PREDICTING MUET QUESTION

BY

CHOI JIEN PING

A REPORT

SUBMITTED TO

University Tunku Abdul Rahman

In partial fulfillment of the requirements

For the degree of

BACHELOR OF INFORMATION SYSTEM (HONS)

INFORMATION SYSTEM ENGINEERING

Faculty of Information Communication Tchnology

(Perak Campus)

JAN 2015

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: ____

____

____

Academic Session: _____________

I ____

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

____

(Author's signature)

____

(Supervisor's signature)

Address:

____

____

____

____

Supervisor's name

Date: _____________________

Date: ____________________

PREDICTING MUET QUESTION

BY

CHOI JIEN PING

A REPORT

SUBMITTED TO

University Tunku Abdul Rahman

In partial fulfillment of the requirements

For the degree of

BACHELOR OF INFORMATION SYSTEM (HONS)

INFORMATION SYSTEM ENGINEERING

Faculty of Information Communication Tchnology

(Perak Campus)

JAN 2015

DECLARATION OF ORIGINALITY

I declare that this report entitled "**PREDICTING MUET QUESTION**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : ______________________

Name : ______________________

Date : ______________________

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisor, Dr. Ng Yen Kaow who has given me a chance to engage in this machine learning project. Besides, he also provided me a lot of constructive feedback and idea for completing this project. It is my first step to establish a career in machine learning field. A million thanks to you.

Apart from that, I would like to thanks the very special person in my life, Ang Wen Kai, for his intelligence, unconditional support and patience, and for always standing by my points of view during hard times. Finally, I would like to say thanks to my parents for their love and support throughout the course.

ABSTRACTS

The aim of this project is to automatically generate (the comprehension part of the) MUET questions for students to practice on, thus allowing them to better prepare themselves for the MUET exam. The questions are generated through the Question Generation (QG) approach, an approach that takes an article or paragraph as input and converts sentences into question form. In our system, the articles are selected according to a method which chooses the most likely article to appear in a given year of examination.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I: INTRODUCTION

1.1) Motivation and Problem Statement

In preparing for the MUET exam, students usually will practice on past MUET questions. This helps to build up their confidence before they enter the exam room, as well as familiarizing them with the examination format. However, one shortcoming in using past MUET questions for this purpose is that these questions are limited in numbers. Also, a student preparing for MUET may hope that the questions they practice on will familiarize themselves with the subject topic that will come up in the exam. However, practicing on past questions has very poor guarantee that the student will be exposed to upcoming subject topics, since it is unlikely for past questions to reappear. Also, it has long been suspected that MUET questions in the comprehension part are somewhat related to current affairs. Hence, at least for the comprehension questions, it would be desirable for an experienced teacher to identify the subject topic that is more likely to appear in the coming exam, and prepare mock exam questions for the MUET student.

On the other hand, the question generation is not an easy task. In order to identify a subject topic to use for the questions, a teacher must be well-informed of past-year questions as well as current affairs. Also, after an article is identified, it is a tedious task to create questions out of the article.

1.2) Project Scope

This project will create and develop an application program which implements the algorithm to generate comprehension type questions based on past MUET questions as well as news articles obtained from the internet. The articles are identified through machine learning methods which attempt to identify the subject topic which is more likely to be relevant for the specific year’s examination. After articles are identified, questions are generated from the questions and displayed in a graphical user interface.

1.3) Project Objective

There are three important requirements of the application program, which are: (1) it should able to predict the coming MUET question, and (2) it should able to generate a set of questions from a complex paragraph or article.

First, the application program will focus on predicting what type of question or which field of question might come out in the coming MUET exam. After that, the application program will generate comprehension type questions set based on a complex paragraph or an article. The questions generated should follow the Standard English language without any syntax or grammatical error, and be helpful in examining the understanding of the student.

1.4) Impact, significance and contribution

Both students and teachers can benefit from this system. Students may use it as a tool for MUET revision. Teachers can use this for preparing teaching material.

1.5) Background information

English language, sometimes described as the foremost global lingua franca (Smith Ross 2005), is also known as the most common language for most of the people. These days, English language plays a major role in many aspects especially during business processes and in the sector of education. Mastery of the language often helps in a person's career since it provides the person with the skill to deal with the international world. English language has been ingrained in many Commonwealth countries such as Malaysia, Brunei, India, *etc*. The use of the language is important to developing countries in order to expand their economy and thicken the relationship bridge with the developed countries.

To evaluate the English proficiency of tertiary students, a special examination called the Malaysian University English Test (MUET) was introduced in 1999. The examination was carried out twice a year from 1999 to 2011, that is, during the middle of the year (April and May) and end of the year (October and November). This frequency is increased to three times a year from the year 2012. The MUET result is a grade, called a Band, which is derived from the total score from the various parts of the exam paper. The highest grade of MUET is Band 6 (around 260 to 300 scores) while the lowest is Band one (below 100 scores) (Chong, 2007). Candidates of the

MUET exam will be considered as having a high level of English language as long as they obtained a Band 4 and above grade (around 180 to 219 scores) from the MUET exam. To reach such level of expertise, most of the candidates or students try their best to obtain the highest score from MUET exam.

The idea of using technology for exam preparation is not new. The technology which I consider in this project, Question Generation, is among the better studied technologies that are being used currently. A study by Luxton-Reilly (2012) has shown that the use of the Question Generation approach has a constructive effect on performance in related exam questions and it is faithfully correlated with exam performance.

CHAPTER II: LITERATURE REVIEW

2.1) Question Generation Phase

As discussed in the previous section, Question Generation (QG) is a well-studied field. According to Vasile (2010), QG is a vital and challenging approach or system where extraction of knowledge and representation in natural language is desired. While Question Generation is a very famous approach, no earlier work has been found which try to use the Question Generation to generate and predict exam questions set especially generate MUET questions.

Since the common Question Generation processes are similar to the present processes for MUET, the common techniques of Question Generation approach can be used for our problem. Nevertheless, depending on the complexity and difficulty of the question generating process, the Question Generation approach may need to be modified. In this section, some common methods for generating and ranking question shall be surveyed.

In the work by Heilman *et al.* (2011), they studied and investigated the methods including generating questions based on a text corpus. Basically, the question generation process is separated into 3 stages, which are:

1. Transformation of Declarative Input Sentences,
2. Question Creation, and
3. Question Ranking

In which each of the stages is separated into several sub-stages which will be discussed one by one in this section.

During the implementation of stage 1, Heilman *et al.* (2011) propose two types of transformation, which are: the extraction of sets of simplified factual statements from the complex structure of input sentences, and pronoun resolution. During the process of extraction simplified factual statement, the input sentences are extracted based on two phenomena: *semantic entailment* and *presupposition*. Appendix A describes the meaning of words in more detail. The semantic entailment that will be extracted includes *adjunct modifiers*, *discourse connectives* and *conjunctions*. The meanings of these words are given in Appendix A.

Both of the discourse connectives and adjunct modifiers can be removed from the clauses, verb phrases, and noun phrases. For instance, from example 2.1, we can extract example 2.2.

(2.1) Nevertheless, John still did not take his lunch, which was prepared by his wife, because he is too busy.

(2.2) John still did not take his lunch because he is too busy.

The extraction of adjunct modifiers and discourse connective is based on some criteria (Heilman *et al.*, 2011). For the conjunctions extraction, the system will split the complex text or paragraph into several short sentences, which might not include the conjunctions with “or” and “nor”.

Next, after extracting the semantic entailment, the process will proceed to the presupposition extraction. Extraction by presupposition aims to convey the information from text. This is a very important part for stage 1, because if the sentences are extracted based only on their semantic entailment, it would miss possibly useful questions about facts. According to Heilman *et al.* (2011), many presuppositions contain triggers, which will assist the extraction process to be more effective and generate the concise questions. However, the work of Michael Heilman does not address all the possible triggers.

Furthermore, the stage 1 will then proceed to the final process, which is transformation of declarative input sentences by resolution the pronoun. Pronoun resolution is also an important process for stage 1 as the generated question will be vague if extracted statements contained unresolved pronouns, such as he, she, they and etc. Note that the pronoun replacement will be implemented by using a co-reference system called *ARKref* which will help the system to identify the antecedents of pronouns. After the pronouns of the sentences being replaced, stage 1 is completely finished and the processed sentences will then bring to stage 2 which is Question creation stage.

The text corpus extraction method of Heilman *et al.* (2011) is much more complex compared to the alternative that used by Agichtein *et al.* (2000) in his research. In their work, Agichtein *et al.* developed a system called *Snowball* that implemented the *Dual Iterative Pattern Expansion (DIPRE)* algorithm to extract structured table from a collection of HTML documents and find out the hidden valuable structured data. His research was more towards extracting and finding out the organization-location pairs. The *DIPRE* of *snowball* will first be trained with a handful of instances. The tuples are extracted with the method called *named-entity tags* for marking the search of new tuples, and then a novel technique is used to generate patterns. However, his work did not focus on extracting all the relevant information from the given input, which is the goal of this project.

Apart from that, Feng (2007) proposed a new and expressive framework which implements *Conditional Random Field (CRF)* models to formalize the procedure of information extraction which provide rapid alternative for the extraction process. However, Feng (2007) focuses more on scientific literature and biomedical data. Otterbacher *et al.* (2005) proposes to develop a question-focused sentence retrieval mechanism by using a topic-sensitive version of the *LexRank* method. Nevertheless, the topic-sensitive system restricts users to identify the query topic themselves. Both these works are not suitable for the goals of this project.

For the question generation (Stage 2), the system of Heilman *et al.* (2011) takes a declarative sentence as input, and then produces a set of possible questions for each sentence. There are six sub-phases included in stage 2, which are:

1. Marking Unmovable Phrase,
2. Generating Possible Question Phrase,
3. Decomposition of Main Verb (optional),
4. Subject-Auxiliary Inversion (optional),
5. Removing Answers and Inserting Question, and
6. Pre-processing.

At the end of the stage 2, some questions will be generated following the question words: who, what, where, when, whose, and how many. According to Heilman et al. (2011), stage 2 aims to over-generate grammatical, though not concise or specific, questions.

In sub-phase 1, the unmovable phrases are marked based on the WH movement constraints which prevent them to generate ungrammatical question. Besides, Heilman *et al.* (2011, p. 61), a set of *Tregex expressions* is used to mark the phrases in an input tree which cannot be answer phrases due to WH movement constraints. The full set of searching expressions is shown in the Table 2.1.

