sentiment Analysis

This analysis defines sentiment or opinion on each sentence. There is no difference between document and sentence level sentiment analysis because documents are composed with short sentences. One of the assumptions is that a sentence usually has a single opinion but this rule is not for all sentences. However, a document typically contains many of opinions. Sentence Level Sentiment Analysis usually contains two steps: [10]

- Subjectivity classification of a sentence: subjective and objective
- Sentiment classification on subjective statements : positive and negative

Objective sentence shows some information but expressing no sentiment or opinion, on the other hand, subjective sentence includes personal views, feelings, beliefs, or emotions. Some of the identification of subjective sentence can be accomplished by using some supervised learning which can classify positive or negative opinions and the others do not. Sentence level sentiment analysis is suitable for simple sentences which has one sentiment. For example, “This camera’s picture quality is amazing”, this can be identified as positive sentence. However, just defining the opinion of that sentence as positive or negative is not sufficient at all. Multiple entities of the problem may be contained in a subjective. For this reason, it needs to dig the deep level of sentiment analysis.

2.1.1.3 Aspect Level Sentiment Analysis

As we learned in the two previous sub-sections, document level or the sentence level sentiment analysis is insufficient for all applications because they do not find targets of opinion and cannot assign opinion result to such targets. It means that for a document that has positive opinion, all aspects of the entity in this document have no positive opinion and also for a document that has negative opinion, every entity in this document is not

negative. For solving this problem, more sophisticated analysis is needed to discover the aspect words and determine whether the sentiment is positive or negative on each aspect.

Aspect level sentiment analysis is intended to identify or extract the features or aspects and determine on each main features whether the opinion is positive, negative, or neutral. This can be decomposed into two steps such as aspect extraction and aspect sentiment classification which are explained (in detail) in the next section.

2.1.2 Aspect-Based Sentiment Analysis

To extract the main features or aspect words from the sentence, it is needed to go aspect level sentiment analysis or opinion mining, which was also called the feature-based opinion mining. In which, the opinion target is parsed into entities and related aspects. The aspect is used to represent the main entity of domain in the result. The two main important tasks in aspect-based sentiment analysis are described as follows:

- **Aspect sentiment classification:** This task determines whether the opinions on different aspects are positive, negative, or neutral. For example, in the sentence, "This phone's voice quality is good" the sentiment result on the aspect word "voice quality" is positive. So, the sentiment result for the entity "phone" is also positive.
- **Aspect extraction:** This task extracts the aspect words that need to be evaluated. For example, in the sentence, "The voice quality of this phone is amazing," the aspect is "voice quality" of the entity represented by "this phone". In this case, the evaluation is not about the phone as a whole, but about its voice quality. In other case, for the sentence "I love this phone", the aspect entity is phone as a whole.

2.1.2.1 Aspect Sentiment Classification

Sentiment classification for each aspect entity has two main approaches such as the supervised learning approach and the lexicon-based approach [11].

For the supervised learning approach, the modelling based methods are used for sentence-level and clause-level sentiment classification. The key issue in this approach is how to determine the scope of sentiment expression. Supervised learning is dependent on the training data. A model or classifier is trained from labeled data in one domain often

performs poorly in another domain. Although domain adaptation is still far from mature, this method is also mainly used for document level sentiment classification because documents are long and contain more features for classification than individual sentences or clauses. Thus, supervised learning has difficulty to scale up to a large number of application domains.

The lexicon-based approach can avoid the issues of supervised learning methods described in the previous paragraph and it performs quite enough for large number of domains. This method is based on unsupervised learning. They use a sentiment lexicon which was composed with a list of sentiment words by attaching their positive and negative scores and the sentence parse tree to determine the sentiment orientation on each aspect in a sentence. Of course, the lexicon-based approach also has its own advantage such as it is fast and it has no necessary for training data. The difficulty of this approach is unable to work for multiple word phrases and unable to deal with multiple word senses. Although the performance of supervised machine learning techniques is better than the unsupervised approaches, supervised approaches require huge amounts of training data with class label. So, unsupervised approaches are also important because collecting the labeled data is very exhausted process and on the other hand, acquiring unlabeled data is easy. Most of the domains' knowledge omits the class label in training data in this case unsupervised method is necessary for application development.

2.1.2.2 Aspect Extraction

To define the opinion for corresponding aspect, it also needs to do the aspect extraction process, which is similar with an information extraction task. However, one of the characteristics of sentiment analysis is that a sentiment result always refers to a target. The target is also known as the aspect word which is needed to extract from input sentence. So, it is necessary to recognize each input sentence with its target or aspect word. For this reason, it can be known as sentiment analysis that can play two important roles which are finding a positive or negative nature of sentiment and defining an aspect or target word. There are four kinds of approaches in aspect extraction: [12].

- **Extraction based on frequent nouns and noun phrases:** This approach finds nouns and noun phrases from review or feedbacks for a given domain. These nouns and noun phrases can be defined by a part-of-speech (POS) tagger of parser. And then, these nouns' occurrence frequencies are counted, and select the important nouns according to the occurrence frequency threshold. The reason that this approach works is that when people comment on different aspects of an entity, the vocabulary that they use usually converges. Thus, those nouns that are frequently talked about are usually genuine and important aspects. Hence, infrequent nouns are likely to be non-aspects or less important aspects. Although this approach is simple, some commercial companies use this approach with several improvements because it is still quite effective.

- **Extraction using supervised learning:** There are many supervised learning algorithm in the past information extraction process and most of these methods are based on sequential learning approach. These approaches require class label for training data set. And it is needed to manually organize the aspects and non-aspects within a corpus.

- **Extraction using topic modeling:** These methods are discovering topics from given text documents. These approaches are kind of unsupervised learning methods. The topic model uses probabilistic procedure to define clusters of words and define the topic for each cluster. It can be applied in aspects extraction. However, there is some difficulty that is topics can over both sentiment and aspect words. So, they need to be separated for making sentiment analysis.

- **Extraction by exploiting opinion and target relations:** This approach based on the idea such as the opinions and their targets are obviously related. Their aspect extraction process can be made based on these relationships. If the aspect words cannot be found in a sentence but some sentiment words can be found in this sentence, the noun or noun phrase of each sentence is extracted as aspect word. For example, the sentence, "The phone is amazing", the "amazing" is a sentiment word

and then “phone” is extracted as an aspect. This thesis mainly depends on this kind of approach and it takes such noun to be confirmed that are included in domain terms.

2.1.2.3 Aspect polarity aggregation

According to the relationships between sentiments or opinions and their targets or aspects, sentiment words has been recognized by identified aspects. Counting the scores of opinion words in which it needs to assign orientation score (+, -) values to all opinion words: positive opinion words and negative opinion words, respectively. In this process, the opinion orientation score of a given piece of text is simply considered to be equal to the sum of orientation scores of all opinion words found. This stage groups the sentiments for aspects together and produces a final summary for the extracted aspect.

2.1.3 Approaches using in Sentiment Analysis

Opinion mining or sentiment analysis can range from sentiment-polarity summarization in reviews to determining the strength of opinions in news articles and to identifying perspectives in political debates. This analysis mainly depend on identifying and extracting some features and aspect from the source data. There are solutions proposed to these problem domains as organizing different aspects of machine learning approaches. Although these aspects may seem to be general themes underlying most machine learning problems, unsupervised learning approaches are highlighted what is unique for sentiment analysis and opinion mining tasks. The next consideration in opinion mining is using the *SentiWordNet* based publicly available library and specific domain.

2.1.3.1 Supervised Approaches

It trains the labeled data to make a learning model. These approaches learn to find the expected targets that are mapped the input examples. The supervised learning algorithm should generate the training data after the training process has been done with correct implementation. So, it can attach the new data which has never seen before with appropriate target.

- **Naïve Bayes:** The classifier, Naïve Bayes is a probabilistic model. It is feature independent method for classifying input data which is mostly used in text classification process. It is simpler computation with low computational cost and relatively high accuracy. The algorithm will calculate the probability of every word in the training set to indicate the positive or negative sense. Then the algorithm is ready to classify new data [12].
- **Support Vector Machine:** This approach is non-probabilistic model which is also known as binary linear classifier. It plots the training data on multidimensional space and then separates the classes with a hyperplane. If the classes are not separable in the multidimensional space, this method will add a new dimension in order to separate the classes. The data that placed on each plane are called the supports. The main difficulty of the SVM method is that the more adding extra dimensions, the more increase the size of the feature space [13].
- **Decision trees:** This approach is well known machine learning algorithms. It partitions the training data base on information gain to obtain smaller parts which are used for identifying patterns. The knowledge generated from this approach is represented like a tree structure. In this decision tree structure, it represents the entire data set as a root node, the intermediate small sets of data as decision nodes, which perform the computation and the last node which can define the class as leaf nodes. The new data will eventually pass through these decision nodes to reach a leaf node and will be assigned with class [14].

2.1.3.2 Unsupervised Approaches

It can mine the data without prior knowledge like training data. It measures how far a word is inclined as positive and negative. This approach makes use the lexical resource to find the similar orientation intend to have similarity.

- **Unsupervised learning using POS**: This approach finds the sentiment related words of each statement by using the rules of part-of-speech (POS) patterns. And then, the predictive opinions of unknown sentiment phrases are computed based on the collected data from sentiment lexicon. So, the sentiments result is predicted based on the sentiments label of nearby known sentiment words inside the same

group. They assume that the opinionated sentences contain at least one sentimental word or phrase for opinion mining. This unsupervised opinion mining is performed by using POS pattern [15].

- **K-means clustering algorithm**: This algorithm uses the TF-IDF (term frequency –inverse document frequency) weighting technique or some other similarity methods to cluster the raw data. And then, it uses voting mechanism to extract more stable clusters. The good result is based on many of the implementation of clustering process. It collects the documents into positive group and negative group [16].

This thesis relies on unsupervised learning using POS tagging approach to make sentiment analysis.

2.1.3.3 Domain Adaption

The applied domain of the items can influence the accuracy of sentiment classification. To make use of this approach, it firstly extracts the sentiment related terms from the sentences and then defines their scores by confirming with the *SentiWordNet* lexical resource. The scoring schemes have applied on three kinds of linguistic feature words, namely adjectives, adverb and verbs. These are used in order to obtain the better result. For evaluating the accuracy and performance of this *SentiWordNet* based approaches, it computed the measure of Accuracy, F-measure and Entropy base on the standard performance metrics.

SentiWordNet is a kind of library which is publicly available. It is used in sentiment classification when using the approaches combined with linguistic model. *SentiWordNet* defines and organizes the words with their scores and it can classify the sentences as positive, negative or neutral opinion based on these score. Aspect level sentiment classification is needed when considering specific feature of domain [9]. The main idea of this thesis follows the flows of this approach.

2.2 Tools used in Sentiment Analysis

There are varieties of open-source tools for text-analytics such as natural-language processing tool used in information extraction and classification approaches that can be applied in sentiment analysis. The following listed tools can be applied on text related sources [28].

- **LingPipe** – It is a suite of speedy, scalability and stability java tools for lingual based processing on text such as part of speech tagging (pos), keyword extraction, classification and clustering etc. It is a kind of mature open source NLP toolkits and widely used in industry.
- **OpenNLP** – This tool can make advanced text processing and is hosted with a variety of java-based tools which perform NLP tasks, such as sentence segmentation, tokenization, named entity extraction, part-of-speech (POS) tagging, co-reference resolution, parsing and chunking.
- **Stanford Parser** – It is a tool composed with java packages and it can support to make part of speech tagging and sentence parsing. This tool is provided from the Stanford NLP group
- **Opinion Finder** – It has multi-stages to do NLP process. It finds subjective sentences and extracts all possible aspects from these sentences such as the subjectivity source and words which can express the positive or negative opinion.
- **Textir** – Its 'mnlm' function sparse the multinomial logistic regression such as the topic functions and a concise partial least square routine, for obtaining the efficient dimension selection in latent topic models and estimation.
- **NLP Tool suite** – It is a useful tool suite for information extraction, text mining and semantic search. Most of the expanded tool suite was developed based on the machine learning techniques and so, they are language domain independent.

This thesis used OpenNLP tools for POS tagging because it is open source software and provides many tutorial to use it.

2.3 Ontology based Sentiment Analysis

Aspect based sentiment analysis breakdowns its work into two parts which are sentiment identification process and aspect detection process. Sentiment analysis defines

the opinion of the whole sentence such as positive or negative; and aspect detection digs the targeted word from input sentence. So, the opinion of the input sentence is identified upon other word in this sentence which means that the emotion of users on the targeted aspect word. In order to confirm the extracted words are aspect or not, the domain specific ontology can help this confirmation process and it well organizes and expresses specific things of the domain knowledge.

2.3.1 Structures of Ontology

Ontology is organized with four major components which are instances, concepts, relations and axioms. The following definitions refer to these components: [18].

- **Concept:** It can be seen as a class in ontology representation and it can be seen as objects collection. This element refers to the domain class in which members are included and can share common properties between them. This concept has object oriented structure and so, it is represented as a "super-class", or "parent class", and a "subclass" which is so-called "child class".
- **Instance**: It can be seen as an individual in ontology representation which represents a specific object of a concept class.
- **Relation**: It can be seen as a slot in ontology representation and this refers to the relationships between concepts. Specifically, the first concept of this relationship is represented as the domain, and the second concept represented as the range
- **Axiom**: It refers to constraints assigned on the classes or individual instances defined by first order logic and which are used to check the consistency of the ontology.

2.3.2 Roles of Ontology

The major objective of ontology is to support useful knowledge concerned with specific domains which are accessible by both the developers and computers. Ontology can improve the information retrieval process and data interoperability between different domain applications [19].

An ontological model is capable of expressing significant categories in the context of the entity embodied by an ontology structure. Conceptual structures are correlated with

a structure of types of situations to reflect the level of concepts for various kinds of entity. Modern approaches of sentiment analysis benefit from this ontological structure in order to do the aspect-extraction process. Ontologies can provide facility for knowledge acquisition and representation in solving complex form of domains

For determining the aspect included in a text is, one approach which consists of similarity measurement between terms and ontology classes. The confirming process of ontology class or instances with the corresponding aspect for a given text is called semantic annotation. In order to perform the semantic annotation process, NLP techniques can help to identify entity or part-of-speech in the structures of a sentence. After identifying these lingual units, their associated concepts and relationships can be identified. According to these purposes of ontology, a sentiment can be assigned as concept in ontology.

2.4 Lexical Resource for Sentiment Analysis

Lexical resources are created with database, spreadsheet or text-based files in which sentiment word are organized with appropriate format. There are various kinds of sentiments on each word such as strong, weak, positive or negative. The polarity of words is important in order to perform sentiment analysis and classify sentiments into neutral, positive or negative opinion. The sentiment lexicons can help to accomplish this kind of task. There are different kinds of sentiment lexicons which can be seen in the following sections [20].

2.4.1 Bing Liu's Opinion Lexicon

This lexical resource uses adjective Synsets of WordNet and their antonyms. It collects 6,800 words drawn from product reviews, labeled these words using a bootstrapping method. It organized 2006 positive words and 4783 negative words [23].

2.4.2 SentiWordNet

This lexical resource can provide sentiment scores to the applications of opinion mining and sentiment classification. It defines all WordNet Synsets with their respective degrees of neutrality, positivity and negativity. The SentiWordNet version 3.0 is based on

WordNet 3.0 and it refines the scores for each opinion word by using a random-walk step and semi-supervised learning step [24].

Table 2.1 A fragment of the SentiWordNet resource

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	00001740	0.125	0	able#1	(usually follow by 'to')having the necessary means or[…]
a	00002098	0	0.75	unable#1	(usually follow by 'to') not having the necessary means or[…]
a	00002312	0	0	dorsal#2 abaxial#1	facing away from the axis of an organ or organism; […]
a	00002527	0	0	ventral#2 dadxial#1	nearest to or facing to ward the axis of an organism; […]
a	00002730	0	0	acroscopic#1	Facing to on the side toward the apex
a	00002843	0	0	acroscopic#1	facing on the side toward the base
a	00002956	0	0	abduction#1 abducent#1	especially of muscles; […]
a	00003131	0	0	adductive#1 adduction#1 adducent#1	especially of muscles; […]
a	00003356	0	0	nascent#	being borm or begining; […]
a	00003553	0	0	emerging#2 emergent#2	coming into existence; […]

2.4.3 The General Inquirer

This is the oldest sentiment lexicons and that can be published as publicly. This work is based on in content analysis and cognitive psychology. It has 1915 positive words and 2291 negative words. It includes further semantic dimensions, such as strength or active/passive orientation. This resource organized the words as semantic, syntactic and pragmatic information in to the words represent part-of-speech tagged. It has created for mainframe computer and it is not suitable for desktop computers [22].

Table 2.2 A fragment of the General Inquirer resource

	Entry	Positiv	Negativ	Hostile	184 classes	Othtags	Defined
1	A					DET ART	...
2	ABANDON		Negativ			SUPV	
3	ABANDONMENT		Negativ			Noun	
4	ABATE		Negativ			SUPV	
5	ABATEMENT					Noun	
35	ABSENT#1		Negativ			Modif	
36	ABSENT#2					SUPV	
...							
111788	ZONE					Noun	

2.4.4 MPQA Subjective Cues Lexicon

Multi-Perspective Question Answering lexicon is a kind of Subjectivity Lexicon. This lexical resource collects 2718 positive words and 4,912 negative words from a combination of some resources which are hand-labeled for sentiment, the lists of General Inquirer and subjective clues in a bootstrapped list. The lexicon also includes four tags refer to polarity words such as neutral, positive and negative and labels for improving reliability such as strongly subjective words or weakly subjective words [23].

