

Islamic University of Gaza
Deanery of Higher Studies
Faculty of Information Technology
Information Technology Program



الجامعة الإسلامية - غزة
عمادة الدراسات العليا
كلية تكنولوجيا المعلومات
برنامج تكنولوجيا المعلومات

Filtering Features and Recommender Model For Messaging System of MOODLE FRMM

Prepared by

Rasha R. Atallah
220100007

Supervised by

Dr. Tawfiq S. Barhoom

**A Thesis Submitted in Partial Fulfillment of the Requirements for the
Master Degree of Science in Information Technology**

2014 م — 1435 هـ

بسم الله الرحمن الرحيم

{ وقل رب زدني علماً { سورة طه 114 }

{ اقرأ باسم ربك الذي خلق (1) خلق الإنسان من علق (2) اقرأ وربك الأكرم (3) الذي علم بالقلم (4) علم الإنسان ما لم يعلم (5) { سورة العلق }

{ الله الذي خلق السموات والأرض وما بينهما في ستة أيام ثم استوى على العرش ما لكم من دونه من ولي ولا شفيع أفلا تتذكرون (4) يدبر الأمر من السماء إلى الأرض ثم يفرج إليه في يوم كان مقداره ألف سنة مما تعدون (5) { سورة السجدة }

صدق الله العظيم

ACKNOWLEDGMENT

*First, I thank and I'm grateful for **Allah** for taking care of me and protecting me all the time.*

I would Thanks everyone who helped me in reaching this stage

***My Father and My Mother** thank you for everything you have done to me my entire life*

*My supervisor **Dr. Tawfiq S. Barhoom** for his encouragement and support, without his effort I couldn't finish my thesis,*

*My Brothers (**Hossam, Mohanned**) & my sisters (**Rana, Haneen, Randa**) for their love*

I would like to express my appreciation to the academic staff of information technology program at the Islamic University-Gaza.

***Dr.Iyad Al Agha** for helping me in writing my thesis basis*

***Eng. Azmi Al Saqqa, Mr. Ahmed Abu mosameh, Miss Sarah Kohail, Miss Najwa Baraka, Mr. Ezz Moshtaha, Mr. Raef Baraka , Mr. Ahmed Deeb** for their helping in finishing my work*

Special Thanks for Mr. Mahmoud Abu Ghosh

***Dr. Salem Sabah** the President of the University of Palestine who gave me the permission to applicant my thesis at the University*

*My Friends **Rana Atalla** and **Shireen Abd Al Rahman** to encourage me*

ABSTRACT

E-learning systems are powerful way for enriching learning and teaching process. With the spread of this culture, there is a need for Learning management system as MOODLE, aTutor , canvas and Sakai.

MOODLE allows students and instructors to interact easily and effectively in various ways, especially via private messaging system. In order to make this feature more convenient for both students and instructors, this research proposed a model that integrated within messaging system of MOODLE to filter unwanted messages, recommend a material for students, and to help them finding answers to their questions which are founded on the message.

This new model will make it easier for students to find information and materials they are looking for, or asking about and at the same time it will keep instructors' mailboxes clean of any unwanted irrelevant messages. The model applied at University of Palestine on six courses with various natural of materials , some of these materials are isolated topics ,each chapter's topic , in this type of courses the accuracy of the system was high. On the other hand, there are some courses in which their topics are integrated with each other ; in these type of courses, the accuracy is less than the other type of materials , but in general the whole system's accuracy is 96.66% and the reliability of the system is 92.55%.

Finally, the model is founded to improve using message in the MOODLE, and the model actually solved the problem of this research.

KeywordsE-learning, Message system in MOODLE, Filter message, Information retrieval

عنوان البحث:

نموذج بخاصية الفرز والتوصية لنظام الرسائل في المودل

الخلاصة

أنظمة التعليم الإلكتروني تشكل أداة فعالة لتعزيز طرق وإجراءات التعليم. و لهذا السبب، أصبحت الحاجة ملحة إلى أنظمة إدارة التعليم الإلكتروني كالمودل وغيره. نظام المودل يسمح للطلاب والمعلمين للتواصل بسهولة وبفاعلية بطرق متعددة، وعلى وجه الخصوص، ميزة الرسائل الخاصة. وفي سبيل جعل هذه الميزة أكثر ملاءمة لبيئة التعليم، يجب فلترة الرسائل الغير مرغوب فيها واقتراح مواد تعليمية للطالب بناء على أسئلته المطروحة لمساعدته في ايجاد الإجابات بسهولة ودقة في حال انشغال المعلم عن إجابته أسئلته. وفي هذا الإتجاه، هذا البحث يساهم في تعزيز هذه العملية وحل المشكلات التي تواجه تطبيق نظام فلترة دقيق ومثالي لنظام الرسائل القصيرة في المودل. تم تطبيق البحث على نظام المودل الخاص بجامعة فلسطين لست مساقات مختلفة وكانت نتائج النظام المقترح مرضية على مستوى الدقة والإعتمادية بنسبة مئوية 96.66% و 92.55% على التوالي.