Purpose	Expression
To identify the main verb, which the system decomposes into a form of *do* followed by the base form of verb	ROOT < (S=clause < (VP =mainvp [< (/VB.?/=tensed !< is|was|were|am|are|has|have|had|do|does|did) | < /VB.?/=tensed !< VP]))
To identify the main clause of for subject-auxiliary inversion	ROOT=root < (S=clause <+(?VP.∗/) (VP < / (MD|VB.?)/=aux < (VP < /VB.?/=verb)))
To identify the main clause for subject auxiliary inversion in sentences with a copula and no auxiliary (e.g., *The currency's value is falling*).	ROOT=root < (S=clause <+ (/VP.∗/) (VP < (/VB.?/=copula < is|are|was|were|am) !< VP))

Table 2.1: Full set of searching expressions

Next, the sub-phase 2 will generate the possible questions for each marked unmovable phrases. The system will first mark the source sentence with set of high-level semantic types and uses these semantic types along with the syntactic structure, and then generate the possible questions. Note that software called *supersense tagger* is used for annotating the source sentences. The selection of WH word for generating question is based on the condition listed in Appendix A.

After the sub-phase 2, sub-phase 3 and 4 will be skipped if the *auxiliary verb* is presented inside the main verb or the answer phrase is a subject noun phrase. For instance, sub-phase 3 will be performed for the following sentence: *Gary played piano.* This sentence will be decomposed into: *Gary did play piano* before converted to question. Note that both sub-phase 3 and 4 also implemented the *Tregex expression.*

In the next sub-phase, the current input tree generated in the previous sub-phase is copied to produce a new candidate question. The selected answer phrase is removed and the previously generated question phrase is inserted into the candidate question. Next, in sub-phase 6, some of post-processing tasks are done to the candidate question, such as adding question mark to the last position of the question, remove extra white space, filter out the illogic question and etc.

Previous work of Kunichika *et al.* (2003) used different methods for question generation, which is syntactic information. Figure 1 indicates the example of syntactic information.

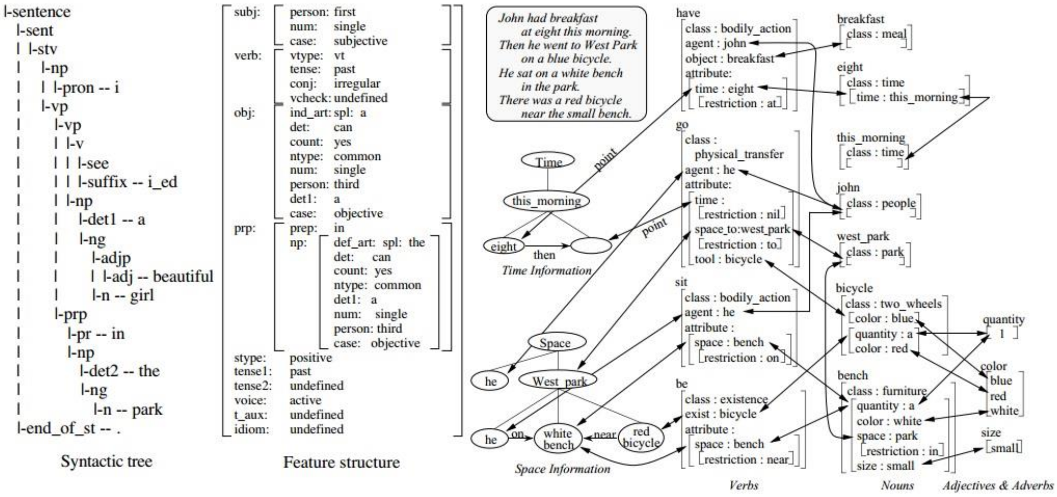


Figure 2.1: The example of syntactic information

The methods of Kunichika *et al.* (2003) are as follow:

1. To ask about the content of one sentence
2. To use synonyms or antonyms
3. To use modifiers appeared in plural sentences
4. To ask the contents by using relative pronoun
5. To ask relationship of space and time

These processes for question generation are fairly similar to the methods implemented in the work of Heilman *et al.* (2011) discussed above; while the fifth method has the distinguishably function compared to others. Kunichika *et al.* (2003) use this method to generate the "when" and "where" question sentences using inclusion relationship and referring to the partial ordering in time information. This is different with the method of Heilman *et al.* (2011) for generating "where" and "when" question based on the condition listed in Appendix A. However, Kunichika and his team did not describe the algorithms for each of the method and pseudo codes were not provided in their research. This makes the methods become ambiguous.

2.2) Question Ranking Phase

After the process of question generation, question ranking is a vital process in order to check whether the generated questions are valuable or not. Usually, ranked questions can analyze the consistency of the system while gaining the confidence from users and become more acceptable. Heilman *et al.* (2011) proposed a *least squares linear regression* that model the questions quality and the question levels acceptability based on the linguistic factor. The acceptability of a question is modeled as follows:

$$y = \mathrm{w}^{\top}\mathrm{f}(x)$$

Where y indicates the acceptability, x indicates question, f indicates a feature function, and w indicates a vector of real-valued weight parameter. The feature function, f, takes a question, x, as an input and return a vector of real-valued numbers. In order to minimize the sum of the squared errors on training data, an equation is as below:

$$\hat{\mathrm{W}} = \mathrm{argmin}_{w} \ \sum_{i=1}^{N}(y_i - w^{\top}\mathrm{f}(x_i))^2 + \lambda \ \|w\| \, 2_2 \ ,$$

where x_i indicates a single instance, and N indicates the total number of training example. This method uses a set of feature functions to evaluate the value of each question sentence. The example of the feature functions used to rank the question is listed at Appendix A.

Crammer *et al.* (2006) also propose two kinds of Passive Aggressive (PA) based ranking methods: PA-regression and PA-binary, where PA-regression is an online linear regression with PA version; while PA-binary is a ranking method that ranks question by using the binary classifier to sort the scores of question sentence for whether it is acceptable or not. Table 2.2 and equation below show the algorithms for both PA-regression and PA-binary.

Algorithm 3 PassiveAggressiveLinearRegression((**X,Y**),C,T):
A passive aggressive learning algorithm for linear regression, following Crammer et al. (2006). H_t,
ℓ_t, and $G_t$can be modified for binary classification or pairwise ranking (see text).

```
$W_1$ = (0,…,0)
$w_{avg}$ = (0,…,0)
for $t$ = 1,2,…,$T$ do
    sample: ($x_t$ , $y_t$) Є (X, Y)
    compute prediction for sample: $H_t$ = $w^{\top}\mathrm{f}(x_t)$
    compute loss based on prediction: $\ell_t$ = max {0,| $w^{\top}\mathrm{f}(x_i)$ - $y_t$| - Є}
    compute update based on loss: $G_t$ = min { $C$, $\frac{\ell_t}{||f(x_i)||^2}$ } sign($y_t - H_t$) f($x_i$)
    update parameters: $w_{t+1}$ = $w_t$ + $G_t$
    update average: $w_{avg}$ = $w_{avg}$ + $w_{t+1}/T$
end for
return $w_{avg}$
```

Table 2.2: The algorithm of PA-regression

$$H_t = \text{sign}(w^\top \text{f}(x_t))$$

$$G_t = \min\left\{ C, \frac{\ell_t}{||\text{f}(x_i)||^2}\right\} y_t \text{f}(x_t)$$

$$\ell_t = \max\{0,1 - y_t w^\top \text{f}(x_t)$$

Heilman *et al.* (2011) carried out an experiment to test the performance between their ranking method and the ranking methods of Crammer *et al*. (2006). They conclude that their *least squares linear regression* outperforms PA-regression and PA-binary.

QG approach has recently become more widely available. It is used in a variety of applications, which include QG for literature review writing support (Liu, Calvo & Rus, 2006), QG of multiple choice questions from domain ontologies (Papasalouros, Kanaris & Kotis, 2008) and QG from Web Based semantically treated search result (Liske, 2011). Since the QG applications employed on a wide range of domains, the interest of accuracy on the deliverables or parsers of QG become acute.

Black *et al.* (1991) proposed a prediction method. This method builds the parse tree from the interest domain by human annotators and uses them to compute a *PARSEVAL* value which indicates the performance of the parser. However, this method is claimed to be too expensive. Besides, Marcus *et al.* (1993) came out with an idea of measuring the parser performance on existing *treebanks* such as *WSJ.*

Nevertheless, this method cannot guarantee on the performance of a new domain and will give incorrect indications.

2.3) Prediction Phase

Our task in this phase is to predict a subject topic where the MUET question of a given year is likely to be under, given the prior knowledge of past year MUET questions as well as the corresponding current affairs for every year. To the best of my search, I could find no previous publication, or any references on how such a question prediction task can be performed. Thus, we devise our own technique using standard machine learning methods. Our prediction method is discussed in Chapter IV.

CHAPER III: PROPOSED METHOD / APPROACH

3.1) Design Specifications

3.1.1) Methodologies and General Work Procedures

Four main processes are required to perform for the purpose of achieving the primary objectives of this project, which are text corpus extraction, question creation, question ranking, and question prediction. Figure 3.1 below illustrates the flow of the processes.