Table 2.3 A fragment of the MPQA resource

	Strength	Length	Word	Part-of-speech	Stemmed	Polarity
1.	type=weaksubj	len=1	word1=abandoned	pos1=adj	stemmed1=n	priorpolarity=negative
2.	type=weaksubj	len=1	word1=abandonment	pos1=noun	stemmed1=n	priorpolarity=negative
3.	type=weaksubj	len=1	word1=abandon	pos1=verb	stemmed1=y	priorpolarity=negative
4.	type=strongsubj	len=1	word1=abase	pos1=verb	stemmed1=y	priorpolarity=negative
5.	type=strongsubj	len=1	word1=abasement	pos1=anypos	stemmed1=y	priorpolarity=negative
6.	type=strongsubj	len=1	word1=abash	pos1=verb	stemmed1=y	priorpolarity=negative
7.	type=weaksubj	len=1	word1=abate	pos1=verb	stemmed1=y	priorpolarity=negative
8.	type=weaksubj	len=1	word1=abdicate	pos1=verb	stemmed1=y	priorpolarity=negative
9.	type=strongsubj	len=1	word1=aberration	pos1=adj	stemmed1=n	priorpolarity=negative
10.	type=strongsubj	len=1	word1=aberration	pos1=noun	stemmed1=n	priorpolarity=negative
...						
8221.	type=strongsubj	len=1	word1=zest	pos1=noun	stemmed1=n	priorpolarity=positive

2.4.5 LIWC (Linguistic Inquiry and word count)

LIWC is a kind of tools to make text analysis and it is widely used in the social sciences of computing. It applied natural language processing beyond the classification. It is firstly created with a small numbers of emotion-related words. This initial success led

the program to collect a more categories of words from which it was develop to the first version of the LIWC for commercial use [21].

Table 2.4 A fragment of the LIWC resource

Category	**Examples**
Negate	anint, anin't, arent, aren't, cannot, cant, couldn't,……..
Swear	arse, arsehole*, arses. Ass, asses, asshoe*, bastard*, …
Social	acquainta*, admit, admits, admitted, admitting, adult, adults, advice, advis*
Affect	abandon*,abuse*, abusi*, accept, accepta*, accepted, accepting, accepts, ache*
Posemo	accept, accepta*, accepted, accepting, accepts, active*, admir*, ador*, advantage*
Negemo	abandon*, abuse*, abusi*, ache*, aching, advers*, afraid, aggravate*, aggress*
Anx	afraid, alanm*, anguish*, anaxi*, apprehens*, asham*, aversi*, avoid*, awkward*
Anger	Jelous*, jerk, jerked jerks, kill*, liar*, lied, lies , lours*, ludicrous*, lying, mad lying, mad

Among these lexical resources described above, SENTIWORDNET is freely available for research purposes, and it defines the score for each sentiment words. So, this thesis chooses the SentiWordNet lexical resource to perform the analysis.

2.5 Semantic Similarity

The semantic similarity between texts or words is always needed to find in various text based analysis such as feedback analysis systems, information retrieval, text mining and so on. Words can be determined into similar or not based on the similarity measure between words or phrases. Similarities between documents are the basis of text mining or informal retrieval and so on. There are different levels in similarities measuring methods. [25]

- **Word similarity**: Similarity between words can be measured based on the spelling or the meaning of words. If two words are similar in spelling, they are possible to

get the highest similarity measure. Lexicon dictionary can be used to calculate the meaning similarity between words.

- **Sentence similarity**: The similarities of each word in different sentences lead to have great similarity between two sentences. Words and their orders of sequence in the sentences are two important factors to calculate sentence similarity.
- **Document similarity**. The similarities between each sentence lead to have great similarity between documents. Commonly used approaches are often based on similarity between the keyword sets or similarity between the vectors of keywords.

In this thesis, the system needs to confirm the action words from the feedback sentence with the words organized in lexicon to get the opinion score. So, this thesis uses the word or short string similarity measure.

2.5.1 Methodologies in Semantic Similarity

The process of determining semantic similarity between texts or words has been done by using various similarity measures. The kind of similarity measures can be categorized as follows [26].

- **Topological based similarity:** In order to handle synonymous problem in distributional application, it needs to organize semantic features in lexical sources to form knowledge-based measures. In the knowledge-based measures, a hierarchical model is used to measure the similarity between words based on features extracted from a provided lexical source.
- **Statistical based similarity**: To measure the statistical similarity between sentences, it needs to define symbolic characteristics and structural information. Symbolic or semantics similarity means that words in two sentences are same and structural relations mean that the relations between words and the distances between words. It could measure the similarity between sentences without any prior knowledge but only on the statistical information of sentences. The methods Latent Semantic Analysis (LSA), Explicit Semantic Analysis (ESA), Pointwise Mutual Information, Hyperspace Analogue to Language (HAL), etc. are used in statistical based similarity.

- **Semantic based similarity**: To measure the semantic based similarity between sentences, it can be taken based on the meanings of the words and the syntactic of sentence. Although the sentences have different symbolic and structure information, the same meaning of these sentences can lead to same similarity. These measures are based on discovering the similar concepts in predefined ontologies. Character based or term based similarity measurement methods are used in this approach.
- **Vector space model based similarity**: To measure the similarity for whole documents, Vector space model or term vector model is used which is an algebraic model for representing text documents as vectors of identifiers. It is used in information filtering, information retrieval, indexing and relevancy rankings. Cosine similarity measure is used to find the similarity between two vectors of an inner space.
- **Word alignment based similarity**: Word alignment similarity measure is used when the two sentences have similar meaning. It looks for the correspondence mapping between the words of each sentence which are corresponding to the same meaning within the context. It can be evaluate by using the labelled corpus. In word alignment task, the alignment has to be zero-or-one in case it can be distributed the word weights in a fuzzy way.
- **Machine learning based similarity**: Similarity can also be learned in an area of machine learning in artificial intelligence. This learning is similar as regression and classification, but the goal is to learn how similar or related of two objects are. The application areas using this approach are recommendation systems, visual identity verification, face verification, and so on. It uses the approaches like a nearest neighbor to define the similarity of two or more objects based on distance functions.

In this thesis, the character base semantic similarity measure is used to identify the similarity between extracted aspect word and the words in ontology.

2.5.2 Character based Semantic Similarity

Character-based Similarity is also called sequence-based similarity measurement and it is also a kind of string-based semantic similarity. It takes two strings of characters

and then calculates the distinct between any data points or groups to define the similarity between those data points [27]. The methods used in this approach are listed as follows:

- **Longest Common Substring (LCS):** This measure is used to find the longest substring from a string. When comparing the two strings, it finds the similarity based on the longest common sequence of characters.
- **Damerau-Levenshtein:** This measure builds a distance or string metric for the two strings. It gives a value that required for transforming one string into another. This transformation is done by insertion, deletion or substitution of a single character or a transposition of two adjacent characters.
- **Jaro:** This measure finds the distance for similarity measure of two strings. It is normalized to *0* means no similarity and *1* means an exact match.
- **Jaro–Winkler:** This measure is a semantic similarity measure of two strings. It is a type of edit distance and variant of Jaro distance metric. The higher the Jaro–Winkler distance, the two strings is more similar.
- **Needleman-Wunsch:** This measure is called optimal matching algorithm and global alignment technique. It is a type of dynamic algorithm used in Bioinformatics to align the protein sequences.
- **Smith-Waterman:** This measure is a variation of Needleman-Wunsch algorithm and measure the similarity by comparing within the segments of the string and optimizes the similarity. It cannot be used in large scale problem.
- **N–gram:** This measure is a probabilistic language model used for predicting the next term in a sequence of (n - 1) terms or characters. The main advantage of this model is simplicity and scalability.
- **Syntactic N–gram:** This measure is an advance of the N-gram model. It overcomes the weakness of n-gram model in which the n-gram elements are not in the order in a text, but the order is in the corresponding syntactic tree.

Among the methods listed in above, this thesis chooses the jaro similarity because it is suitable for short string and this methods is detail explained in the next section.

2.5.3 Jaro Semantic Similarity Measurement

As described above section, it is also known that Jaro Similarity is the measure of distinct to identify similarity between two strings. The value of Jaro distance ranges from 0 to 1, where 1 means the strings are equal and 0 means no similarity between the two strings. Jaro Similarity is calculated by using the following formula.

$$sim_j = \begin{cases} 0 & \text{if m=0} \\ \frac{1}{3}\left(\frac{m}{|s1|} + \frac{m}{|s2|} + \frac{m-t}{m}\right) & otherwise \end{cases} \quad \text{Eq (2.1)}$$

Where:

- $|Si|$ is the length of the string Si;
- m is the number of matching characters;
- t is half the number of transpositions. (words with different order)

The characters are said to be matching if they are the same and the characters are not further than (max ($|s1|$, $|s2|$) / 2) – 1 and transpositions means half the number of matching characters in both strings but in a different order.

Example:

Let s1=”arnab”, s2=”raanb” are the two strings.

- In this case, both the strings have 5 matching characters, but the order is not the same.
- The number of characters that are not in order is 4 and so the number of transpositions is calculated as 4/2 and the result is 2.
- Base on the facts calculated in above, Jaro similarity can be calculated as follows:

  Sim_j = (1/3) * {(5/5) + (5/5) + (5-2)/5 } = 0.86667

This result shows that these two strings do not exactly match but if it is no need to consider their characters order, these two strings are almost the same. The system in this thesis

applied the Jaro Similarity Measurement because most of the aspect words are needed to be exact match.

2.6 Evaluation Methods

For achieving the accuracy in any analyses is a complex task because it requires the identification, estimation, and elimination of sources to reduce analytical error. When considering the concepts of accuracy, it will include choosing measurement process for corresponding analytical method and stating the statistical terms such as precision, imprecision, accuracy, inaccuracy, systematic error, overall or total error, true value, traceability, and compatibility. These concepts provide the basis factors to measure the performance of analytical methods. There are many evaluation methods used in machine learning such as crossed validation and finding evaluation index. Crossed validation methods depend on analytical technique and finds accuracy together with this technique. So, finding evaluation index for accuracy is the best solution for the lexicon based unsupervised opinion mining. The common way for computing the required indexes is based on the confusion matrix as shown below: [17]

Table 2.5 The Confusion Matrix

	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Where,

- TP means True Positive
- FP means False Positive
- FN means False Negative and
- TN mean True Negative

The accuracy measure is calculated as follows: Firstly, it needs to find the rates for each true positive, false positive, false negative and true negative.

- True positive rate (TPR) : TP/(TP + FN)

- True negative rate (TNR): TN/(TN + FP)
- False positive rate (FPR): FP/(FP + TN)
- False negative rate (FNR): FN/(FN + TP)

Accuracy = (TPR + TNR) / (TPR + TNR + FPR + FNR) Eq (2.2)

Accuracy is the portion of all true predicted instances against all predicted instances. An accuracy of 100% means that the predicted instances are exactly the same as the actual instances.

CHAPTER 3

SYSTEM ARCHITECTURE

As the information source on the web is more and more obtained nowadays, opinion mining or sentiment analysis becomes popular. Its work is to dig the important information from review statements or feedbacks. In regular opinion mining techniques, the system can determine the peoples' feeling for a given feedback or comment such as positive or negative. But the current trend of opinion mining is to seek the aspect word which means that it is fine grained sentiment information on different kinds of statements. In this thesis, the aim of this proposed system is to do aspect level sentiment analysis on student feedback system. University of Computer Studies, Taungoo(UCSTaungoo) supposed the required feedback data. For the operation of this system, OpenNLP parser is used for part of speech (POS) tagging and sentiWordNet lexical resources is used for identifying the sentiment score for each word. As the preprocessing stage of this system, Domain Specific Ontology is created by using the main elements of UCSTaungoo which is the main component of Aspect Detection. At the evaluation stage, the evaluation index, accuracy measure is calculated for this system based on the data on confusion matrix. This system will support the admin level to determine the University's performance.

3.1 Overview of the Proposed System

In feedback analysis system, the users or students need to send the feedback comments for the entities of university campus such as the status of rooms, courses, lab rooms and canteens. And then, the system analyzes the feedbacks in order to find what entity of the campus it intends to. After that, the students' opinion relating to the specific entity can be identified and the admin can easily make a decision for which entities of university campus is needed to upgrade. To do this, organizing the specific elements of the campus, also called the collection of aspect words, is important to concern.

For this reason, the domain specific ontology that can organize and express the aspect words is needed to create for this proposed system. In this system, the protégé editor is used to create the ontology with web ontology language (OWL) format. The ontology

can compose the complex representation of knowledge and the relationships between entities. Moreover, the ontology can act as the major resource to extract the aspect words from the feedback statements. This process is done by SPARQL which query language of ontology. How to extract the aspect words from the feedbacks and match with the aspect words in the ontology is still needed to consider.

To solve this problem, OpenNLP parser is used which can provide tokenization and part of speech tagging process, etc. OpenNLP parser parsed the input feedback into three lists such as verbs list, adjectives list and nouns list. Among them, noun forms of words can relate to the aspect words collected in the domain specific ontology. Each word in the nouns list is matched with the entities by tracing the relations within this ontology. By doing this, the system determines whether each noun in the list represents the entity in the university campus or not. After this stage, the opinion score of the related aspect word for a given feedback is still needed to find.

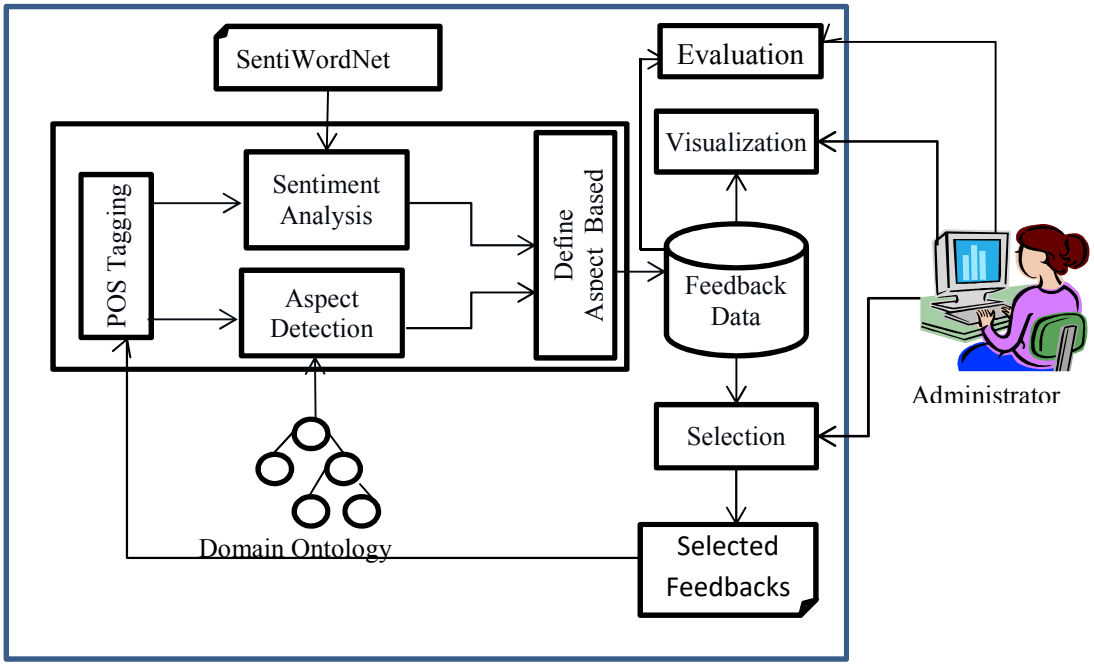


Figure 3.1 Overview of Aspect Detection System

The adjective and verb forms of words reflect the user's opinion and so which are the most suitable forms to define the opinion. This opinion scores identification process can be done by confirming with the lexical resource called *sentiWordNet*. This lexical resource composed with the words and their positive and negative real values as scores.

The type and description of these words are also attached in this lexical resource. So, the system can easily define the scores of each word for a given feedback.

After finishing the processes described above, the feedbacks and their related aspect words and scores are collected as history data sets for the evaluation step. The administrator can choose the appropriate view to see this history records such as bar graph and excel sheet. This proposed system calculates the accuracy index on these pre-collected history data. The overview of proposed system architecture is shown in figure 3.1 and the data flow diagram of this system is described (in detail) in the following figure.