الكلمات الإفتتاحية : التعليم الإلكتروني , نظام الرسائل في المودل , نظام فرز الرسائل , استرجاع

البيانات

TABLE OF CONTENTS

<i>ACKNOWLEDGMENT</i>	vii
<i>English ABSTRACT</i>	viii
<i>Arabic ABSTRACT</i>	ix
<i>TABLE OF CONTENTS</i>	x
<i>LIST OF FIGURES</i>	xii
<i>LIST OF TABLES</i>	xiii
<i>LIST OF ABBREVIATIONS AND GLOSSARIES</i>	xiv
<i>LIST OF APPENDICES</i>	xv
CHAPTER 1 INTRODUCTION	1
1.1 Overview for E-Learning.....	1
1.2 Problem Statement.....	2
1.3 Objectives.....	3
1.3.1 Main Objective	3
1.3.2 Specific Objective	3
1.4 Scope and limitation	3
1.5 Research Methodology.....	4
1.6 Thesis Overview.....	4
CHAPTER 2 TECHNICAL FOUNDATIONS	5
2.1 Pre-processing.....	5
2.1.1 Parsing.....	5
2.1.2 Stop words	6
2.1.3 Stemming	6
2.1.4 The Porter Stemming Algorithm	7
2.2 Information retrieval.....	8
2.2.1 Vector Space Model	10
2.3 Evaluation Method	12
2.3.1 Independent T-Test	12
2.3.2 Cronbach's alpha	12
2.4 Conclusion	13
CHAPTER 3 RELATED WORK	14
3.1 MOODLE as Learning Management System.....	14
3.2 Filter Message Techniques.....	15
3.3 Recommender system	16
3.4 Information retrieval.....	17
3.5 Conclusion.....	20

CHAPTER 4 FILTERING FEATURES AND RECOMMENDER MODEL FOR MOODLE MESSAGING SYSTEM (FRMM)	21
4.1 The Proposed Model FRMM.....	21
4.3 Input the data to the system	28
4.4 Message Pre-processing.....	28
4.5 Taboo words	28
4.5 The Agent for recommender.....	29
4.6 Examples on how FFRM works:	30
4.7 Conclusion	33
CHAPTER 5 EXPERIMENTAL RESULTS AND EVALUATION	34
5.1 Implementation	34
5.2 The users of the system.....	34
5.3 Experimenting Setup	34
5.3.1 Experimental Environment and Tools.....	35
5.3.2 Experimental Procedures	35
5.4 The Material's Nature	35
5.5 System Evaluation:	36
5.5.1 Accuracy Calculation:	36
5.5.2 Using SPSS to calculate the reliability:	37
5.6 The Result Analysis	38
5.6.1 Computer Skill:.....	38
5.6.2 Management Information System.....	38
5.6.3 System Analysis and Design Course	39
5.6.4 Introduction to Programming Course.....	39
5.6.5 Algebra Course.....	40
5.6.6 Software Engineer Course.....	40
5.7 Conclusion.....	41
CHAPTER 6 CONCLUSION AND FUTURE WORK	42
6.1 Conclusion.....	42
6.2 Future Work	43
REFERENCES	44
Appendix A: Students' Questions	47
Appendix B: About University Of Palestine	55
Appendix C : Java Code	58

LIST OF FIGURES

<i>Figure 4. 1 : Explains the structure of the MOODLE and where the FRMM MODEL integrated Model</i>	22
<i>Figure 4. 2 : Explains the model of the System</i>	24
<i>Figure 4. 3 : The proposed FRMM flow chart</i>	27
<i>Figure 4. 4 : Explain the process of searching taboo word on the message Search for Question Mark "?"</i>	28
<i>Figure 4. 5 : Explain the process of searching about the Question mark "?"</i>	29
<i>Figure 4. 6 : An example shows a message which contains taboo word</i>	30
<i>Figure 4. 7 : an example shows a message which contains question mark "?"</i>	30
<i>Figure 4. 8 : An example explains a message that doesn't contain taboo word or question mark "?"</i>	31
<i>Figure 4. 9 : shows a new message which is forward to the academic affair secretary and it is formatted</i>	31
<i>Figure 4. 10 : show the rank for all PowerPoint files in the system</i>	32
<i>Figure 4. 11 : show the chapter's name which send to the student</i>	32
<i>Figure 5. 1 : It shows the T-independent test method</i>	37
<i>Figure 5. 2 : It shows the result</i>	37