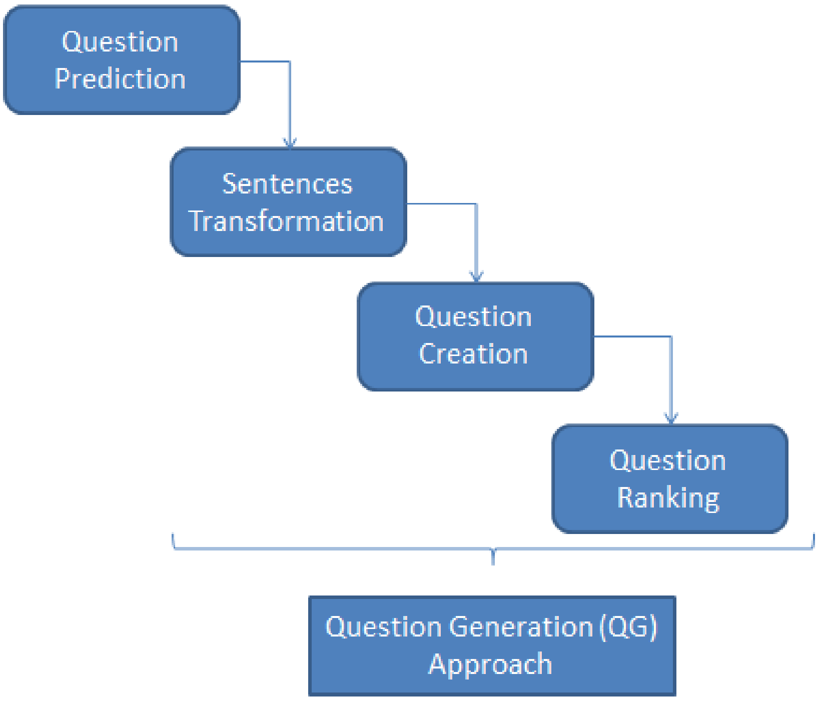
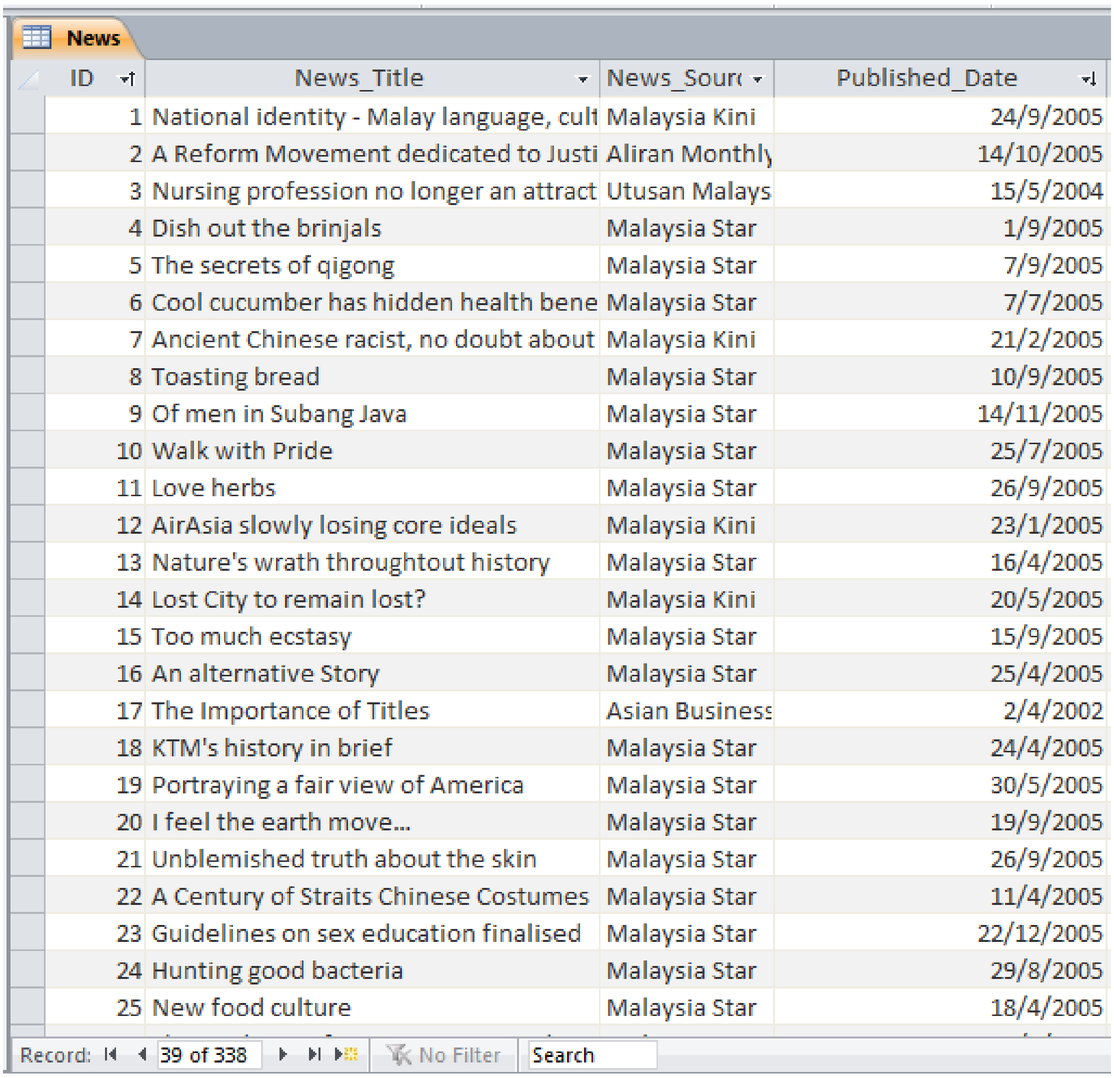


Figure 3.1: The flow of the process

The figure shows the flow of the processes of this project. In order to develop the application program that can solve the problem stated in Session III, the system development process is guided using the concept of waterfall model software development methodology. The progress of the software development process is flowing steadily downwards like a waterfall as show in Figure 3.1. Some of the system developers criticize that waterfall model is not suitable for non-trivial project since the process cannot be moving backward if clients tends to change their requirements frequently. However, it is possible for this project to implement the waterfall model as the process of QG is knowable, fixed and seldom changes.

The process begins with the prediction of the MUET exam question that will be come out in coming MUET exam. To perform so, an assumption was made such that, the question pattern is depending on the issues surrounding it, says the recent news or breaking events. Therefore, the news range from 2005 to 2012 is collected from Google News. Figure 3.2 shows the database of the collected news.



ID	News_Title	News_Sourc	Published_Date
1	National identity - Malay language, cult	Malaysia Kini	24/9/2005
2	A Reform Movement dedicated to Justi	Aliran Monthly	14/10/2005
3	Nursing profession no longer an attract	Utusan Malays	15/5/2004
4	Dish out the brinjals	Malaysia Star	1/9/2005
5	The secrets of qigong	Malaysia Star	7/9/2005
6	Cool cucumber has hidden health bene	Malaysia Star	7/7/2005
7	Ancient Chinese racist, no doubt about	Malaysia Kini	21/2/2005
8	Toasting bread	Malaysia Star	10/9/2005
9	Of men in Subang Java	Malaysia Star	14/11/2005
10	Walk with Pride	Malaysia Star	25/7/2005
11	Love herbs	Malaysia Star	26/9/2005
12	AirAsia slowly losing core ideals	Malaysia Kini	23/1/2005
13	Nature's wrath throughtout history	Malaysia Star	16/4/2005
14	Lost City to remain lost?	Malaysia Kini	20/5/2005
15	Too much ecstasy	Malaysia Star	15/9/2005
16	An alternative Story	Malaysia Star	25/4/2005
17	The Importance of Titles	Asian Business	2/4/2002
18	KTM's history in brief	Malaysia Star	24/4/2005
19	Portraying a fair view of America	Malaysia Star	30/5/2005
20	I feel the earth move...	Malaysia Star	19/9/2005
21	Unblemished truth about the skin	Malaysia Star	26/9/2005
22	A Century of Straits Chinese Costumes	Malaysia Star	11/4/2005
23	Guidelines on sex education finalised	Malaysia Star	22/12/2005
24	Hunting good bacteria	Malaysia Star	29/8/2005
25	New food culture	Malaysia Star	18/4/2005

Figure 3.2: The sample collected news database

In order to classify each of the news topic to different categories correctly, the classification of article in Wiki Vital Article was studied and each features used to classify the article was learned, for instance, features for Social Science category of article including “social”, “environment”, “education” and so on while features under Health article category including “disease”, “nutrition”, and so on.

Support Vector Machine (SVM) classifier was implemented to classify the collected news. SVM is a supervised learning model in machine learning and a discriminative classifier. With given labeled training data in supervised learning process, SVM will output an optimal hyperplane which can further categorize new examples.

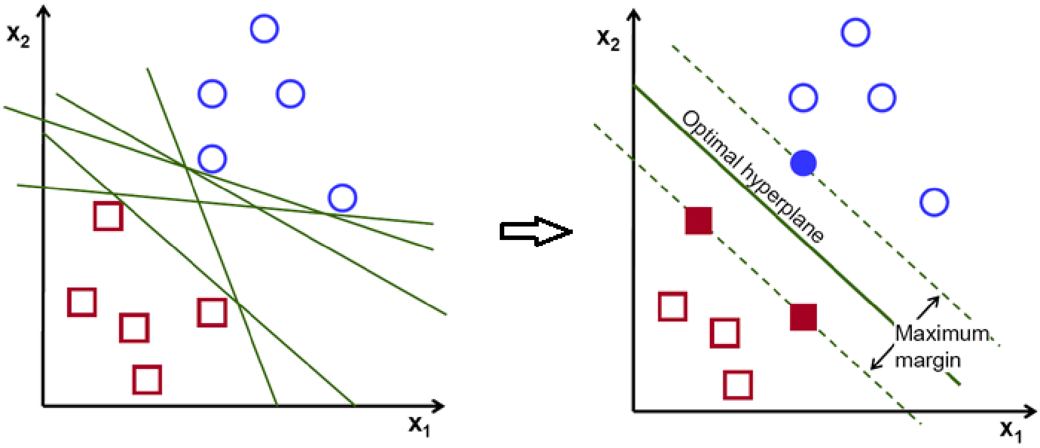


Figure 3.3: SVM obtains the optimal hyperplane

As referring to Figure 3.3, given two classes (in this case, feature and category of article), there are many lines that can separate the two classes, offering different solutions to the classification. The SVM algorithm operation will find the hyperplane that gives the largest minimum distance, known as a margin in SVM's theory, between the classes. This optimal hyperplane maximizes the training data margin to give a better classification solution. The classification process of news will be further discussed in Chapter IV and the medium used to perform this classification will discuss at later section.

After training the SVM classifier, the past year MUET questions and the news collected for each year are classified with the classifier.

In the prediction phase, I look for a function *f* which accepts as input

- The new articles, *S*, from a period of time relevant to the target examination paper, and
- a subject matter, *i*,

and gives as output a value, p, which indicates how likely the subject matter *i* is to come out in the target examination paper. That is, *f*(*S*, *i*) = *p*.

This is how the training samples are constructed. For each MUET paper *X*, I denote the set of articles collected from a period relevant to *X* as S_X. For each subject topic *i* and each S_X we denote the output as p_X^i, which is computed as $\frac{\#i}{N}$, where *N* is the total number of articles that appeared in *X*, and #i is the total number of articles in *X* classified as belonging to topic *i*. For each paper *X* and each topic *i*, (S_X, i, p_X^i) forms a training example. During prediction, the set of news from the period of time relevant to the target paper is used as input *S*. Each of the subject topic *i* is tested to give an output p^i. The subject topics with the highest p^i are given as the predicted topics.

I use a Multilayer Perceptron (MLP) to learn the function *f*. An MLP is one of the common Neural Network model, which is an effective artificial intelligence technology for information processing technique. It performs in such a way that is similar to biological nervous, in which, the signal received in dendrites represents the input, nucleus perform the process such as summation and activation, while the axon that send signal out represents the output (Ng, 2011). The artificial neural network formed when large numbers of neuron work together (McCulloch & Pitts, 1943)

A neural network consists of perceptrons. A perceptron is a linear classifier that computes only a single output (either 0 or 1) from multiple inputs. MLP is also known as a feedforward artificial neural network that consists of multiple layers of nodes in a directed graph which is completely connected from one layer to another without directed backward.