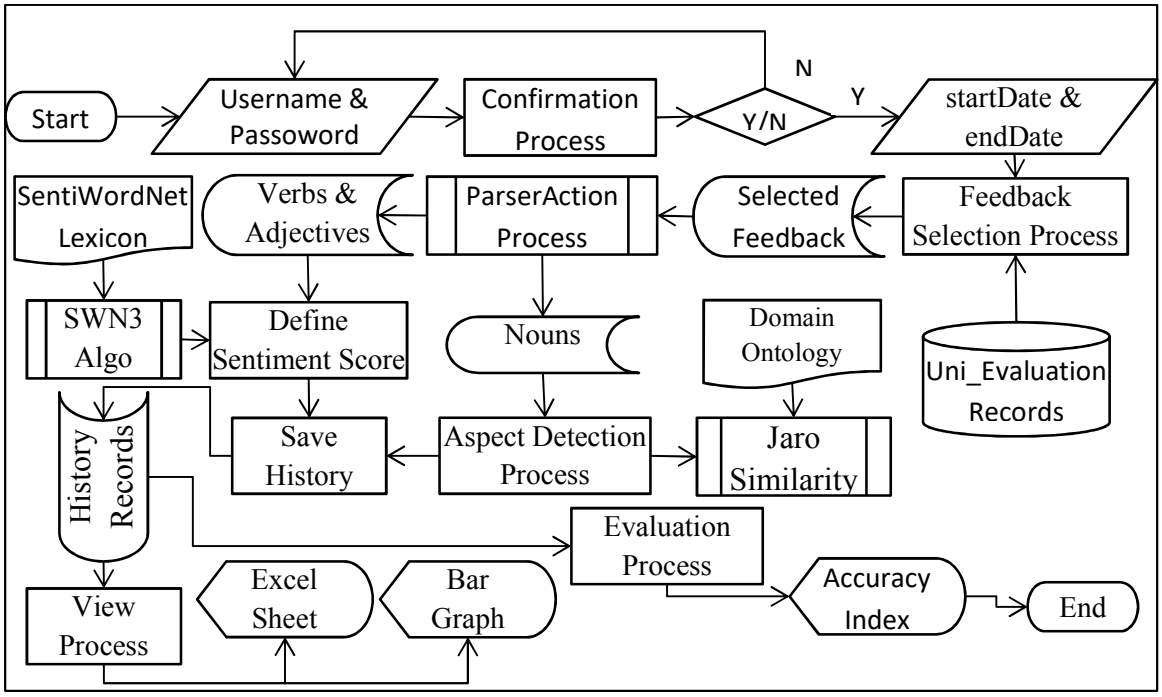


Figure 3.2 Data flow diagram of Aspect Detection System

3.2 Role of Parser

OpenNLP parser can assist to extract the words of feature or opinion from the input feedback. It parses the feedback into each part of speech and attaches with their type. There are many types of words in one sentence such as verb, adjective, adverb, nouns and so on. The algorithm name as *ParserAction* used in this proposed system extracts the verbs,

adjectives and nouns as lists from the feedback sentence. This algorithm can be seen in the following figure:

Algorithm: *ParserAction*. Parse the input sentence into part of speech
Input : feedback statement, *line*
: *OpenNLP* parser model, *model*
Output : the lists of nouns, verbs and adjectives
Begin
(1) *nouns* <= {“NN”,”NNS”,”NNP”,”NNPS”}
(2) *verbs* <= {“VB”, “VBG”, “VBD”, “VBN”}
(3) *adjs* <= {“JJ”,”JJR”,”JJS”,”ADVP”,”RD”}
(4) *tokens* <= parserTool.parse(line, model)
(5) For each *t* of tokens
(6) If *type(t)* included in *nouns*
(7) *nlist* <= *text(t)*
(8) If *type(t)* included in *verbs*
(9) *vlist* <= *text(t)*
(10) If type(*t*) included in *adjs*
(11) *alist* <= *text(t)*
(12) End For
(13) Return collection of *nlist, vlist* and *alist*
End

Figure 3.3 parser_Action algorithm

The function P*arserTool.parse* in this algorithm is already supported as library by *OpenNLP* parser and the main task of this function is to define part of speech (POS) tagging for input sentence. Moreover, this each word in the input sentence is attached with corresponding tag or form of word. *In ParserAction* algorithm, it emphasizes to extract the words which have commonly used tags. The descriptions for these commonly used tags are listed as shown in table 3.1.

Among them, the lists of verbs and adjectives are sent to identify the sentiment score of each word and nouns list are used for aspect detection process. The sentiment analysis task will be explained in the following sections.

Table 3.1 POS tag description

Tag	Group	Description
NN	noun	Noun, singular or mass
NNS	noun	Noun, plural
NNP	noun	Proper noun, singular
NNPS	noun	Proper noun, plural
VB	verb	Verb, base form
VBG	verb	Verb, gerund or present participle
VBD	verb	Verb, past tense
VBN	verb	Verb, past participle
JJ	adjective	Adjective
JJR	adjective	Adjective, comparative
JJS	adjective	Adjective, superlative
ADVP	adjective	Adverb phrases
RD	adjective	Adverb

3.3 Sentiment Analysis

Sentiment analysis can be known as the process of defining the user's opinion from the user comments or feedbacks. Among the three types of sentiment analysis that are already known as sentence level, document level and aspect level, the proposed system intends to make the aspect level sentiment analysis. In order to get the opinion score for the words in two lists: *vlist* and *alist,* the lexical resource S*entiWordNet* can help to define the score for each word in these two lists.

After defining the score for each word with the help of lexical resource, the proposed system finds the total score for the entire sentence by adding the score of each word. In this case, positive values of resulted scores represent the good opinion. And also negative values of resulted scores represent bad opinion and zeros mean neutral opinion. For accessing the scores for each word by confirming with *SentiWordNet* lexical resource, the algorithm known as *SWN3* can support which can be seen in the following figure.

Algorithm: **SWN3.** Define score for each word
Input : list of words which has verb form, *vlist*
: list of words which has adjective form, *alist*
: sentiWordNet lexical resources, *swn*
Output : list of words with score, *scorelist*
: score of the whole sentence, *totalScore*
Begin
(1) create *dictionary* {word, score} from *swn*
(2) *totalScore* ← 0
(3) If *vlist* is not null
(4) Foreach *vword* of *vlist*
(5) *score*← dictionary.get(*vword* +"#"+ "v")
(6) If *score* is not null
(7) *totalScore* ← *totalScore* + *score*
(8) *scoreList.*add(*vword*+":"+*score*)
(9) Endif
(10) Endfor
(11) If *alist* is not null
(12) Foreach *aword* of *alist*
(13) *score*← dictionary.get(*aword* +"#"+ "a")
(14) If *score* is not null
(15) *totalScore* ← *totalScore* + *score*
(16) *scorelist.*add(*aword*+":"+*score*)
(17) Endif
(18) Endfor
(19) Return collection of *scorelist*, *totalScore*
End

Figure 3.4 SWN3 Algorithm

3.4 Aspect Detection with Domain Ontology

Aspect- based sentiment analysis has two parts of process which are sentiment analysis and aspect detection process. The work flow of sentiment analysis is already explained in previous section. In this section, aspect detection process is highlighted which helps in finding the targeted word from this sentence. So, the opinion or sentiment result of the entire sentence refers on one of words in the sentence which means that the emotion of users upon related aspect word. In order to confirm the extracted words whether it is aspect or not, domain specific ontology is needed to prepare. This domain specific ontology collects and composes specific things that are aspect words for the UCST campus.

Ontology composed with categories representing concepts and relationships between concepts. It can also clarify the complexity in structure of domain knowledge and transform this knowledge into concept-class, object-properties (relationships between class and subclass), data-properties (attributes and their types), instances and data values. At the preprocessing stage of this proposed system, it creates the ontology with OWL format using protégé editor. The concept class structure of domain specific ontology used in this system is shown in the following figure.

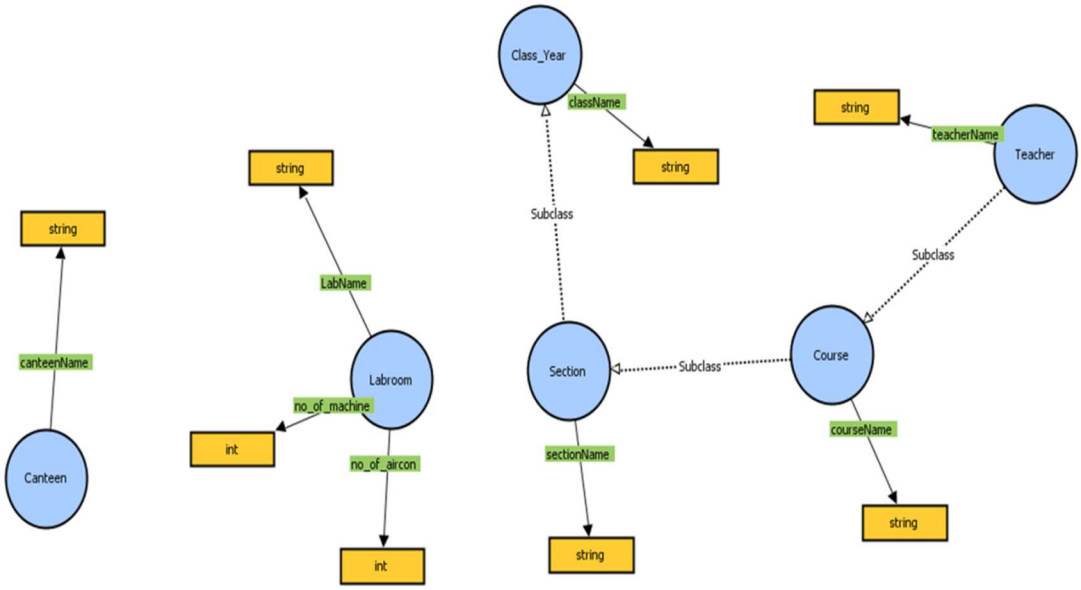


Figure 3.5 Domain Specific Ontology for UCS(Taungnoo)

The feedback sentence's nature from university's student feedback system differs from other feedback statement format. For the feedback sentence such as "Teacher who teaches CST-2112 is excellent at teaching", in this sentence, three noun forms of words s are included such as:

nouns_list: [Teacher, CST-2112, teaching]

Among these three, the two noun forms that are teacher and CST-2112 can be defined as aspect words by confirming with domain specific ontology. Among these two aspects teacher and CST-2112, it needs to choose which one is real aspect. The concept hierarchy of ontology can help to find real aspect. The class name of teacher and CST-2112 is teacher and course, respectively. The word which is included in inner most sub-class is

the most suitable for targeted aspect words. The number of aspects organized in this ontology is generally refers to the number of class nodes in this ontology such as teacher, course, canteen, lab room, class year and section. Moreover, the values under these classes are also included. For which, it is not possible to define these counts because the number of individual teachers depends on the count of teachers assigned in this university and so on.

3.5 Data Visualization

Data visualization is the graphical representation for important data or information. There are many forms of visual elements such as charts, graphs, maps and so on. Among the many forms of data visualization, this system used the bar graph or bar chart to present the sentiment result count for the corresponding aspect word. In this system, the bar graph presents aspect word as categorical data with rectangular bars. These bars' heights or lengths are proportional to the values of sentiment result count for each aspect word. The bars can be plotted vertically or horizontally but this system can only present with vertical format which can be shown in the following figure.

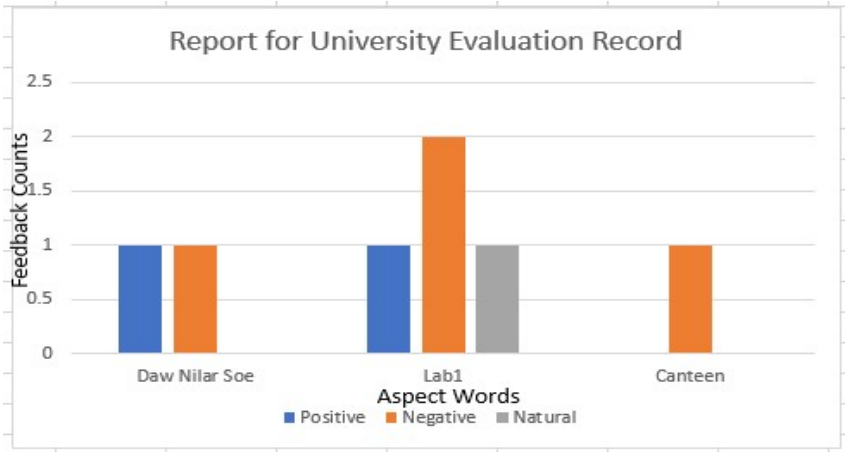


Figure 3.6 Bar graph of sentiment score of each aspect word

For this bar graph, three sets of data presentation, as shown in the above figure, (are necessary) because there are positive, negative and neutral values for related aspect. By seeing this bar graph, the user can easily understand the result of feedback analysis.

3.6 Accuracy Evaluation

As described in Chapter 2, the system in this thesis finds the accuracy index by using the confusion matrix. In this session, how to calculate the accuracy value for the sample set of real data will be presented.

Table 3.2 Sample data of history records for feedback evaluation

Id	Feedbacks	Aspect Words	Opinion Result	Real Opinion
118	CS-102 teacher is good at teaching and she is so kind to her pupils.	Daw Nilar Soe	positive	positive
119	UCSTaungoo canteens are good for students.	canteen	positive	positive
120	In Lab1, machines are so old and they cannot operate well.	Lab1	negative	negative
121	CS-102 teacher is not a well-prepared teacher, we cannot understand her lecture.	Daw Nilar Soe	negative	negative
122	the machines in Lab1 are so bad.	Lab1	negative	negative
123	air conditioning in lab1 is so cool.	Lab1	positive	positive
124	the machines in Lab1 are suitable for database programming.	Lab1	neutral	positive

In this table, it can be seen that there are three kinds of data values such as positive, negative and neutral. But the confusion matrix can handle binary data and so, this system assumes the natural value as the positive values. And then the system counts the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for the whole data set. That can be seen as follows:

- TP means that real opinion ‘positive’ can be defined by the system as ‘positive’ and which count is 3
- TN means that real opinion ‘negative’ can be defined by the system as ‘negative’ and which count is 4

- FP means that real opinion ‘negative’ can be defined by the system as ‘positive’ and which count is 0
- FN means that real opinion ‘positive’ can be defined by the system as ‘negative’ and which count is 0

After that, the system continues to find the true positive rate (TPR), true negative rate (TNR), false positive rate (FPR) and false negative rate (FNR) base on the count of the above which can be seen as follows:

- True positive rate (TPR) : TP/(TP + FN) = 3 / (3 + 0) = 1
- True negative rate (TNR) : TN/(TN + FP) = 4 / (4 + 0) = 1
- False positive rate (FPR) : FP/(FP + TN) = 0 / (0 + 4) = 0
- False negative rate (FNR) : FN/(FN + TP) = 0 / (0 + 3) = 0

Finally, the accuracy for this sample set can be calculated by using the equation 2.2 that was described in section 2.6. The calculation can be seen as follows:

$$\text{Accuracy} = (1 + 1) / (1 + 1 + 0 + 0) = 1.0$$

Based on this result, it can be concluded that the system in this thesis has 100 % of accuracy for this sample data set. In the next chapter, the system implementation result will be shown with attractive screen shots.

CHAPTER 4
SYSTEM IMPLEMENTATION

The system, named as Domain Oriented Aspect Detection, is created to analyze the student feedbacks in order to assist the admin team when evaluating the current status of their university campus. In this chapter, the implementation of this system is presented with step by step processes. (To improve the understanding of the user) when presenting this system, this chapter collects and shows the results with full screen shots. After reviewing all sections in this chapter, the user will be able to know how the system is setup and how to process the analysis task.

In this chapter, how to set up the Domain Oriented Aspect Detection System is firstly described with minimum requirements for this system. And then feedback selection process is presented and the Part of Speech (POS) Tagging results (is shown) for those selected feedbacks. Next the sentiment scores for action words of each feedback that is obtained by Confirming with SentiWordNet lexical resource (are described.). And then aspect word definition is and finally, the experiment result on analyzing sentiment outcome is presented.

4.1 How to Setup the System

In order to implement the Domain Oriented Aspect Detection, a *Mysql database server 5.0.12 or MariaDB database server 10.4.6* is needed to already install and create a database named as *uni_evaluation_records* to store the students' feedbacks data. All the tables from the data source are also needed to import into these database because they are original feedback data table, history records table that stored feedbacks with sentiment result, the appropriate data table for creating bar chart and the table maintain the admin information. The ontology editor software *Protégé 3.4.7* is also needed to install to be able to edit the aspect data collection resources. The configuration is as follows:

- Hardware configurations
    - CPU: at least Core i3 M 560 @ 2.67 GHz
    - Memory (RAM): at least 4 GB

- Hard disk: at least85GB

- Software requirements
  - Window 7 64bit and above
  - JDK 1.8 and above
  - Apache-tomcat 8.5.39
  - Eclipse-jee-mars

4.2 Feedback Selection

The domain oriented aspect detection system is intended to assist the management level in one university for evaluating the student feedbacks and making decision based on these feedbacks. In order to start the feedback analysis task, user or admin needs to login by using the username and password via the following page.

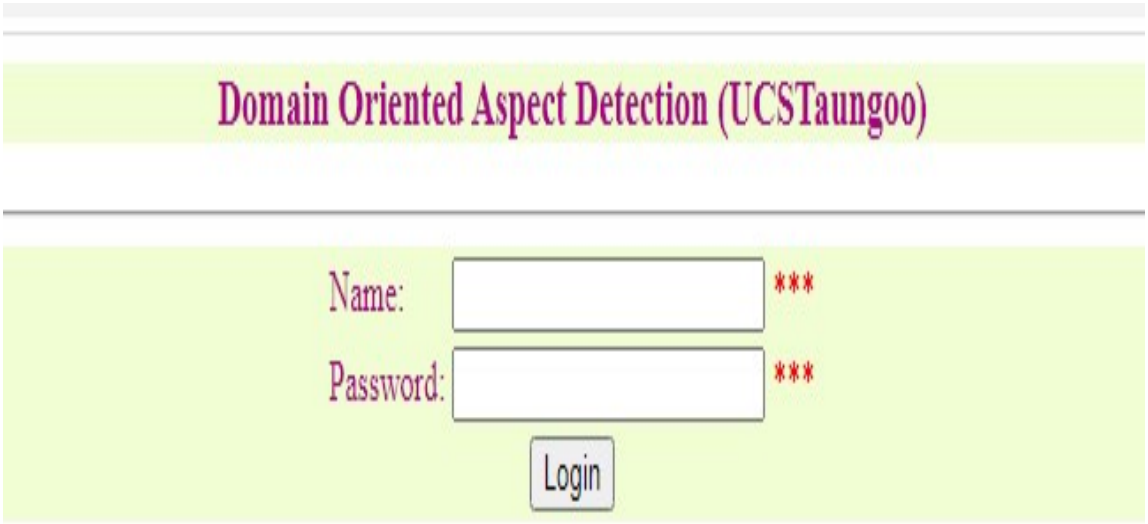


Figure 4.1 The front page of domain oriented aspect detection system

After typing the username and password, click *Login* button that is shown in the following figure. These data *username* and *password* are taken into search in admin table. When the match data is found in admin table, the user or admin can pass through this page.