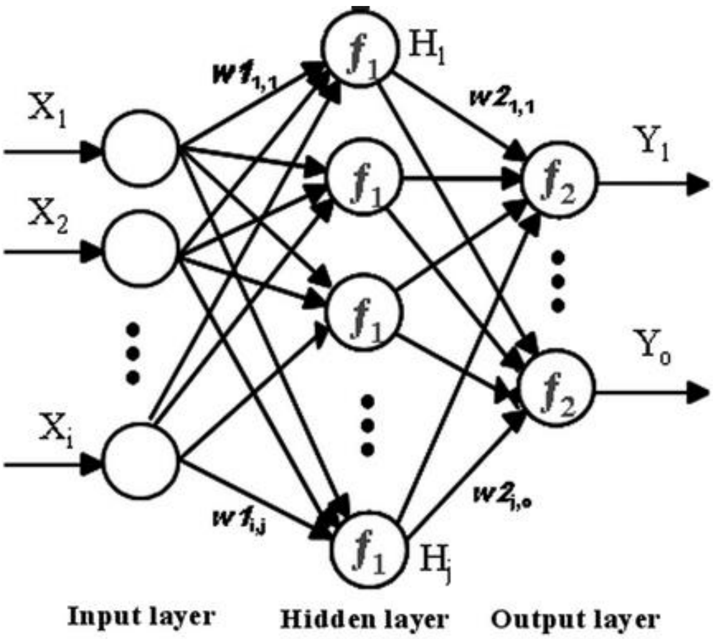


Figure 3.4: A graphical representation of MLP network

For predicting the exam question, MLP was implemented to create a model which will correctly map the input in to the output. The detail of the process will further discussed in Chapter IV.

After the scope or topic of possible MUET questions that might come out are predicted, the process will then proceed with the transformation of sentences, which include the transformations of semantic entailment and presupposition extraction that were employed and implemented in the work of Heilman *et al.* (2011), discussed in Sesction II. In order to perform the semantic entailment and presupposition extraction, the *Stanford Parser,* which will be scussed in later section, is implemented in order to parse input texts, tokenize and split the sentence such as conjunctions. The articles that are related with the predicted topic will be the input for the *Stanford Parser*. Besides, *ARKref* will also be implemented to identify the antecedents of pronouns of a sentence.

Next, after the factual statement is simplified, question will be generated based on it. The question generation process is implemented by using the *Tregex* tree searching language and *supersense tagger* which will be discussed in the later section. *Tregex* tree will "segment" each sentence into separate nodes and link them together into a tree call parse tree; while the *supersense tagger* will labels word tokens with high level of semantic class and labels proper noun. The input sentence will then be changed into a question by using WH-movement constraint to determine which WH-word to use. The question creation process will end with the task of pre-processing on each MUET sample questions to make them more reliable.

Lastly, the process will proceeds to the last part of QG approach, question ranking. To rank the MUET sample question, the method discussed at Session II is implemented, which is the *least squares linear regression* method that used by Michael Heilman et al. (2011). The ranked questions with high marks will be displayed in the front part of the question set followed by the ranked questions with less mark and so on.

3.1.2) Tools to use

In order to develop an application program for this project, my choice of language is Java. Java is an object-oriented and class-based computer programming language. Java mostly used to develop applications or program for a wide range of environment. The syntax rules of Java look much like C's since it is a C-language derivative. Java applications usually can run on any Java Virtual Machine (JVM) regardless of computer architecture. (Yang, 2014)

On the other hand, R Language also has been used to perform the Support Vector Machine Classification. R is an open source software programming language and mostly used for the statistical computing environment. It has the ability to analyze data. Besides, R gives its users access to cutting-edge technology. It allows users to easily visualize the data, perform machine learning algorithms, and perform some statistical testing. With new algorithms for different purpose keep adding to the list of packages, users can always download them and implement for their study. In this project, package "RTextTools" in R is used for new classification.

Apart from that, WEKA, the project developed by The University of Waikato, will also be used for MLP. With plenty of machine learning packages and collection of algorithms which are available publically, WEKA enables machine learning researches to automate the machine learning process. Besides, with simple user interface provided in WEKA program, new researches like students will able to learn and apply in their study.

Furthermore, the integrated development environment (IDE) used to develop this application program is Eclipse. Eclipse IDE able to support multiple languages such as Java, PHP, Python, Ruby and so on. It allows developers to program different program as long as the plugin for the specific language is being installed. On the other hand, in order to implement R Language, R-studio, an open source IDE is used.

Moreover, the tools that being used in this project included *Stanford Phrase Structure Parser*, *Tregex Tree*, *Supersense Tagger*, and *ARKref*. This section will discuss each of the tools one by one.

Stanford Phrase Structure Parser

Stanford Phrase Structure Parser is a kind of statistical natural language parser program which has the function of working out the grammatical sentences structure. It uses language knowledge from hand-parsed sentences and try to produce the most probably analysis of new sentences. (Chris Manning et al, 2006) In this project, I use the latest version of Stanford Phrase Structure Parser, Version 3.3.0, which proposed in 12th Nov 2013. It is an open-source program which is can be downloaded in nlp.stanford.edu/downloads/lex-parser.shtml#Download.

Tregex Tree

Tregex Tree is a tool for querying and manipulating tree data structures. According to Levy & Andrew (2006), "*Tregex* remedies several expressive and implementational limitations of existing query tools". It can be used for matching patterns in tree. It is also a kind of free software which allows users to implement for research purpose. It is available in nlp.stanford.edu/downloads/tregex.shtml. In this project, I use version 3.3.0 of *Tregex* program to transform the sentences input into factual statement for question generation.

ARKref

One of the tools that I used in this project is *ARKref*. *ARKref* is a rule-based system that uses syntactic information from semantic information from an entity recognition component to constrain the set of possible mention candidates such as noun phrases. (Brendan O'Connor et al, 2013). The *ARKref* is an open-source which publicly available at http://www.ark.cs.cmu.edu/ARKref. In this project, I use *ARKref* tools to identify the antecedents of pronouns of a sentence.

Supersense Tagger

Supersense Tagger is a tool for annotating and assigning text to verb, adverb, adjective and noun with 45 standard WordNet supersenses. Its functions included labels word tokens with high level of semantic classes, labels proper nouns and able to tell whether the noun is proper noun or common noun. It is an open-source tool which available in sourceforge.net/projects/supersensetag/.

3.2) System Design / Overview

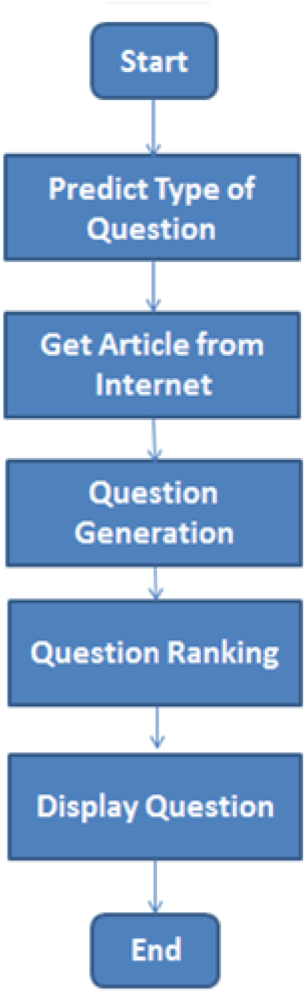
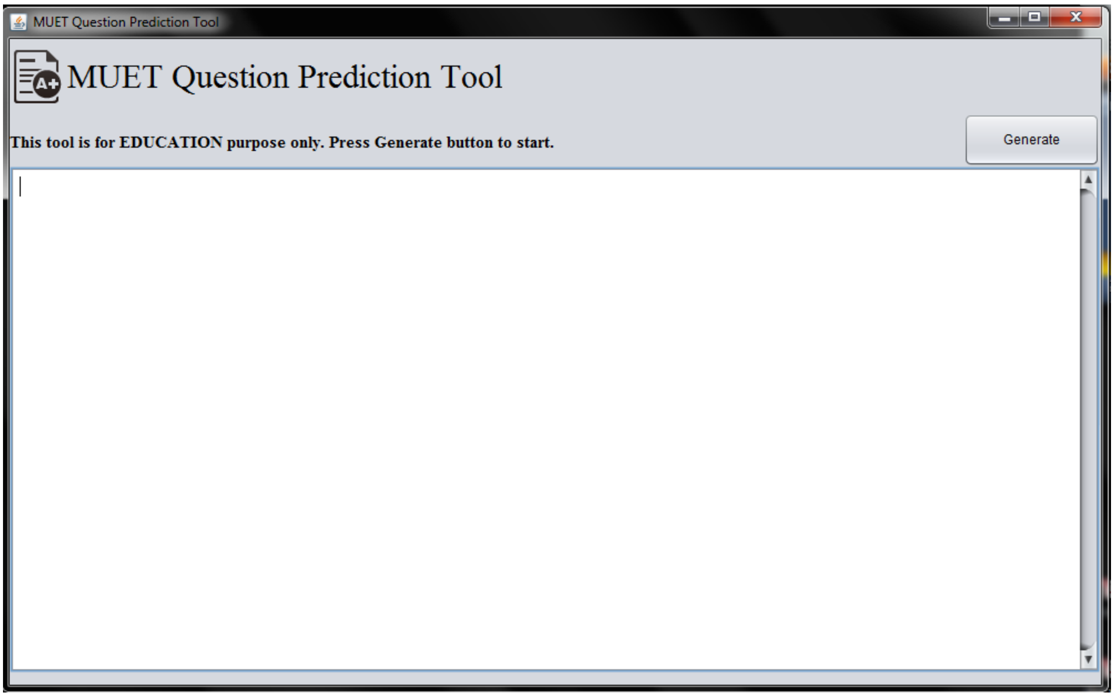


Figure 3.5: The System Flow Chart of the Application Program

Figure 3.5 shows the System Flow Chart of my application program. At the beginning, the system will allows user to select the session (i.e. 2010 May) of the coming MUET exam. Next, based on the selected session, the system will perform question prediction to predict the kinds of question which may come out at the coming MUET exam with the implementation of Bayesian Approach calculation.

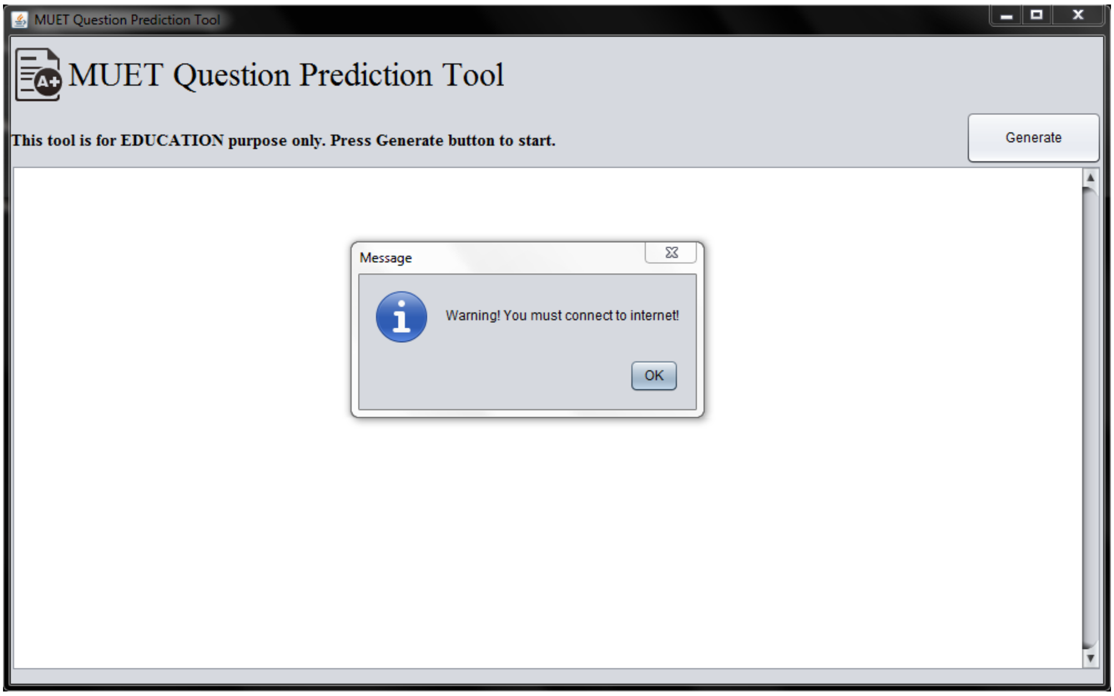
After the prediction, the system will access to internet to get the related article. Question generation process will then performed by the system after the related article is download successfully. The system will generate question based on the downloaded article. Then, the system will further proceed to Question Ranking process in which the generated questions are ranked. Lastly, the ranked questions will be displayed for user review.

Figure 3.6 shows that the main graphical user interfaces of the system. Users are allowed to press the "Generate" button to start the question prediction and question generation processes. Users must be connected to internet before start the processes, if not, an error dialog will prompt out as shown in Figure 3.7.



This tool is for EDUCATION purpose only. Press Generate button to start.

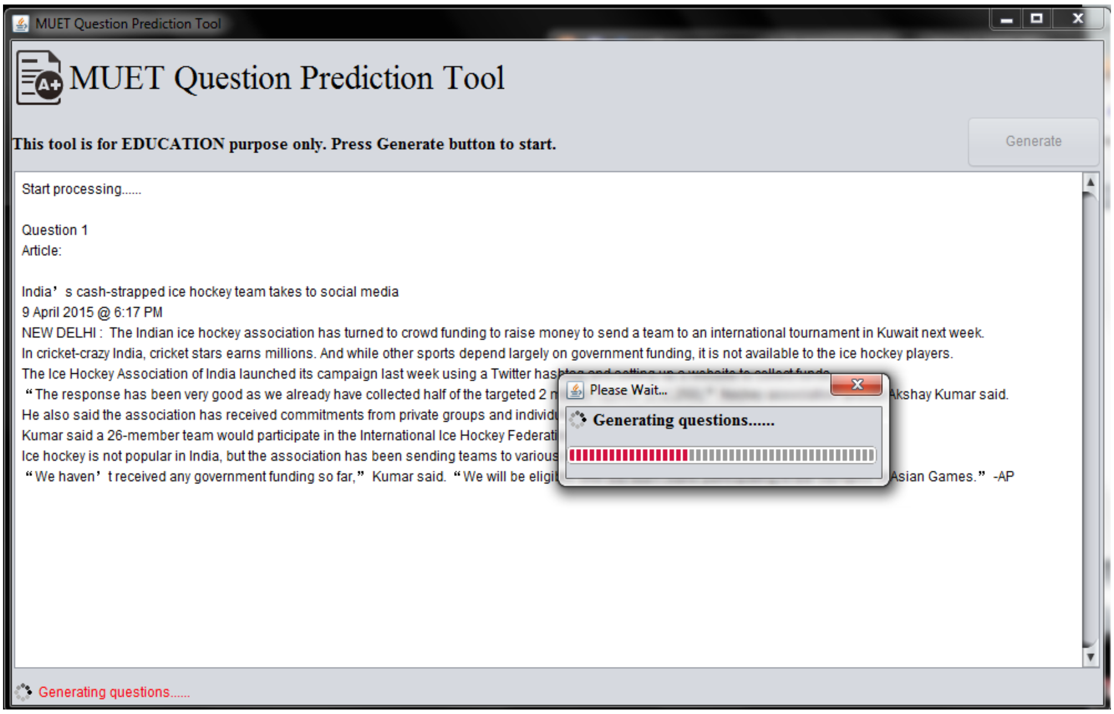
Figure 3.6: The starting graphical user interface of the application



This tool is for EDUCATION purpose only. Press Generate button to start.

Warning! You must connect to internet!

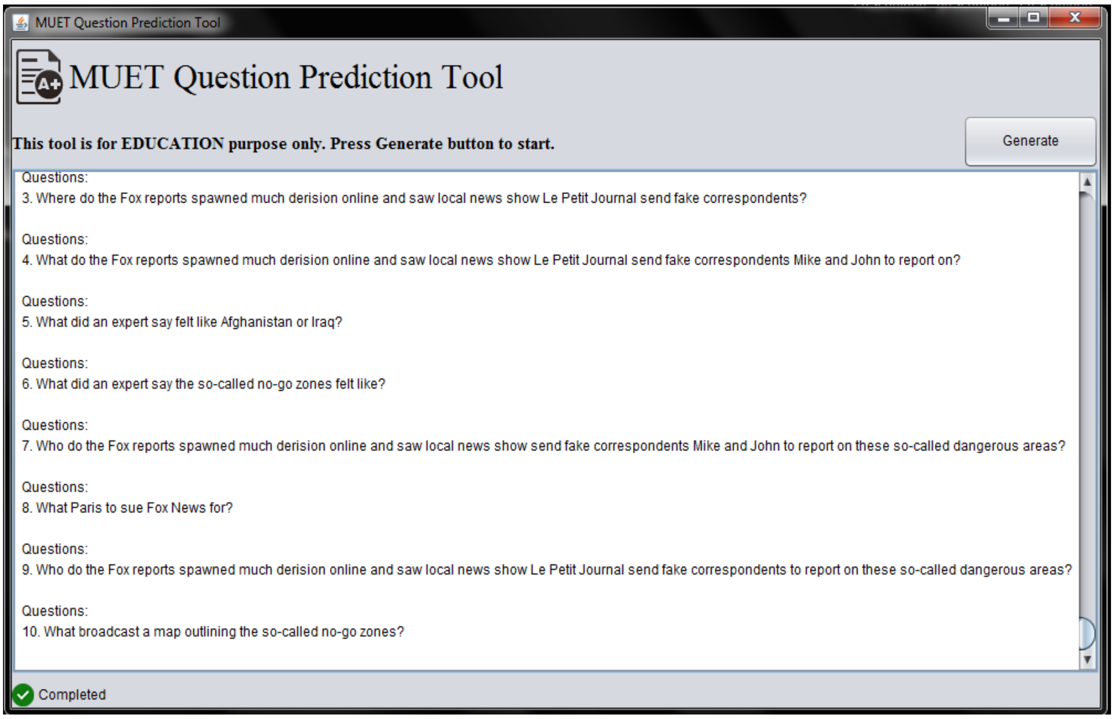
Figure 3.7: The error dialog show up when no internet connection


MUET Question Prediction Tool
MUET Question Prediction Tool
This tool is for EDUCATION purpose only. Press Generate button to start.
Generate
Please Wait...
Generating questions......

```
Start processing......

Question 1
Article:

India’ s cash-strapped ice hockey team takes to social media
9 April 2015 @ 6:17 PM
NEW DELHI : The Indian ice hockey association has turned to crowd funding to raise money to send a team to an international tournament in Kuwait next week.
In cricket-crazy India, cricket stars earns millions. And while other sports depend largely on government funding, it is not available to the ice hockey players.
The Ice Hockey Association of India launched its campaign last week using a Twitter has
 “The response has been very good as we already have collected half of the targeted 2 m                                                  Akshay Kumar said.
He also said the association has received commitments from private groups and individu
Kumar said a 26-member team would participate in the International Ice Hockey Federati
Ice hockey is not popular in India, but the association has been sending teams to various
 “We haven’ t received any government funding so far,” Kumar said. “We will be eligi                                                     Asian Games.” -AP
```

```
Generating questions......
```

Figure 3.8: The graphical user interface of the application after process start perform


MUET Question Prediction Tool
MUET Question Prediction Tool
This tool is for EDUCATION purpose only. Press Generate button to start.
Generate

```
Questions:
3. Where do the Fox reports spawned much derision online and saw local news show Le Petit Journal send fake correspondents?

Questions:
4. What do the Fox reports spawned much derision online and saw local news show Le Petit Journal send fake correspondents Mike and John to report on?

Questions:
5. What did an expert say felt like Afghanistan or Iraq?

Questions:
6. What did an expert say the so-called no-go zones felt like?

Questions:
7. Who do the Fox reports spawned much derision online and saw local news show send fake correspondents Mike and John to report on these so-called dangerous areas?

Questions:
8. What Paris to sue Fox News for?

Questions:
9. Who do the Fox reports spawned much derision online and saw local news show Le Petit Journal send fake correspondents to report on these so-called dangerous areas?

Questions:
10. What broadcast a map outlining the so-called no-go zones?
```

```
✓ Completed
```

Figure 3.9: The graphical user interface of the application after process is done