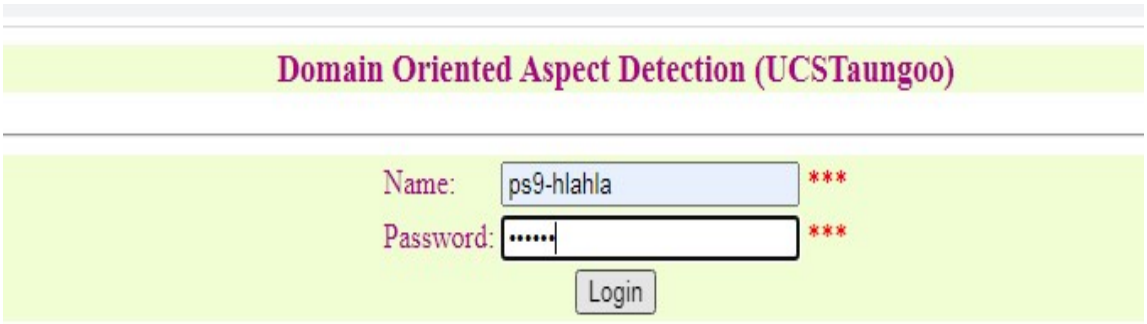


Figure 4.2 User login page

And then the home page of Domain Oriented Aspect Detection system can be seen as shown in following figure. Here, the students' feedbacks records stored in feedbacks table maintain the feedbacks with their submission dates semester by semester.

So, the user or admin needs to select the date in order to obtain the required information.

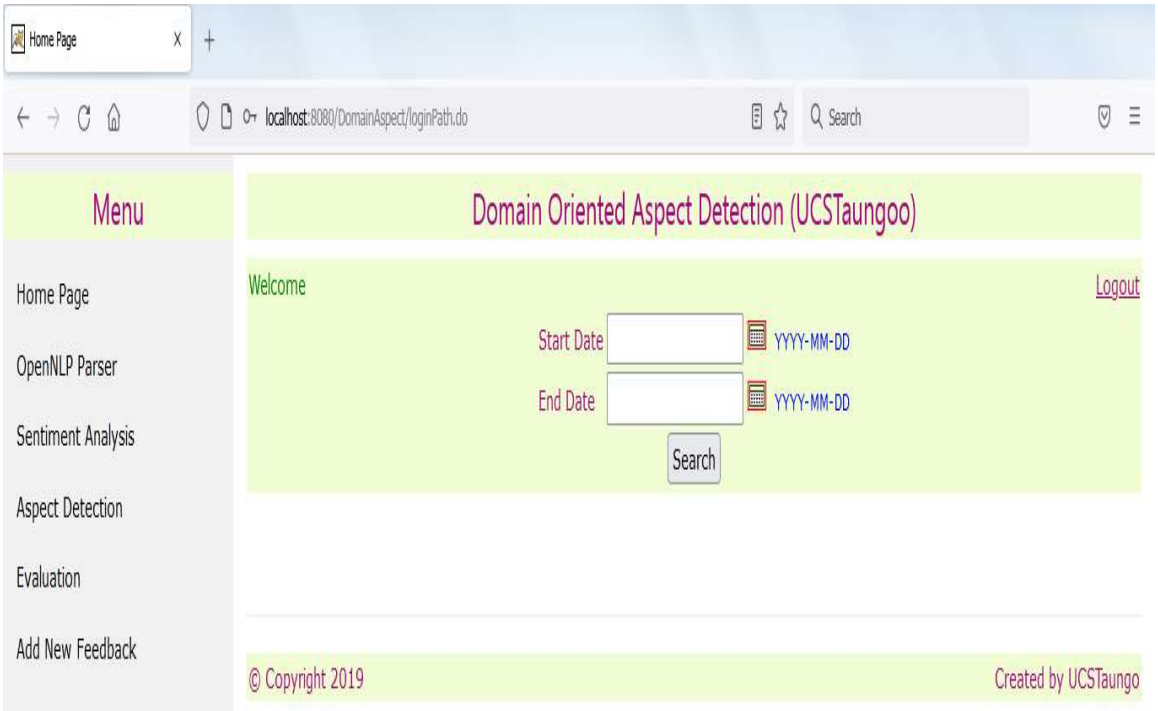


Figure 4.3 Home page of Domain Oriented Aspect Detection System

After selecting the start date or end date of required semester, this home page shows the selected feedbacks with 10 records in one page. If the user or admin wants to see the next record, he or she can press the next button and also *previous* button to go back the first page. These are shown in the following figures.

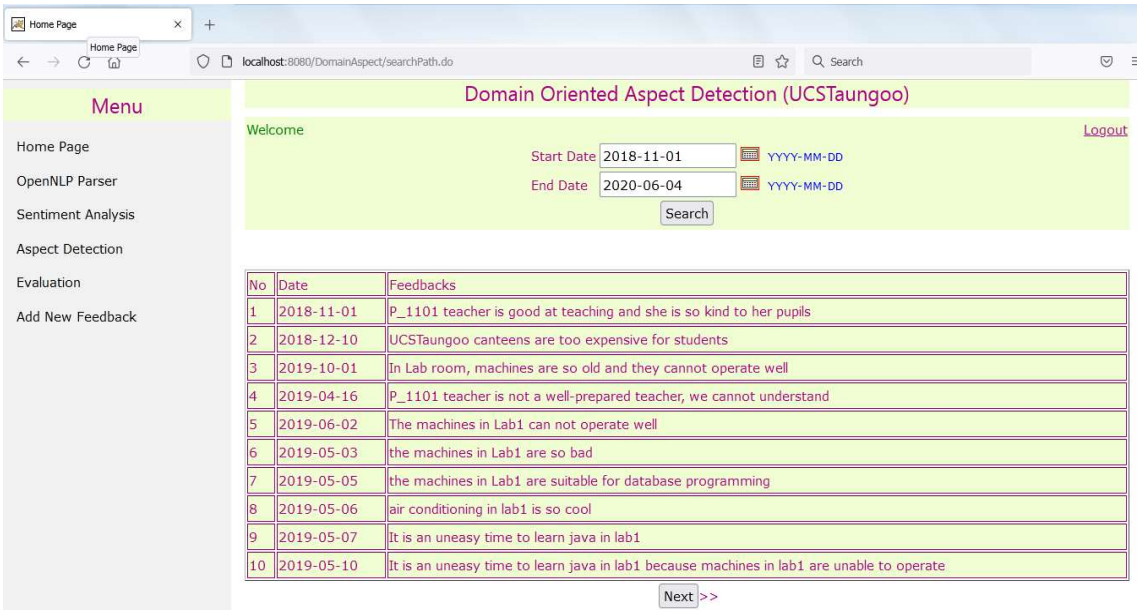


Figure 4.4 Feedbacks selection in home page

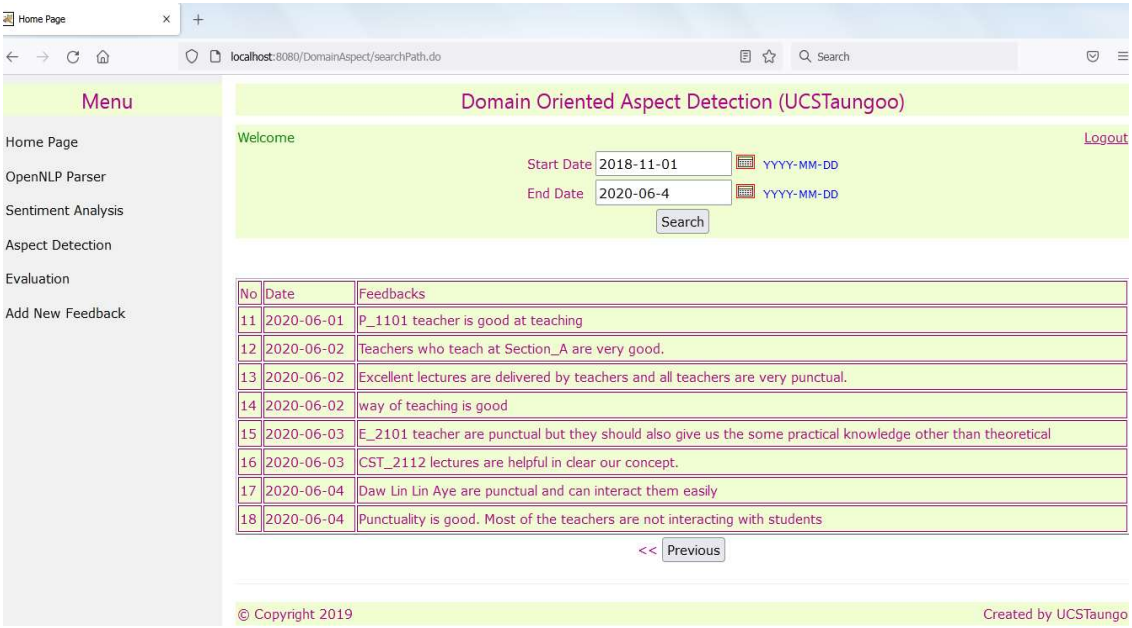


Figure 4.5 Feedbacks selection next page segment

After selecting the start date and end date for desired semester, the user needs to go the *OpenNLP parser* link to parse each of the feedback sentences.

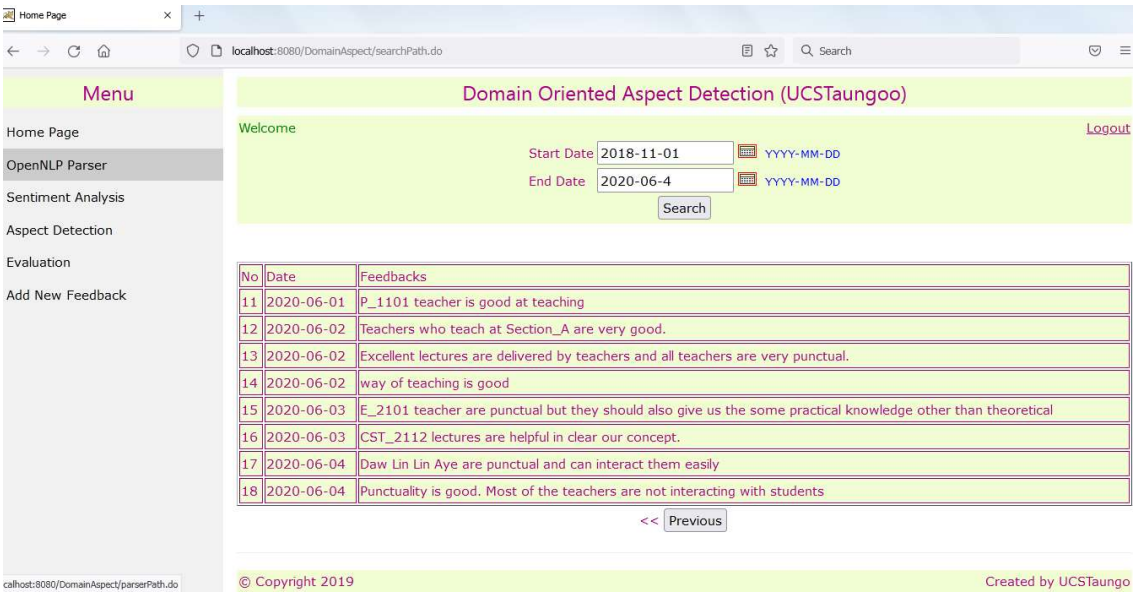


Figure 4.6 *OpenNLP* parser link

4.3 Part of Speech (POS) Tagging

As we already know that, the *OpenNLP* parser parses the sentence into *nouns* group, *adjectives* group and *verbs* group for this analysis. These groups of data are extracted base on most commonly used POS tags that are described in previous chapter. Domain Oriented Aspect Detection system create the *array_list* to store these data group and their related feedbacks. This task is done by using *en-parser-chunking.bin* file and applying parserAction algorithm.

And then this *array_list* is sent to the *parserResult* page to show the result of part_of_speech tagging process. This page shows 7 records in one page and when the user or admin want to see next data records, he or she can press the *next* button and also *previous* button to go back the first page of data records. These results are shown in the following figures.

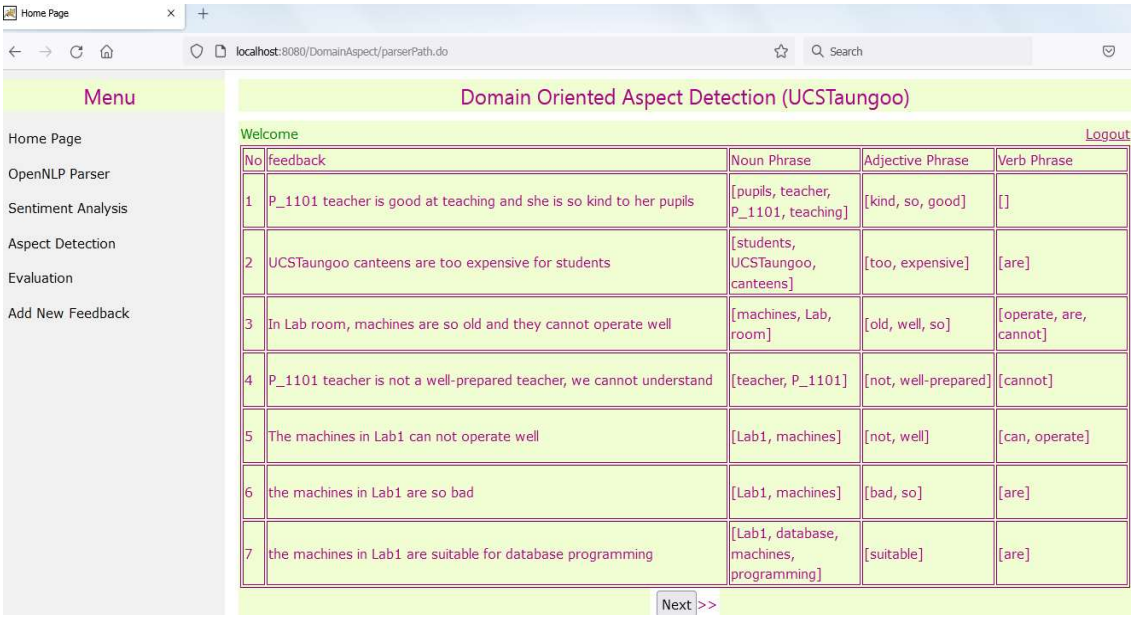


Figure 4.7 The results page of *OpenNLP* parser

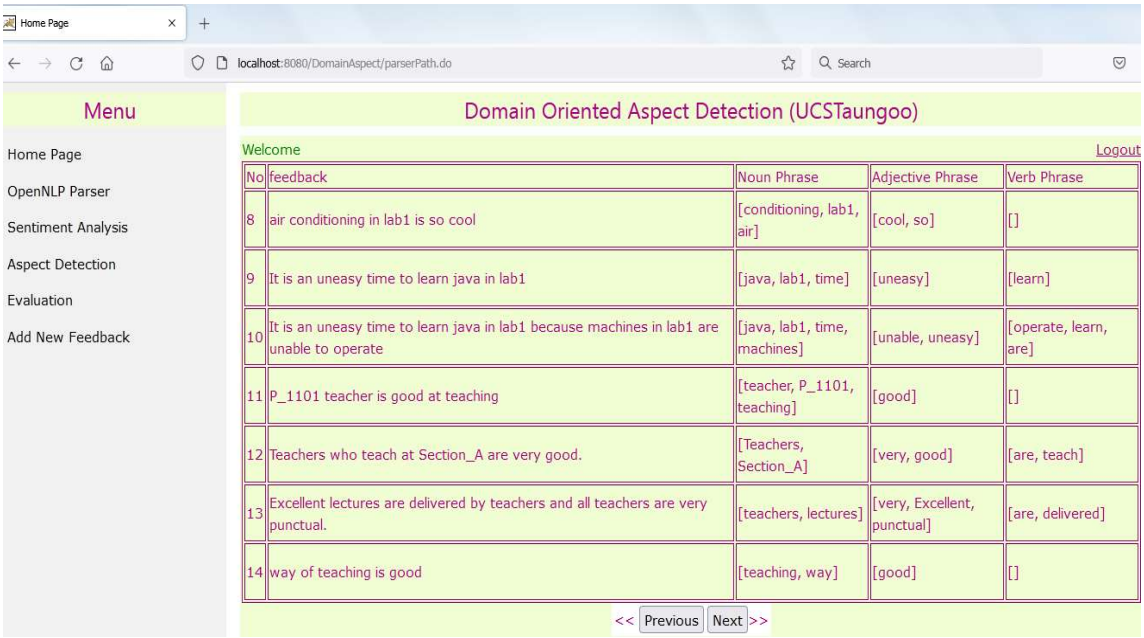


Figure 4.8 The next page segment of *OpenNLP* parser

After taking the part-of-speech tagging process, the user needs to press the *Sentiment Analysis* link to send the *adjectives* and *verbs* group of sentence for finding their related sentiment scores. This link is shown in the following figure.