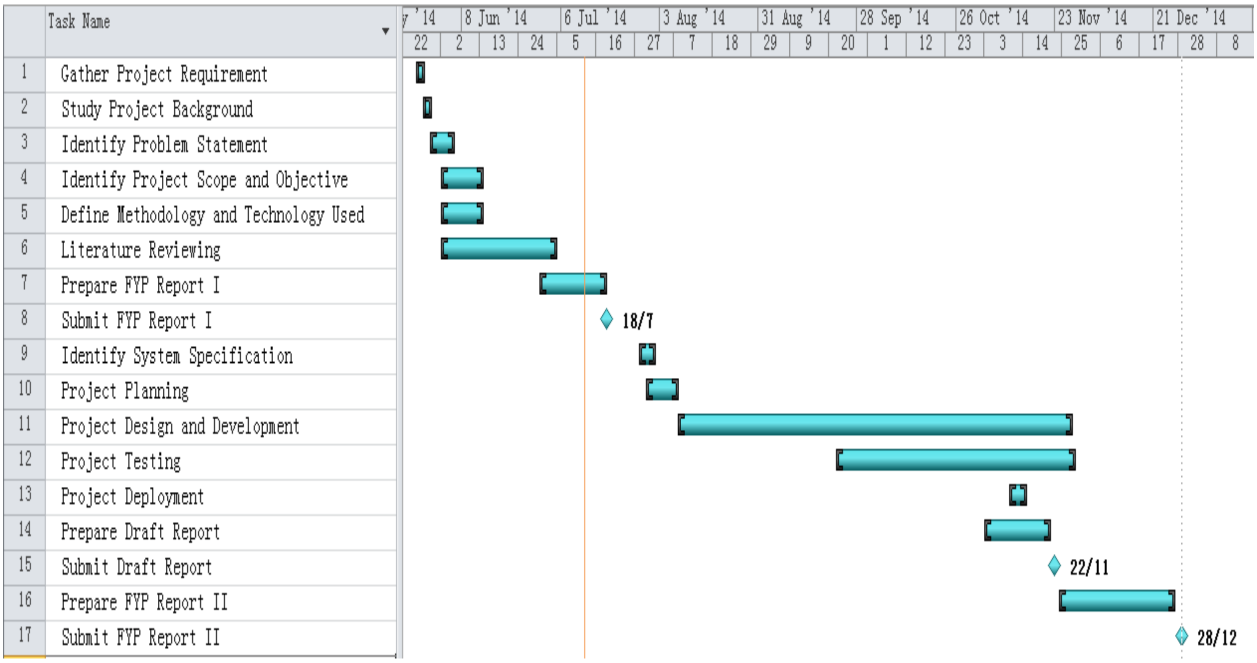
Figure 3.8 shows the graphical user interface when the question prediction and question generation processes are started while Figure 3.9 shows the graphical user interface of the application when the processes are done.

3.3) Implementation Issues and Challenges

Currently there is one critical implementation issues and challenge for the application program, which is, the internet connection. As discussed above, during the operating process of the system, the system is required to connect to the internet in order to surf and download the related article. This may lead to incompleteness of the collection process. The unsuccessful download of article will impede the system from continuing to perform the rest of the process.

3.4) Timeline

In order to complete the development of application program smoothly without any unwanted consequences, a Gantt chart is created to do the project scheduling to make sure all the required tasks are performed as scheduled. Figure 3.5 shows the Gantt chart for the application program.



	Task Name
1	Gather Project Requirement
2	Study Project Background
3	Identify Problem Statement
4	Identify Project Scope and Objective
5	Define Methodology and Technology Used
6	Literature Reviewing
7	Prepare FYP Report I
8	Submit FYP Report I
9	Identify System Specification
10	Project Planning
11	Project Design and Development
12	Project Testing
13	Project Deployment
14	Prepare Draft Report
15	Submit Draft Report
16	Prepare FYP Report II
17	Submit FYP Report II

Figure 3.10: Gantt Chart for application program

CHAPTER IV EXPERIMENT AND RESULT

4.1) Feature Extraction

As discussed at the former chapter, in order to facilitate the news classification, the initial set of measured data used to classify different article in Wiki Vital Article was extracted and built into informative values with no redundancy. In general term, the list of words, that used to define a category, was decomposed into feature. Table 4.1 shows the sample of the feature extraction result.

Keywords	Article Category
education, pollution, environment	Social Science
medical, illness, fitness	Health
pressure, emotion, love	Everyday Life
artificial intelligence, machine learning	Technology
history of, age of, ancient	History
culture, belief, religion	Culture
biology, astronomy, science	Science

Table 4.1: Sample of the feature extraction result

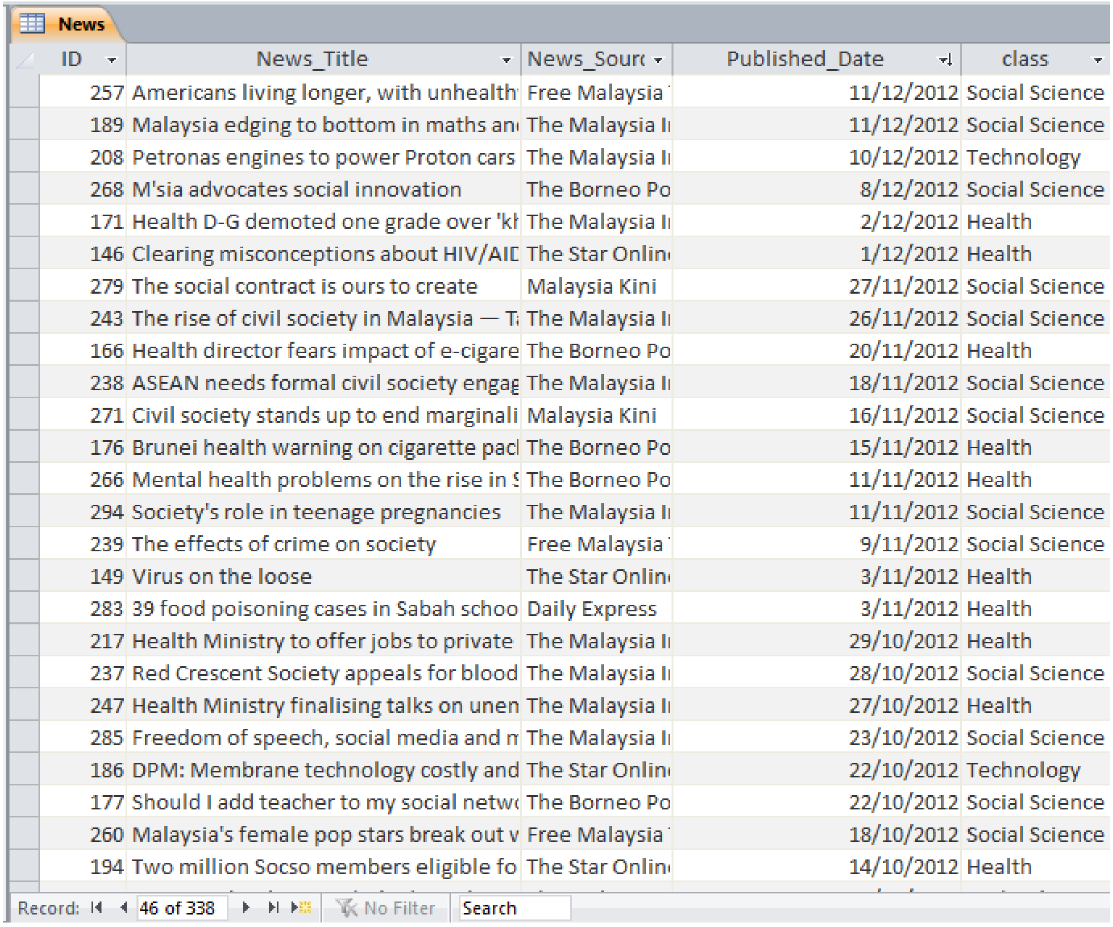
4.2) Training , Testing and Result

4.2.1) SVM Classification

4.2.1.1) Training and Testing

The SVM news classification was performed in R environment. The R script for running the SVM to perform classification will be described in Appendix B. After installed the required package, which is “RTextTools”, the labeled data set, the collection of keywords and article categories, were loaded into the R environment and converted into document term matrix, which is a mathematical matrix that presents the terms frequency that appear in the list of documents. This is because the mathematics properties of matrices will make the classifier to perform well.

Besides, in order to train the SVM model, the document term matrix was put into a container that can be used for training and classification. This is done by using the create_container methods in “RTextTools”. By setting the whole labeled data set as training set, the SVM model was trained with train_model method. The news data required to classify was then loaded into the R environment as test set. Same goes to the train data set, test set was also converted into document term matrix and put into another container. This container was then classified using the trained model with the help of classify_model method. Figure 4.1 shows the classified news.



ID	News_Title	News_Sourc	Published_Date	class
257	Americans living longer, with unhealth	Free Malaysia	11/12/2012	Social Science
189	Malaysia edging to bottom in maths an	The Malaysia I	11/12/2012	Social Science
208	Petronas engines to power Proton cars	The Malaysia I	10/12/2012	Technology
268	M'sia advocates social innovation	The Borneo Po	8/12/2012	Social Science
171	Health D-G demoted one grade over 'kl	The Malaysia I	2/12/2012	Health
146	Clearing misconceptions about HIV/AID	The Star Onlin	1/12/2012	Health
279	The social contract is ours to create	Malaysia Kini	27/11/2012	Social Science
243	The rise of civil society in Malaysia — T	The Malaysia I	26/11/2012	Social Science
166	Health director fears impact of e-cigare	The Borneo Po	20/11/2012	Health
238	ASEAN needs formal civil society engag	The Malaysia I	18/11/2012	Social Science
271	Civil society stands up to end marginali	Malaysia Kini	16/11/2012	Social Science
176	Brunei health warning on cigarette pacl	The Borneo Po	15/11/2012	Health
266	Mental health problems on the rise in S	The Borneo Po	11/11/2012	Health
294	Society's role in teenage pregnancies	The Malaysia I	11/11/2012	Social Science
239	The effects of crime on society	Free Malaysia	9/11/2012	Social Science
149	Virus on the loose	The Star Onlin	3/11/2012	Health
283	39 food poisoning cases in Sabah schoo	Daily Express	3/11/2012	Health
217	Health Ministry to offer jobs to private	The Malaysia I	29/10/2012	Health
237	Red Crescent Society appeals for blood	The Malaysia I	28/10/2012	Social Science
247	Health Ministry finalising talks on unen	The Malaysia I	27/10/2012	Health
285	Freedom of speech, social media and n	The Malaysia I	23/10/2012	Social Science
186	DPM: Membrane technology costly and	The Star Onlin	22/10/2012	Technology
177	Should I add teacher to my social netw	The Borneo Po	22/10/2012	Social Science
260	Malaysia's female pop stars break out v	Free Malaysia	18/10/2012	Social Science
194	Two million Socso members eligible fo	The Star Onlin	14/10/2012	Health

Figure 4.1: Classified News

4.2.1.2) Result

After the news classification, the result was manually compared with the question set for each year to find out correctness of the assumption made that mentioned in earlier section. Figure 4.2 shows the statistic data of the news in each year while Figure 4.3 shows the statistic of question topic that came out in each year.