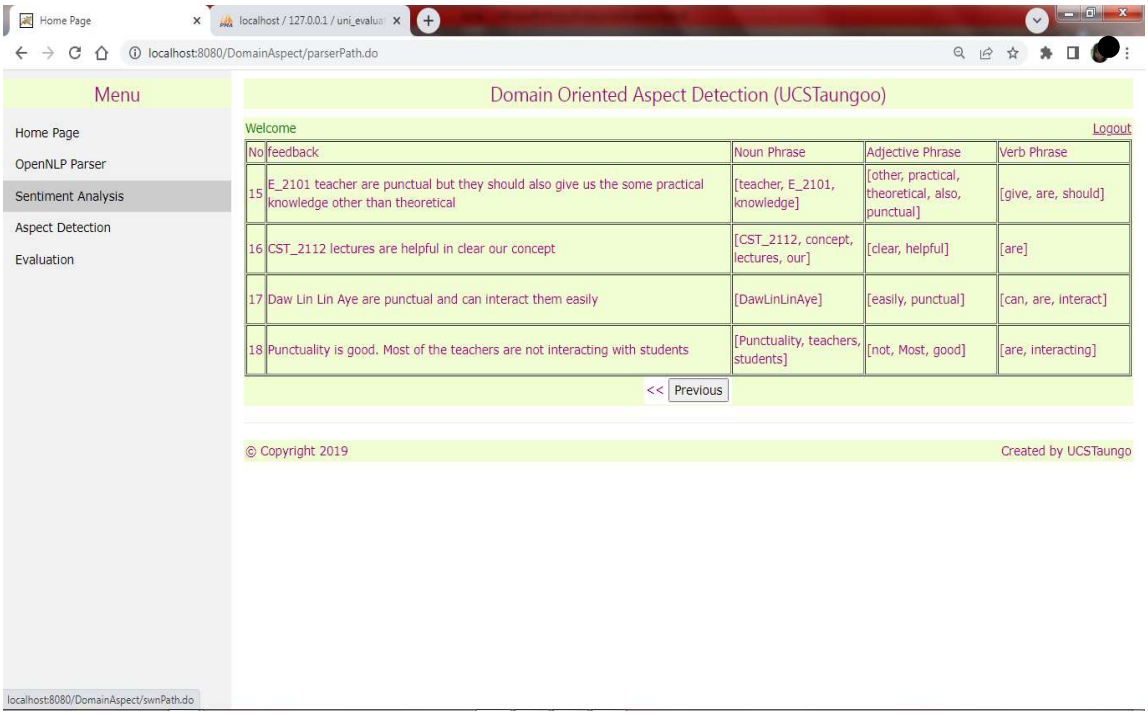


Figure 4.9 Sentiment analysis link

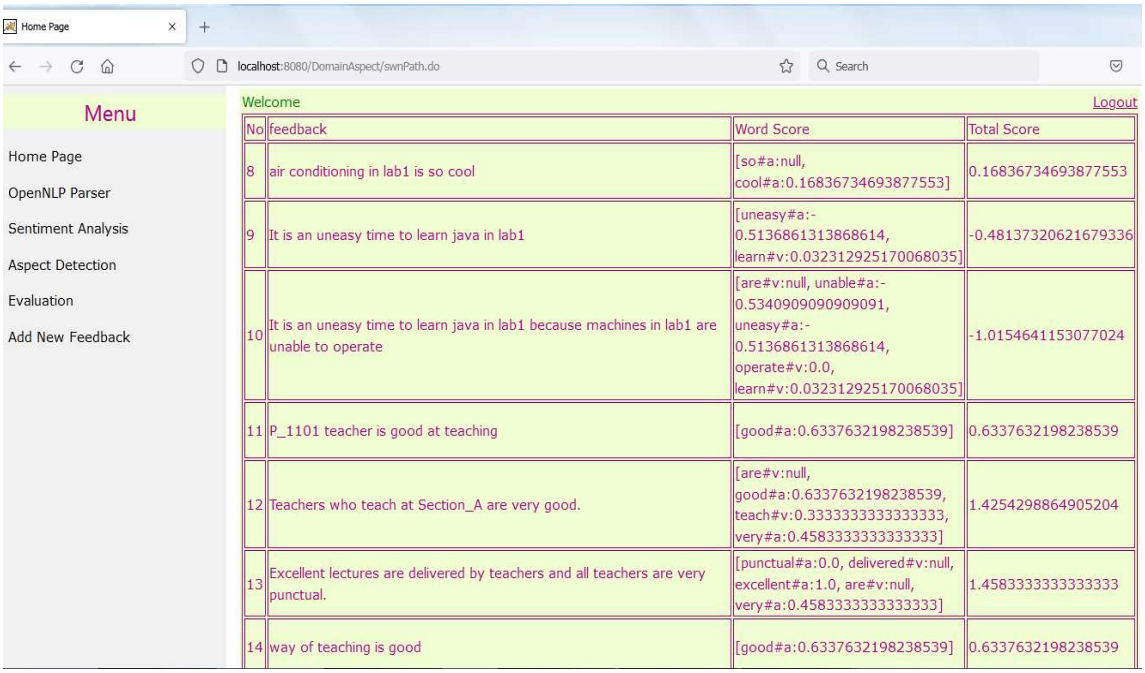
4.4 Confirmation with *SentiWordNet*

This stage takes the *adjectives* and *verbs* group and finds each word in *sentiWordNet* lexical resources to define the sentiment score for each word. This task is done by SWN3 Algorithm and then total score for each sentence is found by summing of all sentiment scores. This stage the firstly create the *array_list* to store the feedback, the scores of each word in this feedback and its related total score. And this stage also defines the sentiment result for this feedback as *neutral* if total score is zero, *positive* if total score is greater than zero and *negative* if total score is less than zero. Finally, this modified records are save as history records into *opinionResult* table of *uni-evaluation-record* database.

The result obtain from this stage is sent to *swnResult* page to show the score result before modifying into *neutral*, *positive* and *negative*. This page shows 7 records per page and if user or admin wants to see the *next* data, he or she can press the next button and also previous to go back the first page of records. These results are shown in the following figures.

No	feedback	Word Score	Total Score
1	P_1101 teacher is good at teaching and she is so kind to her pupils	[so#a:null, good#a:0.6337632198238539, kind#a:0.6477272727272727]	1.2814904925511266
2	UCSTaungoo canteens are too expensive for students	[expensive#a:-0.5, too#a:null, are#v:null]	-0.5
3	In Lab room, machines are so old and they cannot operate well	[so#a:null, are#v:null, cannot#v:-0.75, old#a:-0.17148488830486208, operate#v:0.0, well#a:0.6931818181818181]	-0.22830307012304396
4	P_1101 teacher is not a well-prepared teacher, we cannot understand	[well-prepared#a:0.625, cannot#v:-0.75, not#a:-0.75]	-0.875
5	The machines in Lab1 can not operate well	[can#v:0.0, operate#v:0.0, not#a:-0.75, well#a:0.6931818181818181]	-0.05681818181818188
6	the machines in Lab1 are so bad	[so#a:null, are#v:null, bad#a:-0.5706406664316871]	-0.5706406664316871
7	the machines in Lab1 are suitable for database programming	[are#v:null, suitable#a:0.25]	0.25

Figure 4.10 Sentiment scores for feedback sentences



No	feedback	Word Score	Total Score
8	air conditioning in lab1 is so cool	[so#a:null, cool#a:0.16836734693877553]	0.16836734693877553
9	It is an uneasy time to learn java in lab1	[uneasy#a:-0.5136861313868614, learn#v:0.032312925170068035]	-0.48137320621679336
10	It is an uneasy time to learn java in lab1 because machines in lab1 are unable to operate	[are#v:null, unable#a:-0.5340909090909091, uneasy#a:-0.5136861313868614, operate#v:0.0, learn#v:0.032312925170068035]	-1.0154641153077024
11	P_1101 teacher is good at teaching	[good#a:0.6337632198238539]	0.6337632198238539
12	Teachers who teach at Section_A are very good.	[are#v:null, good#a:0.6337632198238539, teach#v:0.3333333333333333, very#a:0.4583333333333333]	1.4254298864905204
13	Excellent lectures are delivered by teachers and all teachers are very punctual.	[punctual#a:0.0, delivered#v:null, excellent#a:1.0, are#v:null, very#a:0.4583333333333333]	1.4583333333333333
14	way of teaching is good	[good#a:0.6337632198238539]	0.6337632198238539

Figure 4.11 Next page segment of sentiment scores for feedback sentences

In OpenNLP parser action stage that is described in section 4.3, the noun group is also extracted. This noun group is needed to send to aspect detection process and the user or admin clicks the Aspect Detection link shown in the following figure to do this action.

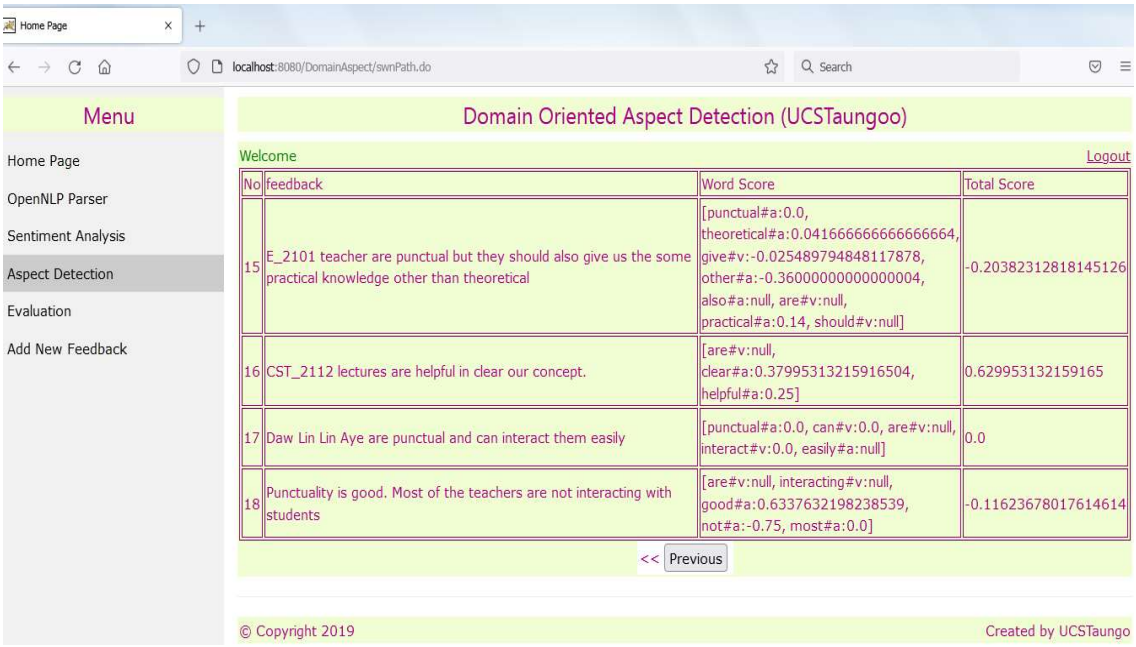


Figure 4.12 Aspect detection link

4.5 Aspect Word Definition

Aspect word definition stage defines the noun forms of words sent from parser action stage as main aspect words that represent the key feature for a feedback sentence. In this process, noun forms of words are confirmed with the domain oriented ontology to define whether these are aspect words or not. This confirmation process uses jaro similarity methods to match the noun and words on ontology. Here, more than one noun may be contained in the data that are pre-organized in ontology. In this case, the ontology greatly provides to determine which words is the most suitable for the required aspect word. Moreover, this system can dig the related value with the help of ontology if the feedback does not contain the main component.

After extracting the original aspect word from the feedback sentence and obtain the relate value of this aspect word, the system saves this words into a matched row of *opinionResult* table. So, the *opinionResult* table can collect the feedbacks and their related aspect word, related value of this aspect word and real opinion data and that is sent to *aspectDetection* page to show the result. This page shows 7 records per page and if the user or admin wants to see the next data record, he or she can press the next button and also can

press the previous button if the user wants go back to the first page. This is shown in the following figures.

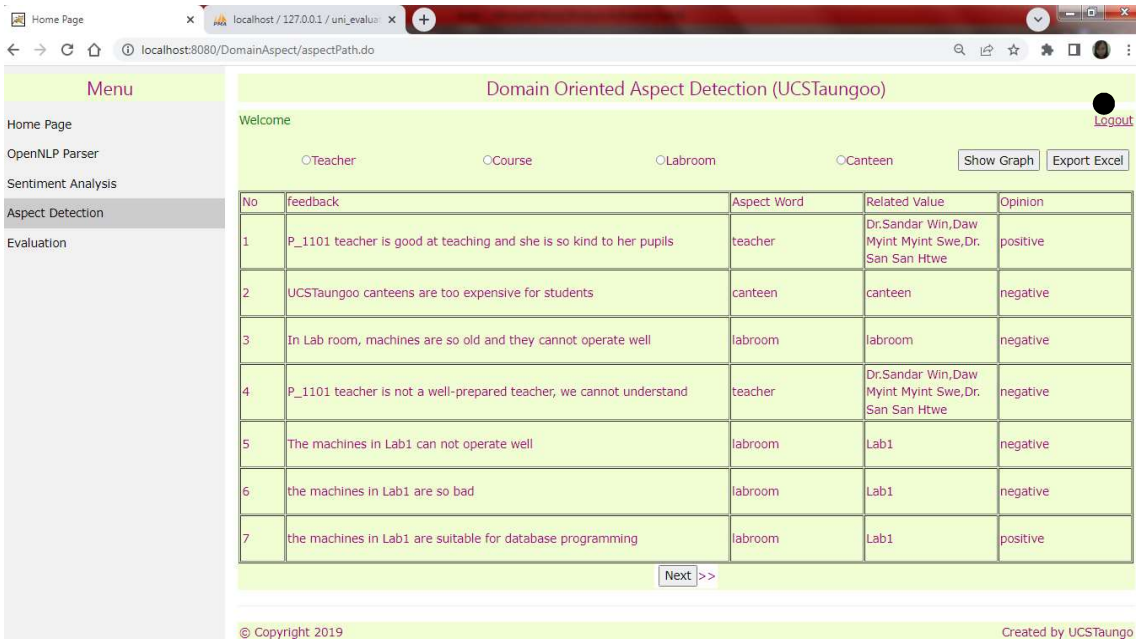


Figure 4.13 Aspects detection from given feedbacks

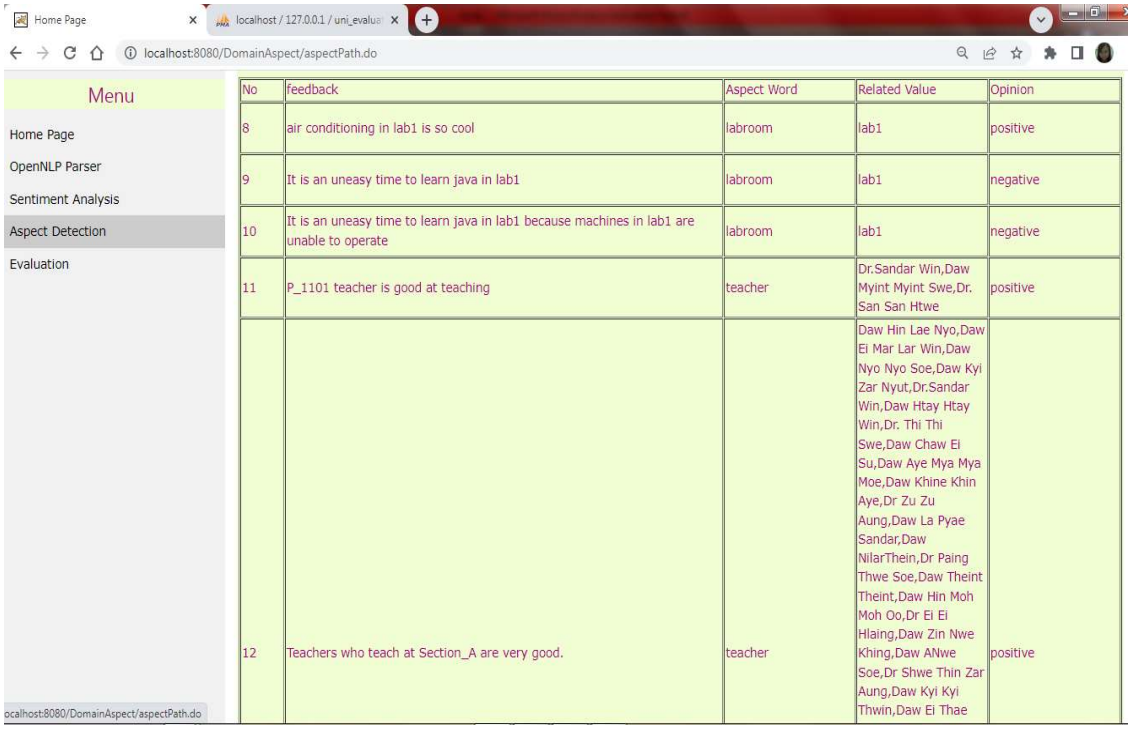


Figure 4.14 Next page segment of Aspects detection from given feedbacks

And then, the user or admin can see the data with visualize form and excel sheet as they desired. In order to see the graph or excel sheet, it is needed to select what kind of aspect because there are various *relateValues* for each kind of aspect word and the graph and excel sheet will be created based on these *relateValues*. To avoid the sophisticated illustration, four main kinds of key aspects are predefined.

When the aspect word *Teacher is selected* and press the show graph button that are shown as the following figure, the system shows bar graph that represents only data for teacher aspect.

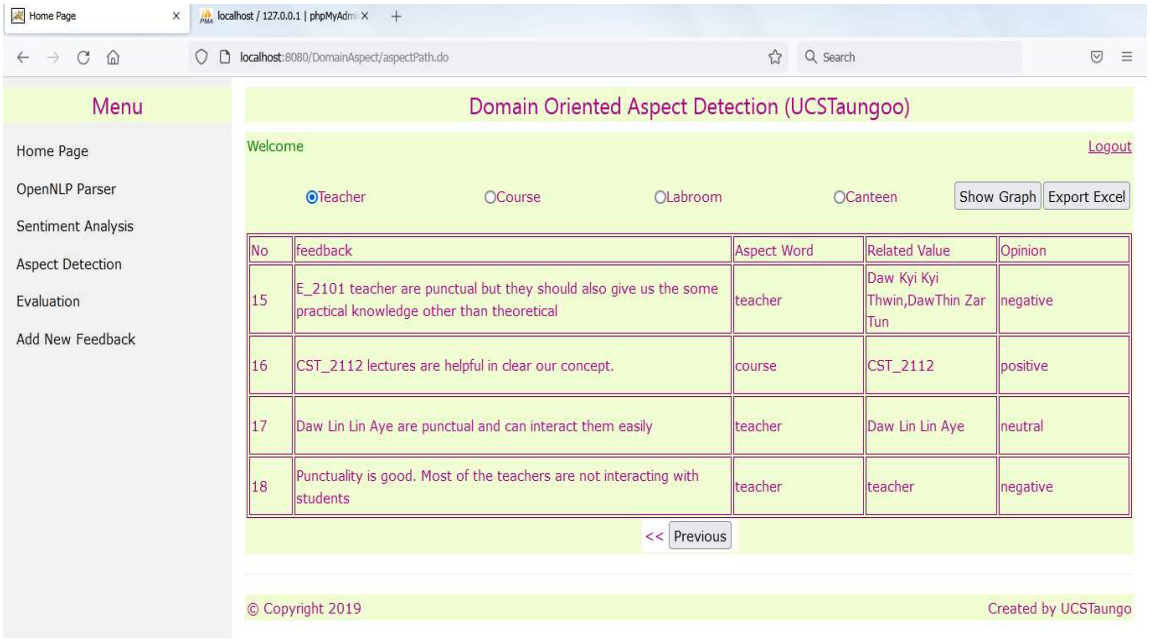


Figure 4.15 Teacher aspect category selections for visualization

For creating the bar graph, this system chooses the vertical format of the chart style. It takes *relateValues* as categories that are shown on X-axis of the chart and the height of the each bin represents the number of feedbacks. The three sets of data representation in this chart such as green color bin, red color bin and blue color bin represent positive sentiment count, negative sentiment count and natural sentiment count respectively. If the user or admin presses the *Export Excel* button, the system downloads this teacher related data as an excel file. These are shown in the following figures.

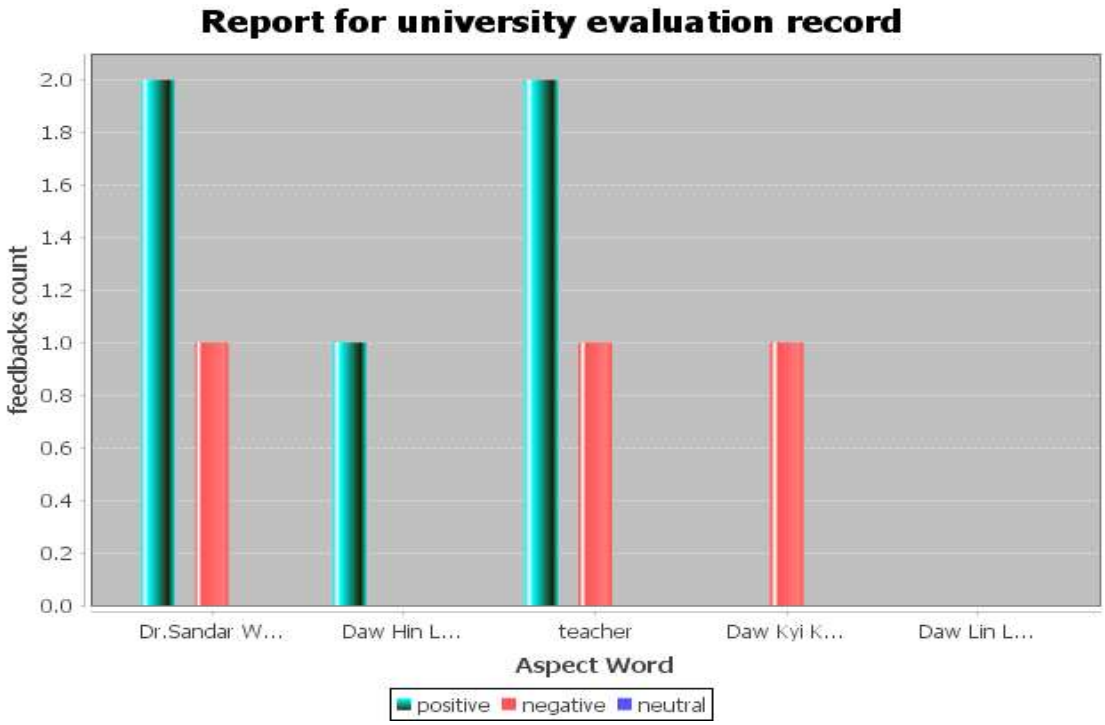
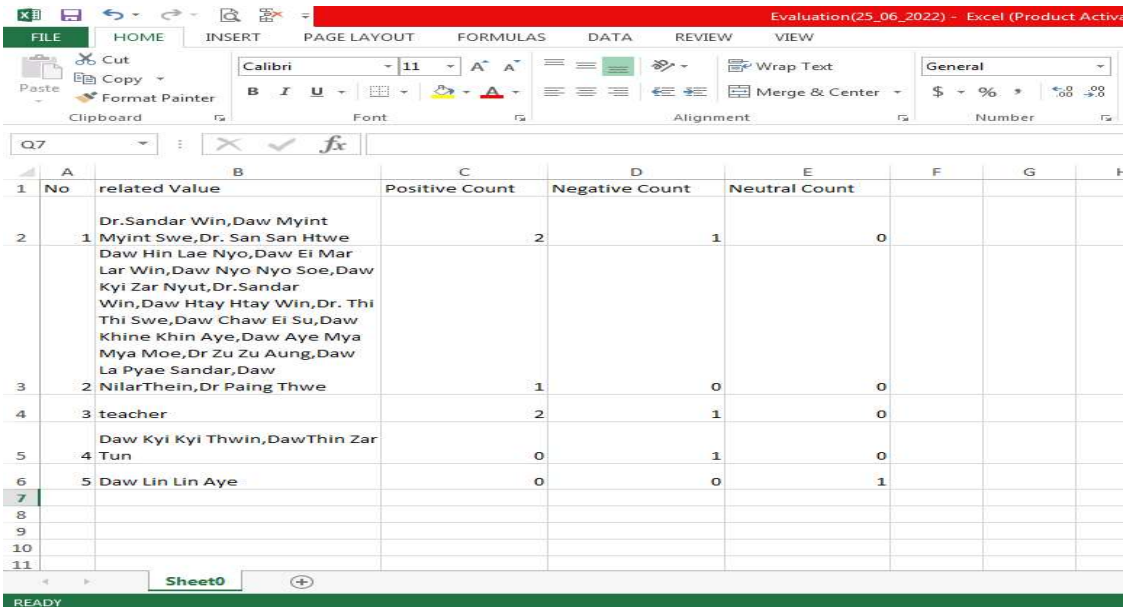


Figure 4.16 Teacher related data visualization



	A	B	C	D	E	F	G
1	No	related Value	Positive Count	Negative Count	Neutral Count		
2	1	Dr.Sandar Win,Daw Myint Myint Swe,Dr. San San Htwe	2	1	0		
3	2	Daw Hin Lae Nyo,Daw Ei Mar Lar Win,Daw Nyo Nyo Soe,Daw Kyi Zar Nyut,Dr.Sandar Win,Daw Htay Htay Win,Dr. Thi Thi Swe,Daw Chaw Ei Su,Daw Khine Khin Aye,Daw Aye Mya Mya Moe,Dr Zu Zu Aung,Daw La Pyae Sandar,Daw NilarThein,Dr Paing Thwe	1	0	0		
4	3	teacher	2	1	0		
5	4	Daw Kyi Kyi Thwin,DawThin Zar Tun	0	1	0		
6	5	Daw Lin Lin Aye	0	0	1		
7							
8							
9							
10							
11							

Figure 4.17 Export excel file for teacher related data

When user or admin chooses the course can press Show Graph button, this system shows the graph with course related data. And if he or she presses the *Export Excel* button,

this system downloads the excel file for course related data. These are shown in the following figures.

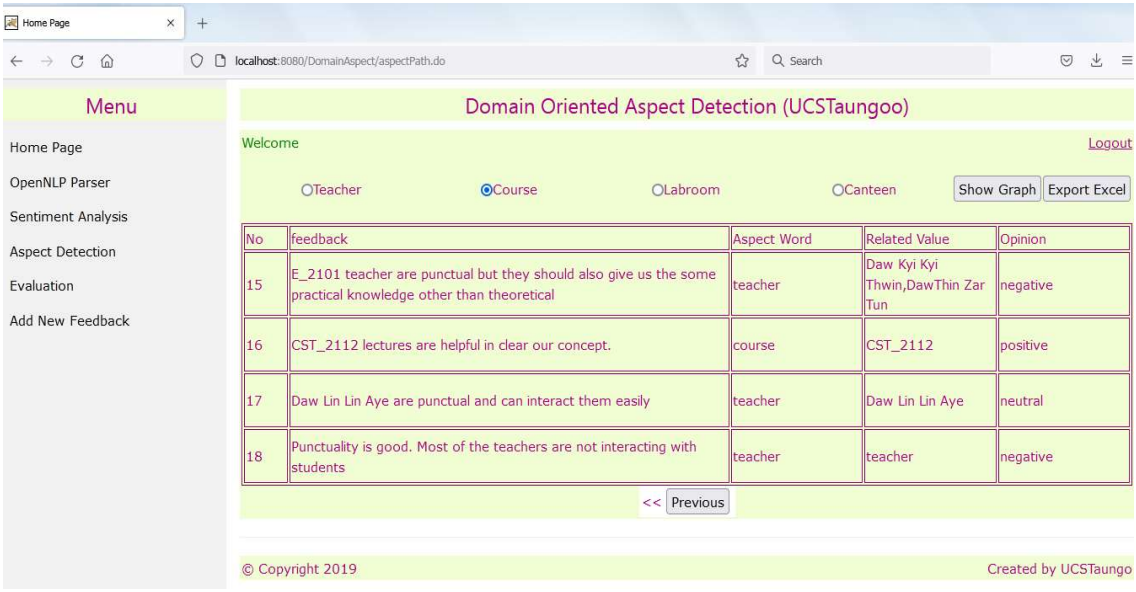


Figure 4.18 Course aspect category selections for visualization

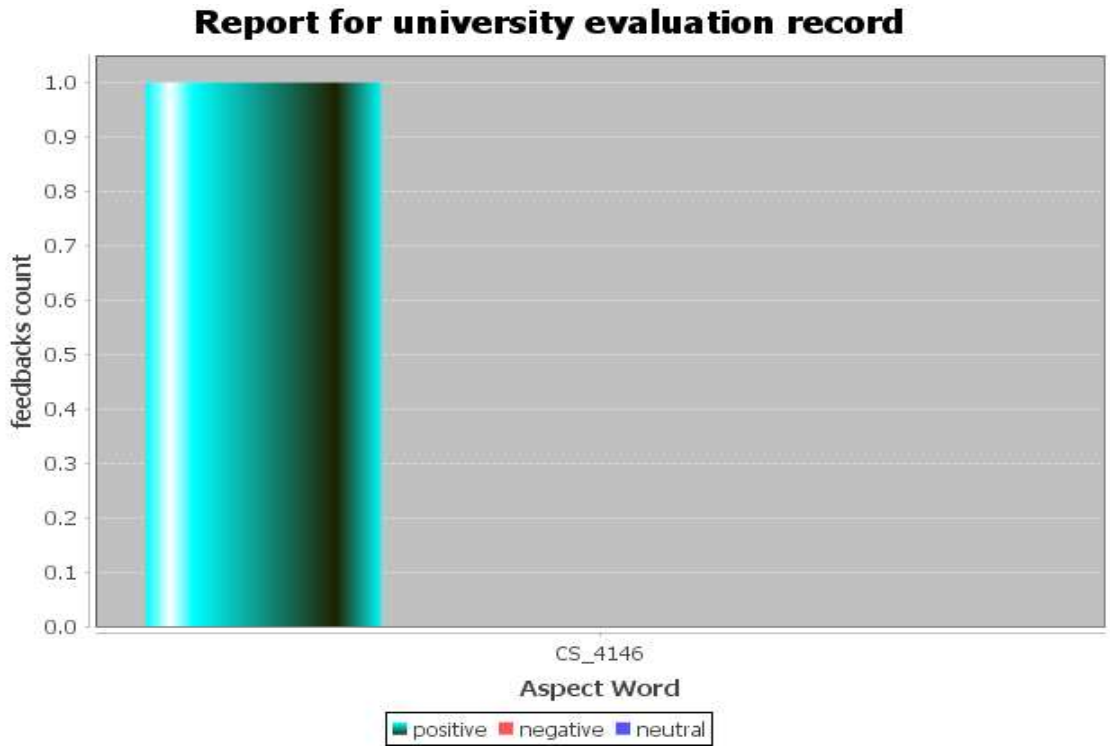
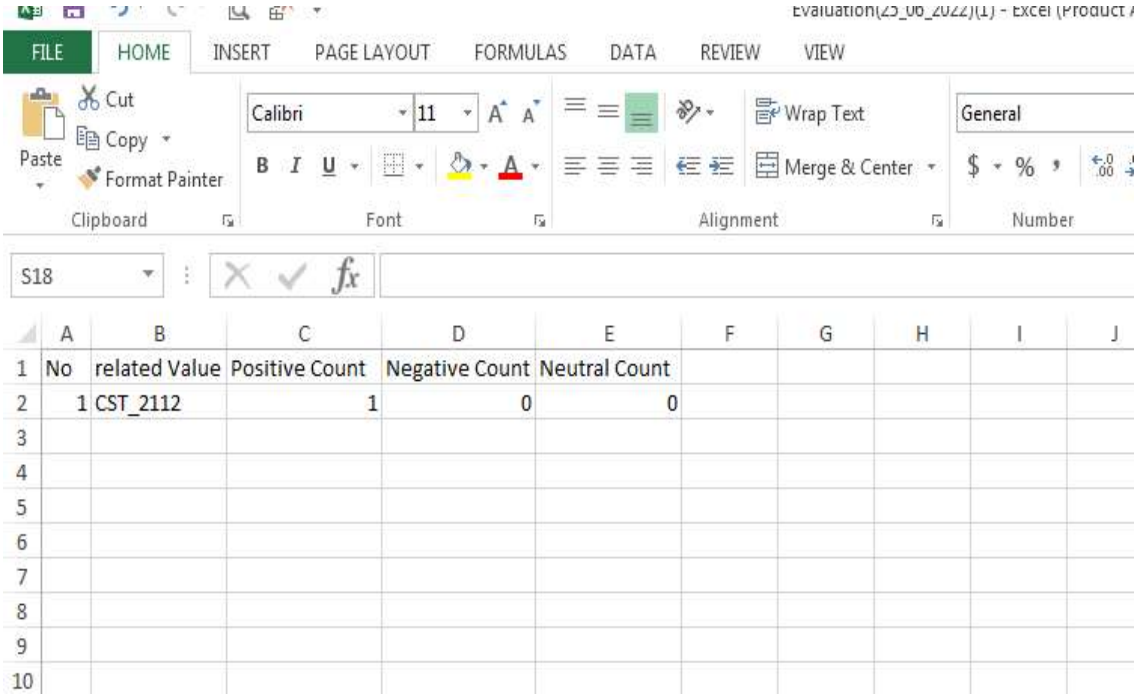


Figure 4.19 Course related data visualization



No	related Value	Positive Count	Negative Count	Neutral Count
1	CST_2112	1	0	0

Figure 4.20 Export excel file for course related data

The following figures show the *Labroom* related data selection, graph presentation and excel file export.