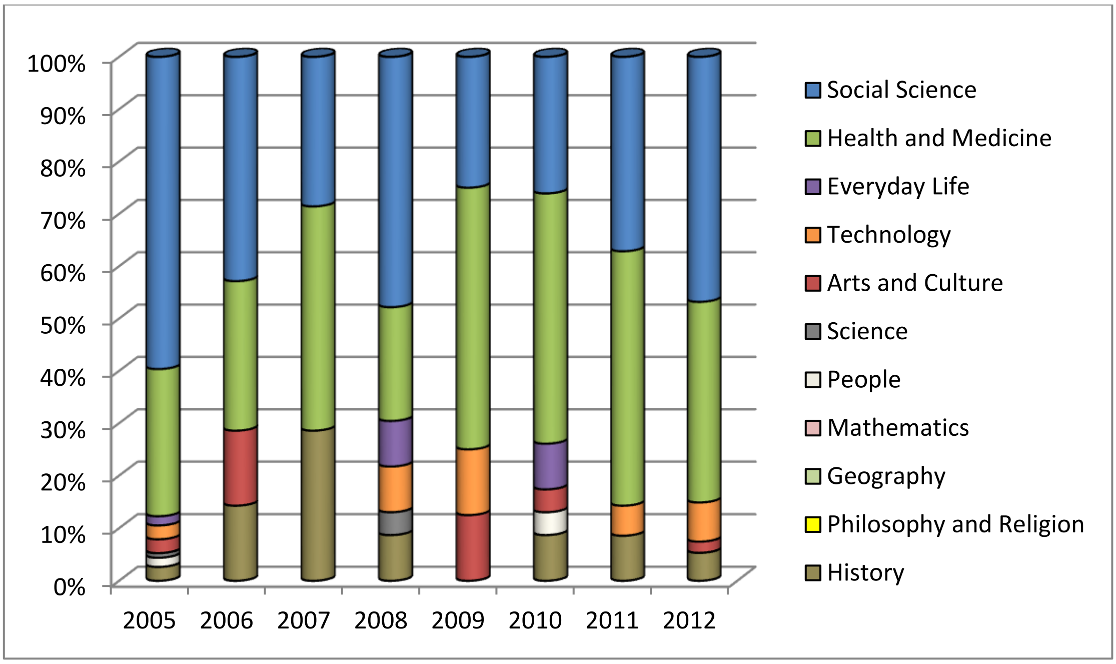


Figure 4.2: Statistic data of news in year 2005 to year 2012

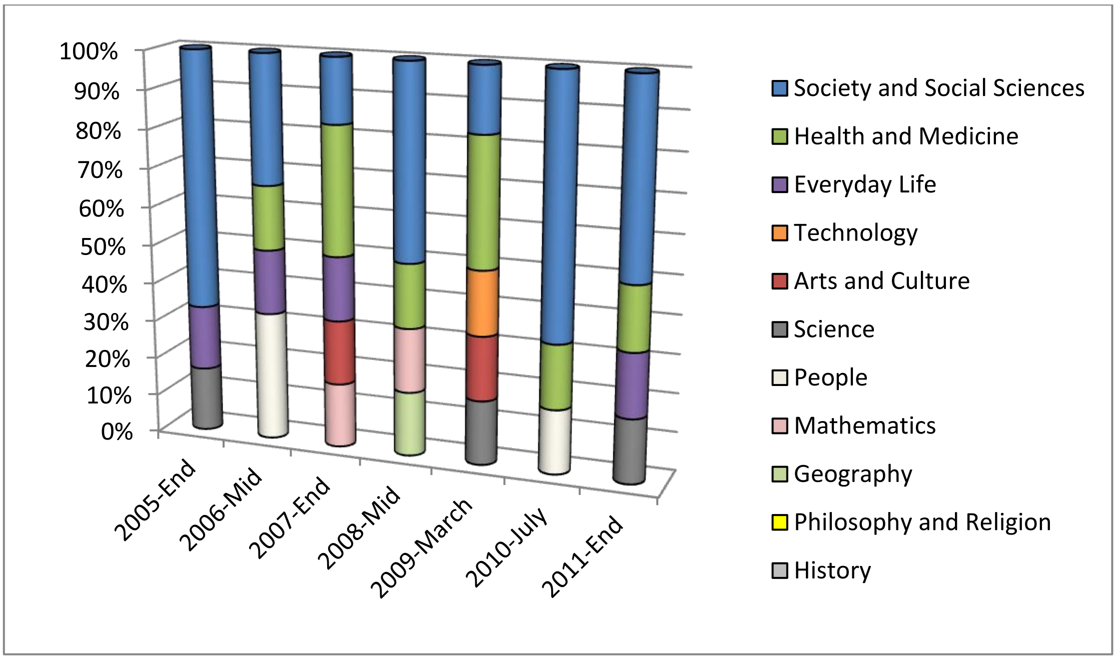


Figure 4.3: Statistic of question topic came out from year 2005 to 2011

Based on the statistic shown in Figure 4.2 and Figure 4.3, it is clear that the assumption was correct, which is, the topic of the question is dependent on the news during the exam period.

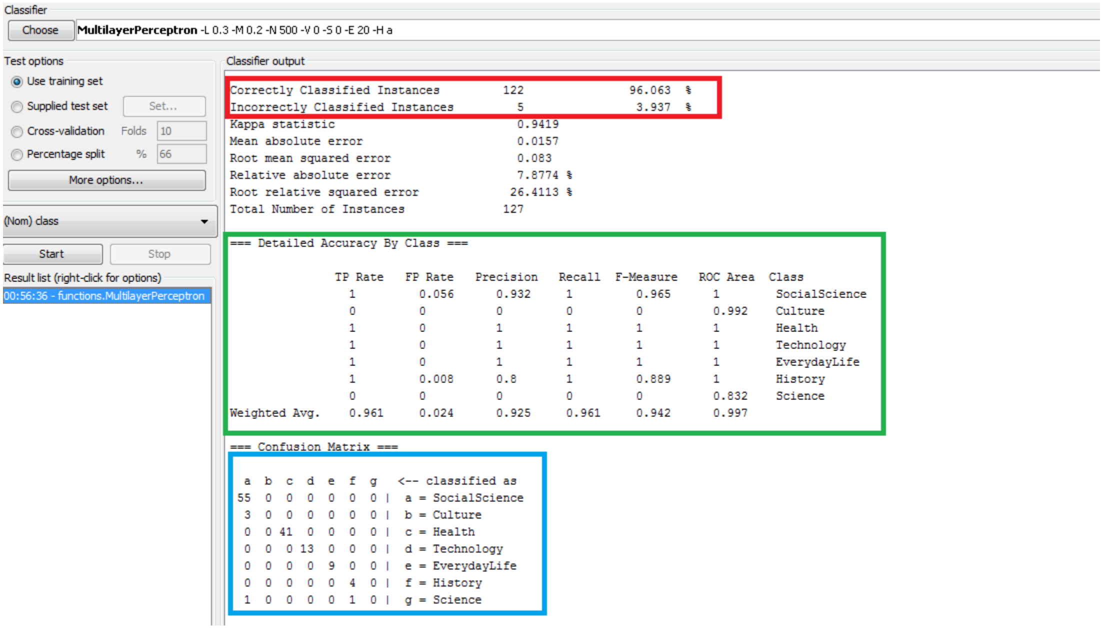
4.2.2) MLP Classification

4.2.2.1) Training and Testing

In order to programmatically find out the relationship between news and question, MLP classification was implemented. The MLP classification was performed in WEKA program. In this classification, in order to find out the question probability that will be come out in the coming exam, the features of news (collection of words) and question topic in the same year with the probability were set as the labeled training data, in which, the former two as the input of MLP while the latter as the output of the MLP. Table 4.2 shows the sample training data for year 2012 of MLP classification.

News features	Question Topic	Probability
suicide, punishment	Social Science	0.965
nursing, medical	Health	0.912
open burning, pollution	Social Science	0.997
fitness, nutrition	Health	0.943
mobile, web services	Technology	0.876
education, university	Social Science	0.961
ancient, islamic	Culture	0.832

Table 4.2: Sample train data for year 2012



Classifier: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options: ● Use training set ○ Supplied test set ○ Cross-validation Folds 10 ○ Percentage split % 66

(Nom) class

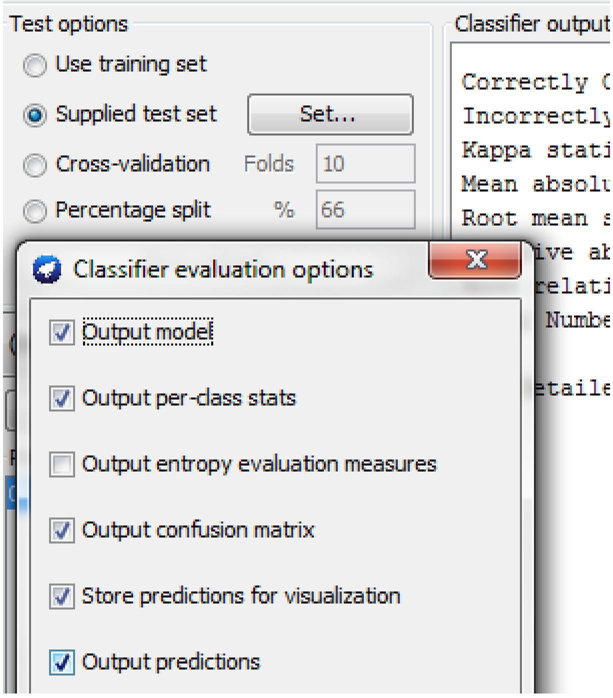
Result list (right-click for options): 00:56:36 - functions.MultilayerPerceptron

Classifier output

```
Correctly Classified Instances         122               96.063  %
Incorrectly Classified Instances         5                3.937  %
Kappa statistic                          0.9419
Mean absolute error                      0.0157
Root mean squared error                  0.083
Relative absolute error                  7.8774 %
Root relative squared error             26.4113 %
Total Number of Instances              127

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.056      0.932     1         0.965      1        SocialScience
                 0         0          0         0         0          0.992    Culture
                 1         0          1         1         1          1        Health
                 1         0          1         1         1          1        Technology
                 1         0          1         1         1          1        EverydayLife
                 1         0.008      0.8       1         0.889      1        History
                 0         0          0         0         0          0.832    Science
Weighted Avg.    0.961     0.024      0.925     0.961     0.942      0.997

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 55  0  0  0  0  0  0 |  a = SocialScience
  3  0  0  0  0  0  0 |  b = Culture
  0  0 41  0  0  0  0 |  c = Health
  0  0  0 13  0  0  0 |  d = Technology
  0  0  0  0  9  0  0 |  e = EverydayLife
  0  0  0  0  0  4  0 |  f = History
  1  0  0  0  0  1  0 |  g = Science
```

Figure 4.4: The Screen shot of WEKA MLP implementation

Figure 4.4 shows that the implementation of MLP classification in training the train dataset. As referring to Figure 4.4, the correctly classifies instances of this training set gained 96.043% of score, which shows that the MLP model is properly trained. While in the green box and blue box portion in Figure 4.4 shows the detailed accuracy of each class and the confusion matrix of the training set.