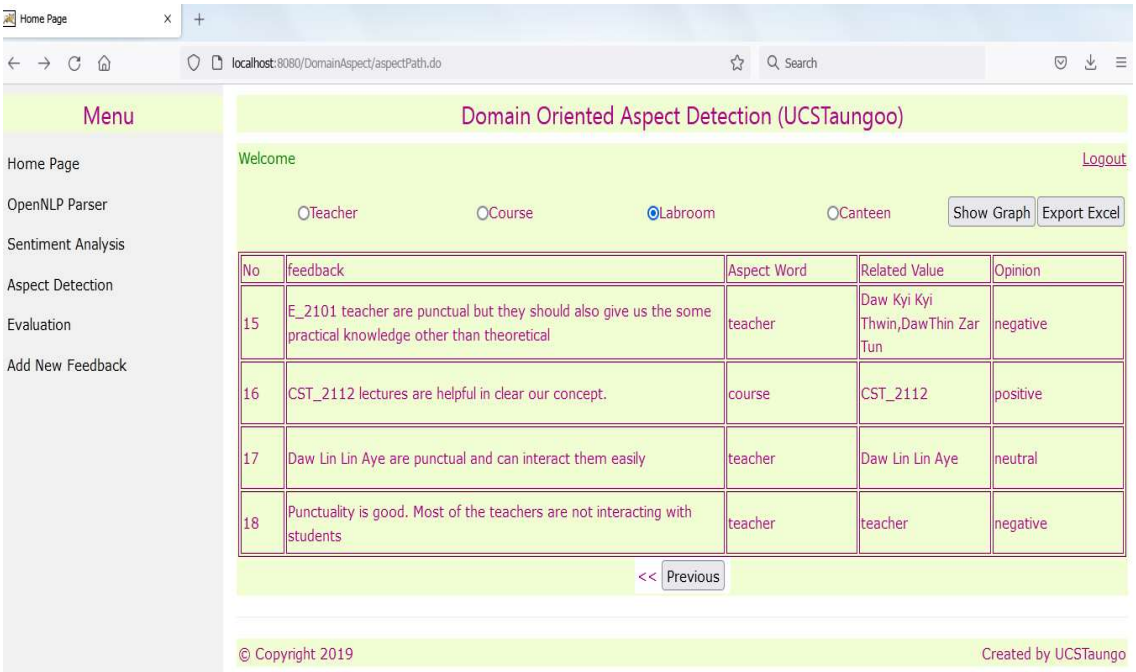


Figure 4.21 Labroom aspect category selections for visualization

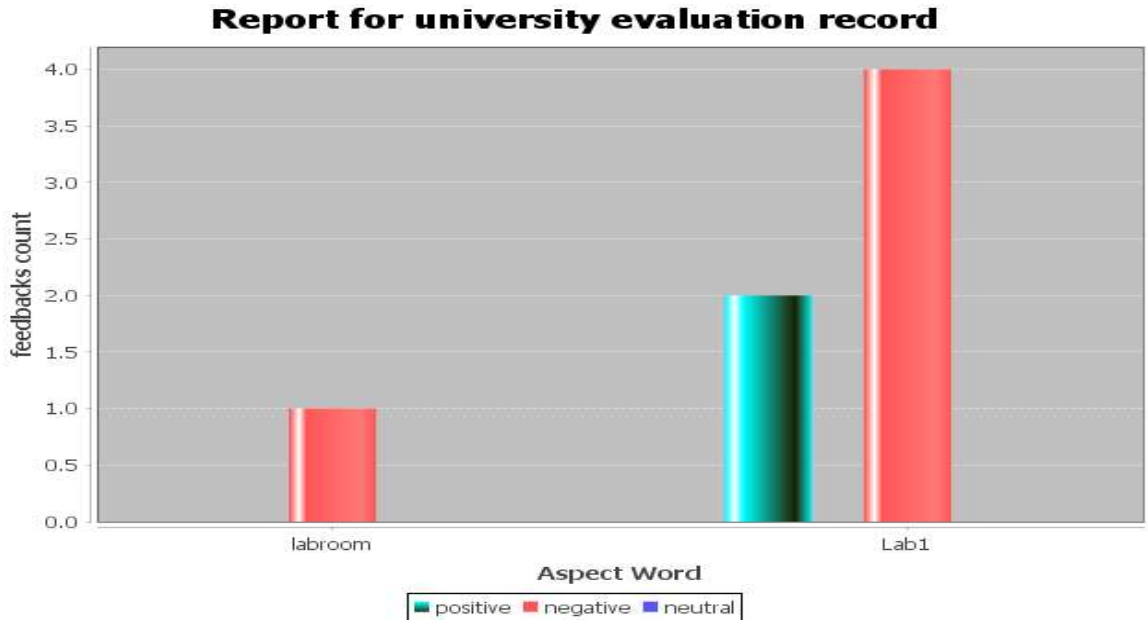
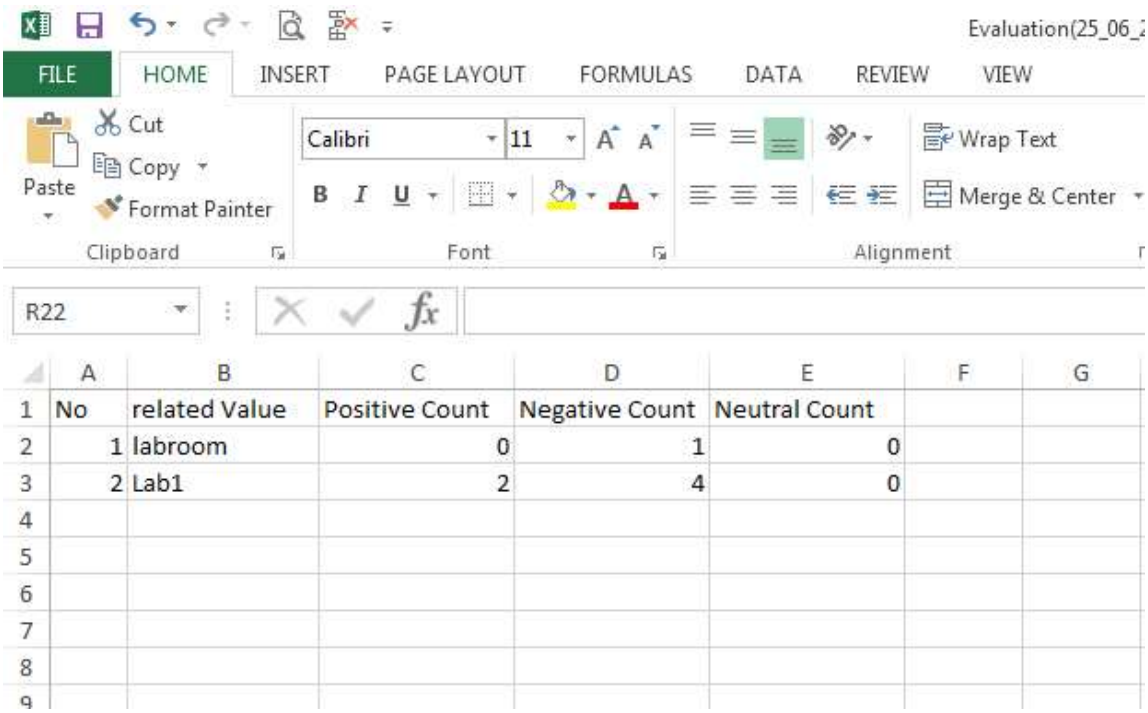


Figure 4.22 Labroom related data visualization



	A	B	C	D	E	F	G
1	No	related Value	Positive Count	Negative Count	Neutral Count		
2	1	labroom	0	1	0		
3	2	Lab1	2	4	0		
4							
5							
6							
7							
8							
9							

Figure 4.23 Export excel file for labroom related data

The following figures also show the *Canteen* related data selection, graph presentation and excel file export.

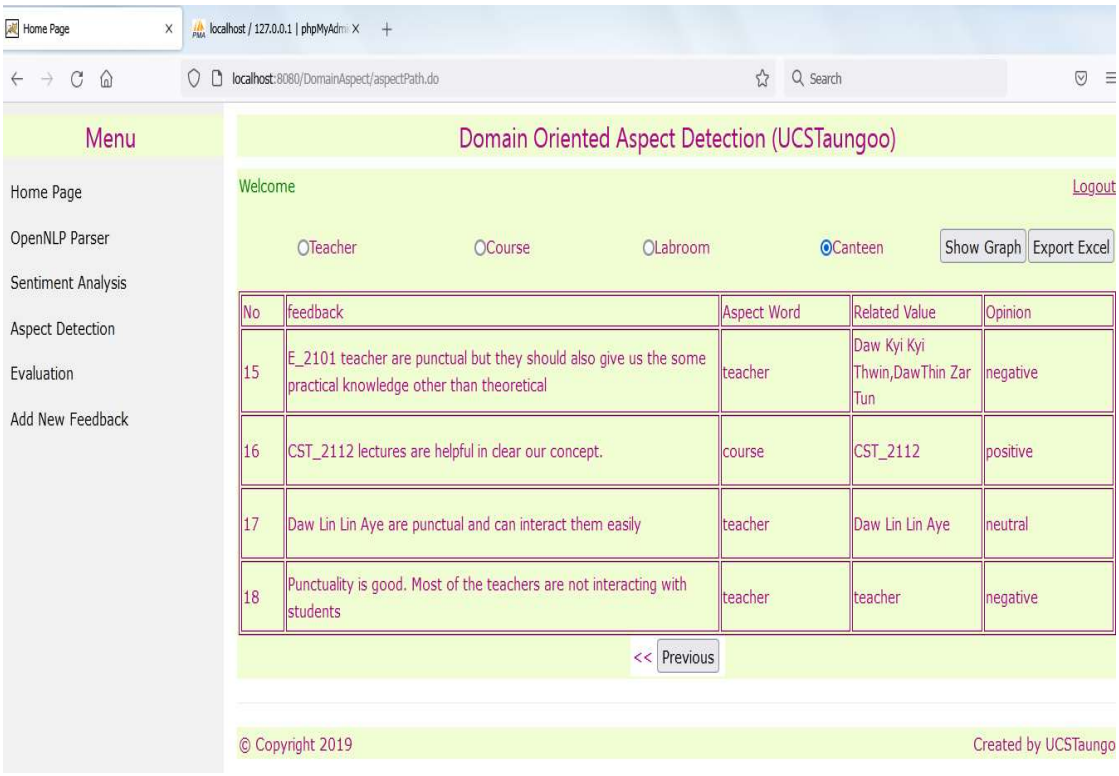


Figure 4.24 Canteen aspect category selections for visualization

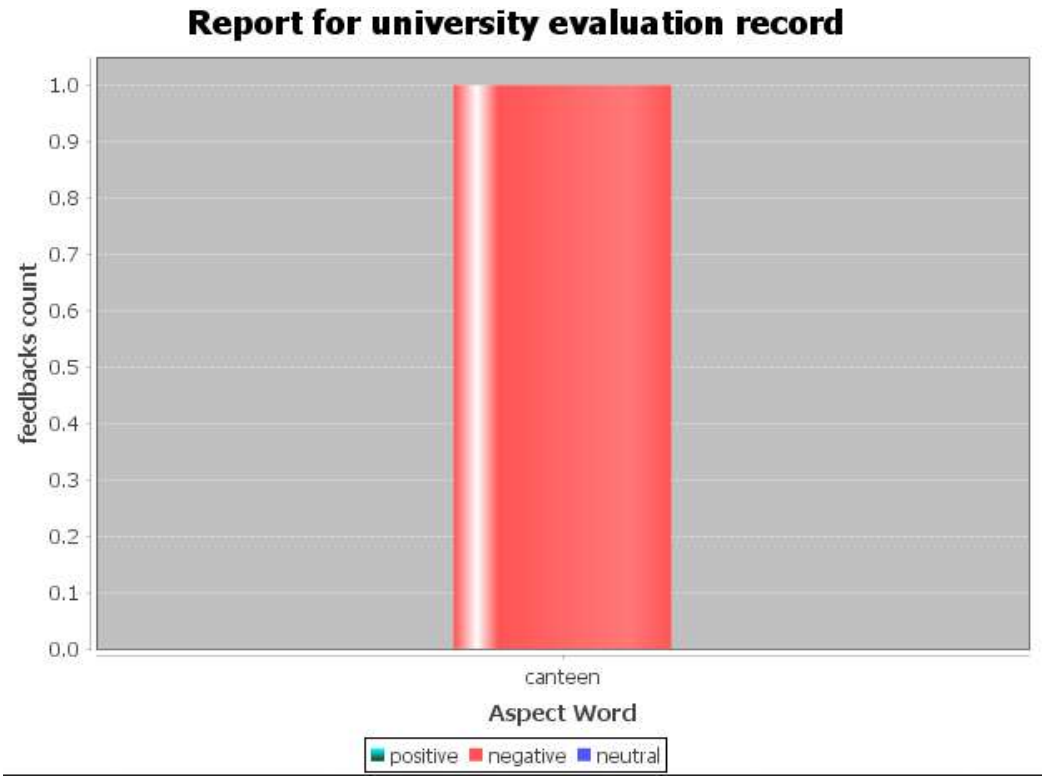


Figure 4.25 Export excel file for canteen related data

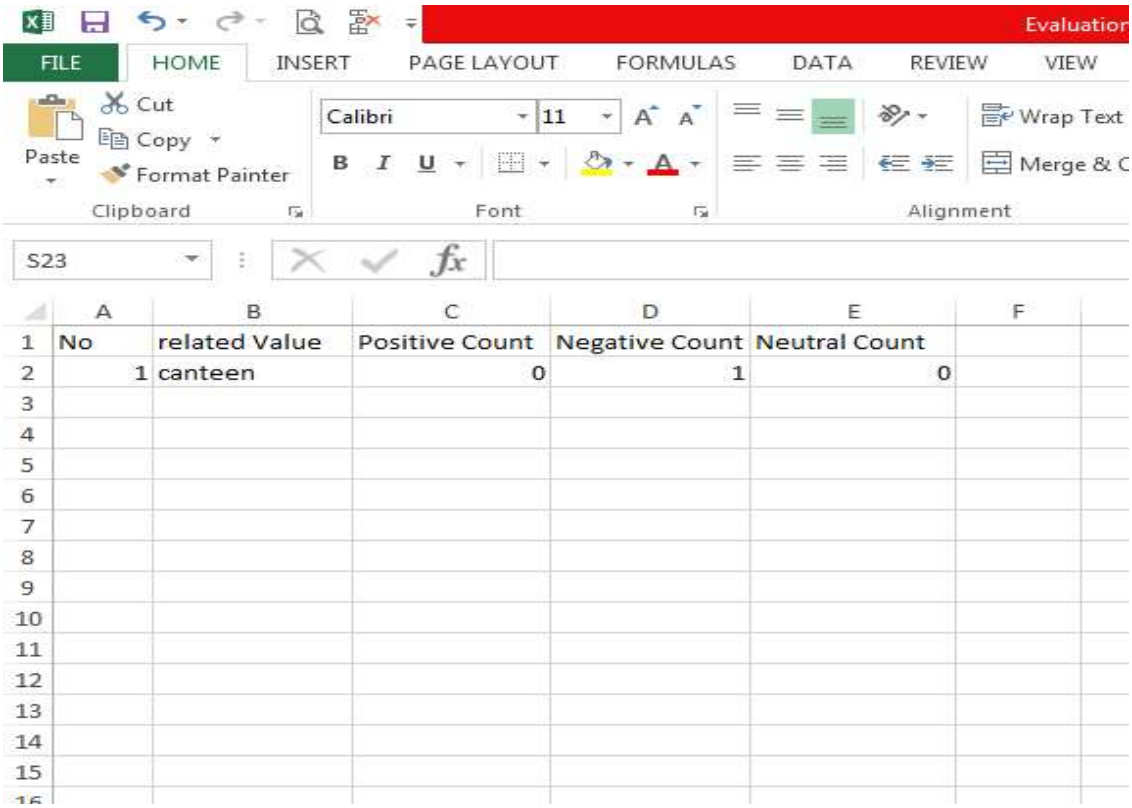


Figure 4.26 Export excel file for canteen related data

4.6 Evaluation Results

In order to see the evaluation result of this system, the user or admin can click the *Evaluation* link. After clicking this link, the webpage *EvaluationPage* shows the history records with 7 records per page. In this page, the user or admin can press the previous or next button to see all history records.

When the user or admin clicks the *Evaluate* button, the system starts to calculate the true positive rate, false positive rate, true negative rate and false negative rate in order to calculate the accuracy index. There are 100 history records and this system also calculates the accuracy index based on these records. After calculating each rate value and the accuracy index, these results are forwarded to the *EvaluationResultPage* and this page continue to show these result and conclude this Domain Oriented Aspect Detection System. The accuracy of this system based on these 100 records is 94%. These resulted pages are shown in the following figures.

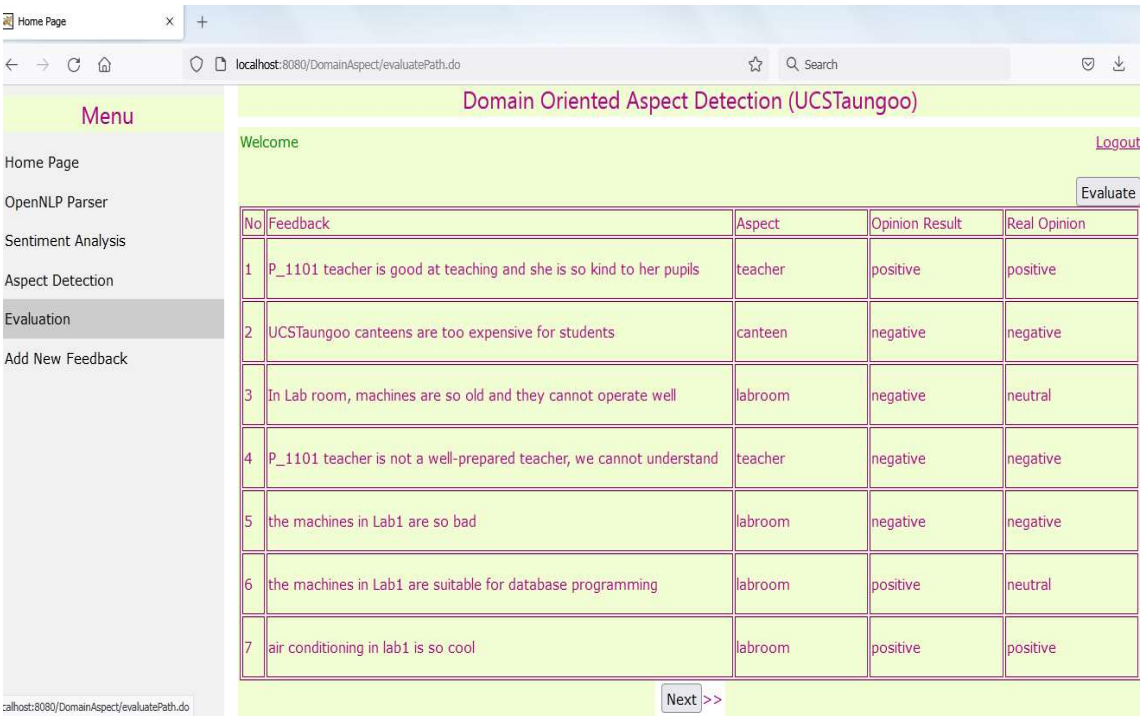


Figure 4.27 History records for system evaluation

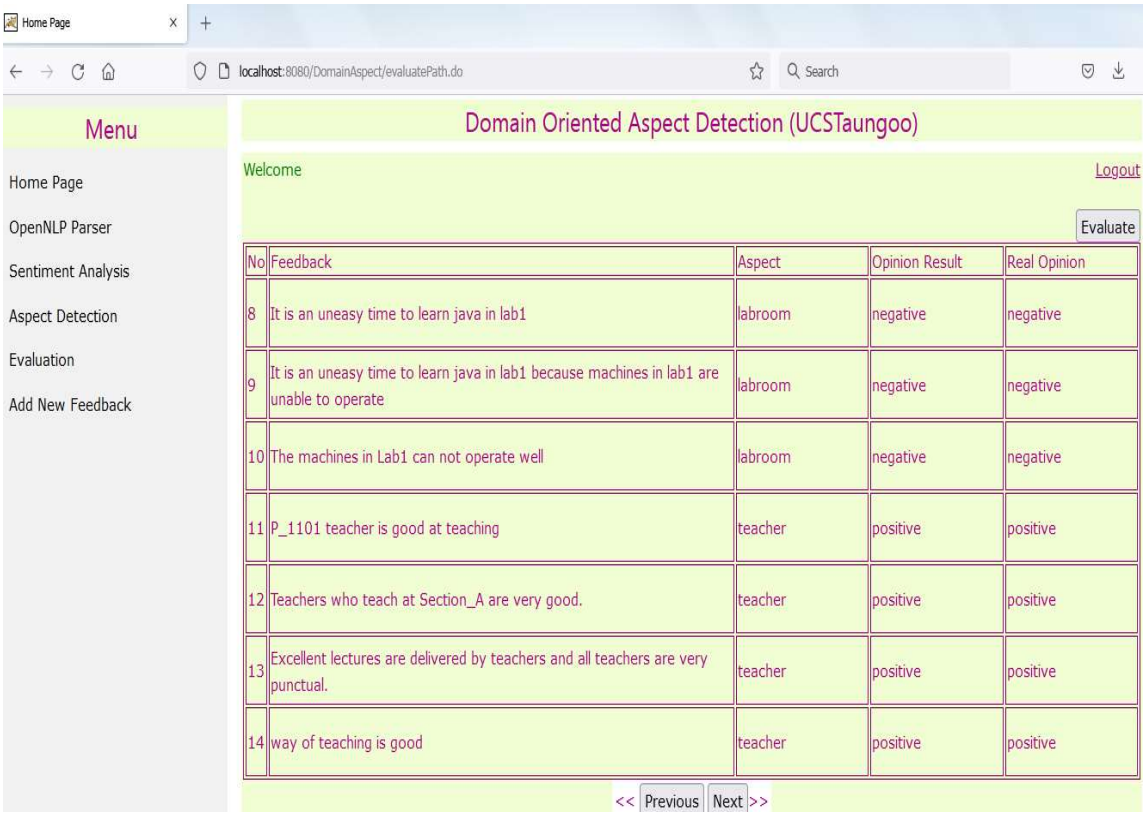


Figure 4.28 Next page segment of history records for system evaluation

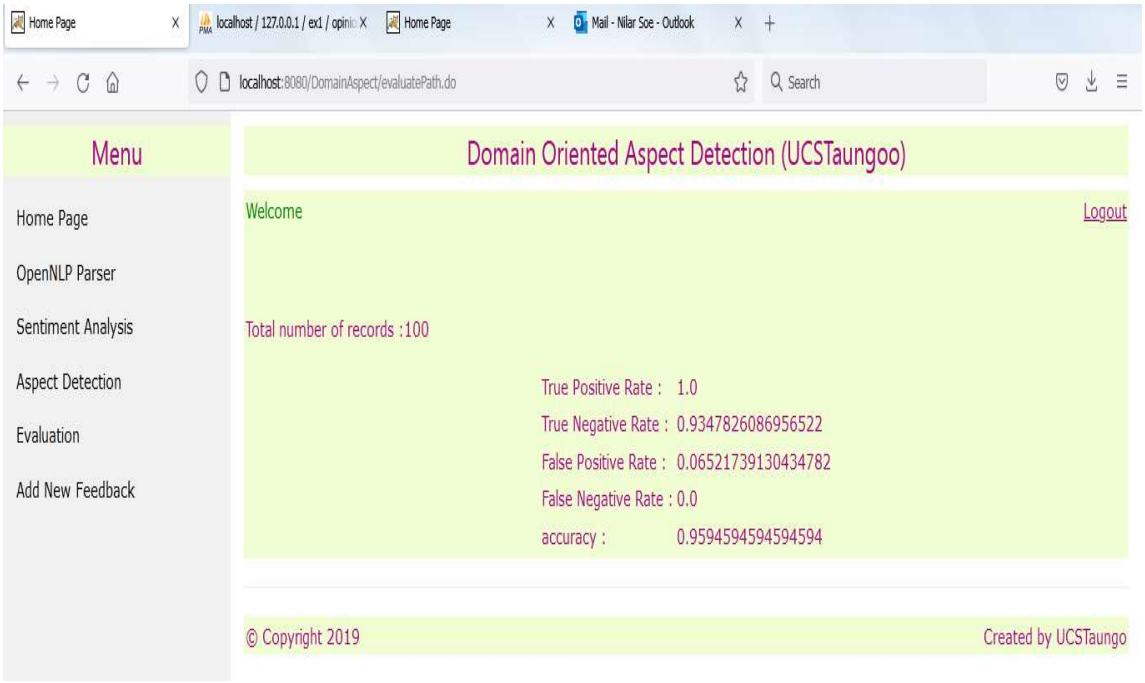


Figure 4.29 Evaluation Result Page

After all or on any page of the system, if the user clicks the *logout* link, the system changes its state to the *login* page.

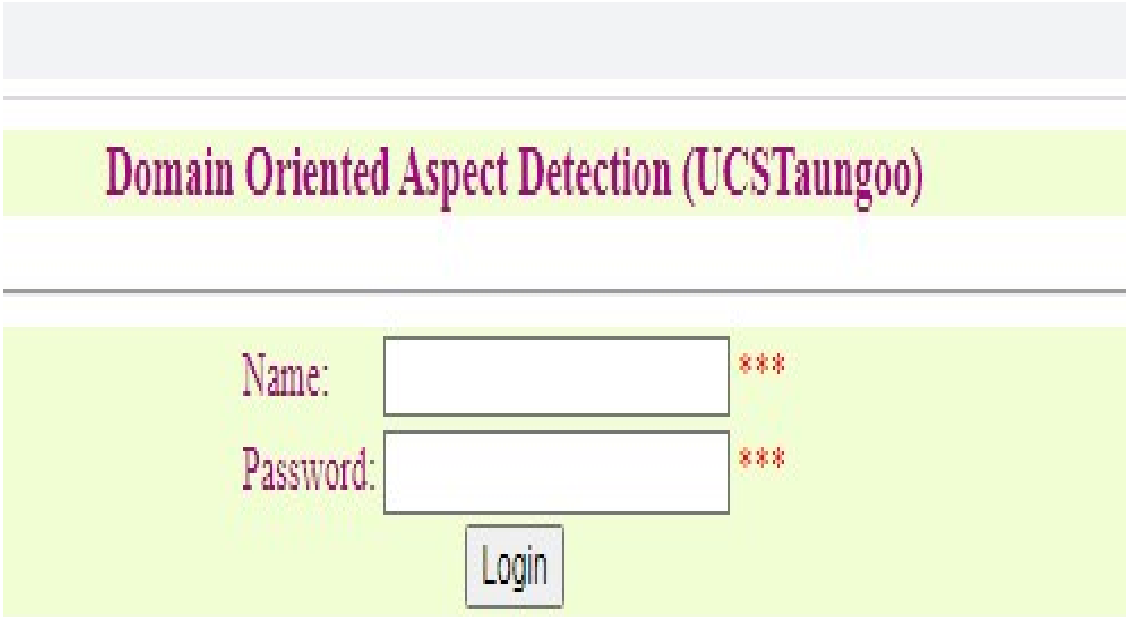


Figure 4.30 Logout state of the system

CHAPTER 5

CONCLUSION

The Domain Oriented Aspect Detection system uses domain oriented ontology to make aspect detection process. It can also identify the emotion of user from their input feedbacks by utilizing the operations of *OpenNLP* parser and features of *SentiWordNet* lexical resource. So, it is a kind of unsupervised opinion mining approach. According to the facts described in this thesis, the system well assists to the administrator of campus who has responsibility to care about the students' feedbacks. By analyzing on these feedbacks, the administrative person can review the opinion of students on specific things in campus in a short time. He can make decision for improving the performance of campus based on the result from analysis. This system can visualize the sentiment result for each aspect words that is corresponding to the class and individual values detail composed in ontology to provide the summarize report. The system accuracy is calculated based on the sentiment result and it is obtained as 95 % and so that is acceptable.

5.1 Advantages of the System

The system in this thesis uses unsupervised lexicon- based opinion mining approach so that it has advantage of no training data is required. The system accuracy of other supervised approach depends on a lot of feedbacks they collected whereas the system accuracy of unsupervised approach does not depend on the numbers of feedbacks that can analyze all kind of feedbacks without prior knowledge. Moreover, the processing speed of lexicon based approach is much shorter than other supervised approach because supervised approaches require long running time for their training process. And also, the usability of this system can assist the administrative person in order to easily determine to improve the performance of their organization. The using of bar graph visualization technique can improve the understandability for this system result. For these reasons, this system can be very useful and applicable for end users, administrative persons.

5.2 Limitations of the System

The system in this thesis has some limitations. The first limitation of this system is that it can examine the opinion scores of the aspect words for only English sentences. The second limitation is that it can extract the aspect terms from the sentence for only aspects which are specifically organized for the campus *UCSTaungoo*. The *POS* tagging process of this system is made on each feedback statements and the bigger the size of feedbacks selection, the more processing time is required for passing *ParserAction* stage. So, this is the third limitation of this system. And the fourth or final important limitation is that this system can solve the aspect words ambiguity but all aspects word in the input sentence are cohesive and intend to only one thing the sentences with conjunctions in which more than one aspect separately intend to their respective things because the aspect words are organized based on the possible relationship between them.

5.3 Further Extension

This system can be improved by adding the more aspect categories into domain specific ontology. It still needs to improve the overall processing time to be convenient with the administrative process but it is now acceptable by using high speed computer. The lexicon with more kinds of opinion terms can make this system to do more perfect sentiment analysis. It still needs to precisely define the action words from a given sentence. Moreover, this will be able to upgrade for defining more than one kind of aspects in a given sentence and their associated opinion.

AUTHOR'S PUBLICATIONS

[1] Nilar Soe, Paing Thwe Soe, "Domain Oriented Aspect Detection for Student Feedback System", to be published in the Proceedings of the 3rd International Conference on Advanced Information Techonologies (ICAIT), Yangon, Nov 6-7, 2019.

REFERENCES

[1] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification", Proceedings of 14th ACM International Conference on Information and Knowledge Management, Bremen, Germany, 2005.

[2] A. Katrekar, "An Introduction to Sentiment Analysis", https://www.globallogic.com, 2019.

[3] A. M. Ortiz, C. P. Hern´andez, "Lingmotif-lex: aWide-coverage, State-of-the-art Lexicon for Sentiment Analysis", In Proceeding of the International Conference on Language Resources and Evaluation, pages 2653–2659, 2011.

[4] A. Rhouati, J. Berrich, M. G. Belkasmi and T. Bouchentouf, "Sentiment Analysis of French Tweets based on Subjective Lexicon Approach: Evaluation of the use of OpenNLP and CoreNLP Tools", Journal of Computer Science 2018, 14 (6): 829.836.

[5] A Rozeva and S. Zerkova, "Assessing Semantic Similarity of Texts – Methods and Algorithms", Proceedings of the 43rd International Conference Applications of Mathematics in Engineering and Economics, pp. 060012-1 to 060012-8, December, 2017;

[6] A. S. Manek, Pallavi R P, V. H. Bhat, P. D. Shenoy, M. C. Mohan, Veenugopal K R and L M Patnaik, "SentReP: Sentiment Classification of Movie Reviews using Efficient Repetitive Pre-Processing", 978-1-4799-2827-9/13 IEEE, 2013.

[7] B. Li[10]u. "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.

[8] B. Pang and L. Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, Vol. 2, No 1-2, pp. 1–135, 2008.

[9] C,L. Zhou, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches", Proceedings of the 38th Hawaii International Conference on System Sciences, 2005.

[10] D.Kanbur, M. S. Aktas, “Ontology Based Aspect Oriented Opinion Summary Methodology”, The proceeding of 5th International Conference on Building and Exploring Web Based Environments, Turkey, 2017.

[11] G. Majumder, P. Pakray, A. Gelbukh and D. Pinto, “Semantic Textual Similarity Methods, Tools, and Applications: A Survey”, Proceedings of the 43rd International Conference Applications of Mathematics in Engineering and Economics, AIP Conf. pp. 647–665, No. 4, Vol. 20, 2016.

[12] J. Zhang, Y. Sun, H. Wang and Y. He, “Calculating Statistical Similarity between Sentences”, Journal of Convergence Information Technology, Volume 6, Number 2. February, 2011.

[13] K. Denecke, “Using SentiWordNet for Multilingual Sentiment Analysis”, The proceeding of 24th International conference on Data Engineering Workshop, 2008.

[14] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, “Lexicon-BasedMethods for Sentiment Analysis”, Computational Linguistics Volume 37, Number 2, pp.267-307, September, 2010.

[15] M. M. Taye, “Understanding Semantic Web and Ontologies: Theory and Applications”, Journal of Computing, Volume 2, Issue 6, June 2010.

[19] Sridevi. U.K, Shanthi. P, “An Ontology-based Sentiment Analysis Model towards
Classification of Drug Reviews”, International Semantic Intelligence Conference, Delhi, India, pp. 25-27, February, 2021.

[16] Ohana, Bruno, "Opinion mining with the SentWordNet lexical resource", Dissertations, https://arrow.tudublin.ie/scschcomdis, 2009.

[17] O. Hartig, “Foundation of RDF* and SPARQL* (Alternative Approach to Statement Level Metadata in RDF) ”, The proceeding of International Semantic Web Conference, Sweden, 2017.

[18] Ohana, Bruno, "Opinion mining with the SentWordNet lexical resource", Dissertations, https://arrow.tudublin.ie/scschcomdis, 2009.

[19] P.K.Singh and M.S.Husain, "Methodological Study Of Opinion Mining and Sentiment Analysis Techniques", International Journal on Soft Computing(IJSC), Vol. 5, No. 1, February 2014.

[20] P. Patil and P. Yalagi, "Sentiment Analysis Levels and Techniques: A Survey", International Journal of Innovations in Engineering and Technology (IJIET), Volume 6 Issue 4 April 2016. pp. 523–528.

[21] Sophie de Kok, Linda Punt, Rosita van den Puttelaar, Karoliina Ranta, Kim Schouten, Flavius Frasincar "Review-Level Aspect-Based Sentiment Analysis Using an Ontology" Erasmus University Rotterdam Rotterdam, the Netherlands SAC 2018, April 9–13, 2018.

[22] S. Padmaja and Prof. S. S. Fatima, "Opinion Mining and Sentiment Analysis – An Assessment of Peoples' Belief: A Survey", International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.4, No.1, February 2013.

[23] Sridevi. U.K, Shanthi. P, "An Ontology-based Sentiment Analysis Model towards
Classification of Drug Reviews", International Semantic Intelligence Conference, Delhi, India, pp. 25-27, February, 2021.

[24] T.C. Peng and C.C. Shih , "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs", IEEE/WIC/ACM International Conference on Web Intelligence and intelligent Agent Technology JOURNAL, 2010.

[25] Z. T. T. Myint, K. K. Win, "Triple Pattern Extraction for Accessing Data on Ontology", International Journal of Future Computer and Computer Science (IJFCC)", Thailand, Vol. 3, No. 1, February 2014, pp.40-44.