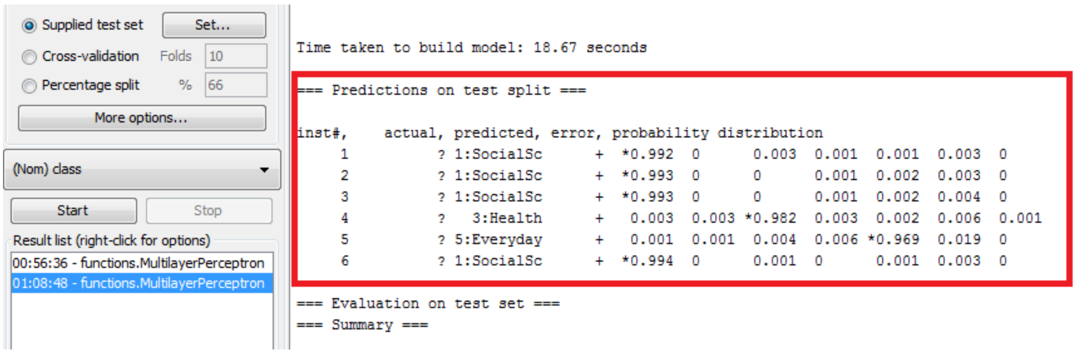


Test options: ○ Use training set ● Supplied test set ○ Cross-validation Folds 10 ○ Percentage split % 66

Classifier evaluation options

☑ Output model
☑ Output per-class stats
□ Output entropy evaluation measures
☑ Output confusion matrix
☑ Store predictions for visualization
☑ Output predictions

Figure 4.5: Configuration on the test data in WEKA



● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) class

Result list (right-click for options)
00:56:36 - functions.MultilayerPerceptron
01:08:48 - functions.MultilayerPerceptron

```
Time taken to build model: 18.67 seconds

=== Predictions on test split ===

inst#,    actual, predicted, error, probability distribution
     1          ? 1:SocialSc      +  *0.992  0      0.003  0.001  0.001  0.003  0
     2          ? 1:SocialSc      +  *0.993  0      0      0.001  0.002  0.003  0
     3          ? 1:SocialSc      +  *0.993  0      0      0.001  0.002  0.004  0
     4          ?   3:Health      +   0.003  0.003 *0.982  0.003  0.002  0.006  0.001
     5          ? 5:Everyday      +   0.001  0.001  0.004  0.006 *0.969  0.019  0
     6          ? 1:SocialSc      +  *0.994  0      0.001  0      0.001  0.003  0

=== Evaluation on test set ===
=== Summary ===
```

Figure 4.6: The result of the MLP Classification

Besides, by loading the testing data in to WEKA environment and configure the test option to let the MLP model to output the prediction as shown in Figure 4.5, the outcome of the testing process is shown in Figure 4.6. The red box portion in Figure 4.6 shows the predictions on test split, in which, the prediction score was printed under each class.

4.2.2.2) Result

To predict the question in the coming exam, the MLP model of year 2015 will be used. The output probability will be sorted in descending order as the high score of output represent the highest probability of the topic that will be come out.

CHAPTER V: CONCLUSION

5.1) Conclusion and Future Work

In this report, the MUET question prediction and generation application program is presented. This application program is developed with the purpose of solving the problem as discussed at the former section, which is, to assist student for preparing their MUET exam revision and relieve English teachers from the routine and tedious work of preparing mock exam questions.

The application program will perform a set of operations in order to achieve the objective, which include, question prediction, question generation and question ranking. Apart from that, in order to improve the accuracy of predicted result, pass 8 years of MUET question sets are collected and further analyzed. From the analysis, every article within the question set is classified into several article categories which have been discussed at the section III. The Support Vector Machine (SVM) was implemented to perform the news classification process. Apart from this, Multilayer Perceptron (MLP) also applied to predict the upcoming question topic.

The application program will first wait for the user's response by providing a single button. Once the button was being clicked, the system will start perform the prediction. After the topic has been predicted, the system will proceed and connect to internet to find and download the related article for question generating process. The question generation action is performed once the related article has been downloaded and followed by the question ranking process. After the generated questions are being ranked, the application program will display to the user for review and study.

There is an implementation issue that has been faced by this application program which is internet connection. The internet connection will affect and prevent the system from downloading related article and this may directly stop the system from proceeding to the rest of the processes and actions.

Another problem encountered in this project was that the difficulty faced in obtaining past MUET questions. These questions are not available publicly on the Internet. It is hard to collect the whole complete set of question. One has to buy the related books and manually input these questions into electronic format. Even then, the content of these books are often incomplete.

Moreover, while the main objective was focusing on predict, generate and rank the question, the system is having poor user interface and not able to generate the possible answers for users especially students. In the future, the system can be further enhanced to generate question together with the answer to allow users to check their answer after they finish attempting the generated questions. Besides, the system can be further designed and developed to be more users friendly.

In addition, as the system has the ability to predict the trend of upcoming question topic, with proper enhancement in the future, it might extend the functionality to predict other kind of question trends, says O-level, A-level or even SPM and STPM question trends.

REFERENCE

Baron, M (2007). *Probability and Statistics for Computer Science.* London.

Bredan O'Connor et al (2013) ARKref: a rule-based coreference resolution system

Chong. (2007, July 28). *Malaysian University English Test (MUET) Guide*. Retrieved from Malaysia-Student: http://www.malaysia-students.com/2007/07/malaysian-university-english-test-muet.html

Crammer. K et al (2006) Online Passive-Aggression Algorithms

Donghui. F (2007) Factorizing Information Extraction from Text Corpora

Eugene. A. et al, 2000. *Snowball*: Extracting Relations from Large Plain-Text Collection

Herbert S (1998) C++ The Complete Reference, 3rd edn. Osborne McGraw-Hill, Sydney

Kaow, N. Y (2011). *Multilayer Perceptron.* Malaysia.

Kunichika. H et al (2003) Automated Question Generation Methods for Intelligent English Learning Systems and its Evaluation

Luxton-Reilly, A. (2012). The Impact of Question Generation Activities on Performance

Otterbacher. J et al (2005) Using Random Walks for Question-focused Sentence Retrieval

Rus Vasile et al (2010). The first question generation shared task evaluation challenge. In Proceedings of the Sixth International Natural Language Generation Conference (pp. 251-257)

Smith Ross (2005) "Global English: gift or curse?" English Today 21 (2): 56

Yang, D. H. (n.d.). *What is Java, JVM, JRE and JDK?* Retrieved from Herong's Tutorial Examples: http://www.herongyang.com/Windows/Java-What-Is-Java-JVM-JRE-JDK.html

Appendix A

Meaning of the Word

This appendix will describe each word in detail. In this project, in order to extract sentence that comprises the combination of different components and structures, several specific terms are used to describe each component. The following are the terms that used in Question Generation (QG) process.

1. Semantic entailment

   The semantic entailment of a sentence indicates the deepest message that conceal inside that sentence. For instance, *Mason lives in Kuala Lumpur* entails *Mason lives in Malaysia*. However, it is not true in the reverse way. It is not the case that *Mason lives in Malaysia* entails *Mason lives in Kuala Lumpur*. Sometimes, an entailment can be a connection between set of sentences. For instance, *Alan washed his car happily.* From this sentence, some entailment might be: *(a) Alan owns a car; (b) Alan likes to wash his car;* and *(c) Alan's car was being washed*.

2. Presupposition

   The presupposition of a sentence is an implicit assumption or prediction about the background story of that sentence. Normally, a sentence's presupposition relates to a person whose fact or truth is taken for granted in discourse. For instance, D*o you still sleeping?* has the presupposition of *the person had slept for some time*.

3. Adjunct modifiers

   Adjunct modifiers can be described as a part of sentences that can be removed without making another sentence ungrammatical. In other words, adjunct modifiers used to descript another sentence to make it more understandable, but will not make the sentence become weird even after being removed. For example, *John is driving a car, which belongs to his father, to Ipoh*. The adjunct modifier of this sentence is "*which belongs to his father"*. It makes thelistener or reader to know that *the car being drove by John is under his father's property*.

4. Discourse connective

   Discourse connective is a word or a set of sentences that make a set of sentences or paragraph connected to each other. The examples of popular discourse connective that usually will be used in paragraph are *however, nevertheless, on the other hand, apart from that* and *etc*.

5. Conjunctions

   Usually, conjunction is part of sentences that connects two words or sentences. It is different as discourse connective. It is an invariable grammatical particle which may or may not stand between the items it conjoins. For example*, and, or, but, because, since, so* and etc.

Appendix B

R script for performing classification

```
#install packages

install.packages(RTextTools,dependencies=TRUE)

require("RTextTools")

# set the directory

dataDirectory <- "C:\\Users\\JPCHOI\\Desktop\\Marchine Learning\\"

# load data

data <- read.csv(paste(dataDirectory, 'features.csv', sep=""),header=TRUE)

# Create document term matrix

dtMatrix <- create_matrix(data["Keyword"])

# Configure training data

container <- create_container(dtMatrix, data$Class, trainSize=1:135, virgin=FALSE)

# train SVM Model

model <- train_model(container, "SVM", kernel="linear", cost=1)

# load prediction data

predictionData <- read.csv(paste(dataDirectory, 'sample.csv', sep=""), header=FALSE)

# create prediction document term matrix

predMatrix <- create_matrix(predictionData, originalMatrix=dtMatrix)

# create container for the prediction data

predSize = length(predictionData)

predictionContainer <- create_container(predMatrix, labels=rep(0,predSize),
testSize=1:predSize, virgin=FALSE)
```

```
# predict

results <- classify_model(predictionContainer, model)
```