LIST OF TABLES

<i>Table 1 : Summaries the process of the model.....</i>	<i>24</i>
<i>Table 2 : the figure shows the system's accuracy.....</i>	<i>36</i>
<i>Table 3 : The table shows 45 questions for programming 1 course sent by students</i>	<i>47</i>
<i>Table 4 : The table shows 45 questions for Computer Skill course sent by students</i>	<i>48</i>
<i>Table 5 : The shown table 45 questions for Algebra course sent by students.....</i>	<i>49</i>
<i>Table 6 : The table shown 45 questions for Management Information System course sent by students.....</i>	<i>50</i>
<i>Table 7 : The table shown 45 questions for System Analysis and Design course sent by students</i>	<i>52</i>
<i>Table 8 : The table shown 45 questions for Software engineer course sent by students</i>	<i>53</i>
<i>Table 9 : Examples for message which contains taboo words</i>	<i>54</i>

LIST OF ABBREVIATIONS AND GLOSSARIES

- **MOODLE** : *Modular Object-Oriented Dynamic Learning Environment*
- **LMS**: *Learning Management system*
- **E-learning** : *Electronic learning*
- **VSM** : *Vector space Model*
- **FRMM**: *Filter features and recommender model for message system of MOODLE*
- **TFIDF** : *Term Frequency – Inverse Document Frequency*

LIST OF APPENDICES

<i>Appendix A: Students' Questions</i>	47
<i>Appendix B: About University Of Palestine</i>	55
<i>Appendix C : Java Code</i>	58

CHAPTER 1

INTRODUCTION

This chapter introduces the thesis, which explains the underlying concepts of e-learning, tools used at managing e-learning process especially MOODLE which applicant on the thesis ,also it talks about the thesis problem, the research objectives, the research importance, the research scope and limitations, as well as the research methodology.

1.1 Overview for E-Learning

Learning is “the process by which people acquire new skills or knowledge for the purpose of enhancing their performance” according to Rosenberg definition, when adding the letter ‘e’ for learning, it means electronic learning "e-learning" that also he defines it as “a networked phenomenon allowing for instant revisions and distribution”. In addition, it is delivered using standard Internet technology; it is used for delivering information and tools to improve performance [1].

E-learning has many benefits such as cost-effectiveness, enhanced responsiveness to change, consistency, timely content, flexible accessibility and provision of customer value. However, it needs a special architecture and various tools to support the concept of e-learning on the organizations as Learning Management System and Course Management System [2].

A Learning Management System (LMS) is a tool for administration, documentation, tracking, reporting and delivery of online education courses or training programs. Most LMSs are web-based made to facilitate both access to learning content and administration, and enhance and support classroom teaching and offering courses to a larger population of learners such as MOODLE, Alpha LMS, and Ziiva Prosperity™ Learning Management System [3].

The MOODLE (Modular Object-Oriented Dynamic Learning Environment) is a platform to help educators to create online courses with a focus on interaction and collaborative construction of content, and is in continual evolution. [4] The first version of MOODLE was released on 20 August 2002. It can be used in many types of environments such as in

education, training, development, and business settings. It has some typical features such as assignment submission, discussion forum, files download, grading system, and MOODLE instant messages. MOODLE is also an open source, which means that developers can always develop new features and modular according to their needs free-of-charge. [5]

The Instant Messaging refers to both: automatic alerts from MOODLE about new forum posts, a assignment submission notifications etc., and the conversations that uses the instant messaging feature.

This research focuses on the instant messaging that students can exchange with their instructors for asking about the course content purpose.

According to the University of Palestine's experience using the MOODLE as part of its learning system since 2006 (for more detail about the university see appendix B), the system administrator has repeatedly encountered a problem with messaging feature. A large number of unwanted messages are being sent to instructors, which often caused a blockage in the service. Consequently, the accuracy and reliability of the system decreased, especially when it comes to communications between instructors and students, therefore, to overcome this issue a filtering system needs to be established.

1.2 Problem Statement

One of the features offered by e-learning systems is messaging system, through this feature students and instructors can exchange message between each other, huge messages sent from students to instructors, which caused a big delay in the teacher response and requires extensive efforts from instructors.

The problem of this research is how to build a model to prevent message which contains taboo words and send recommended material to students according to their questions.

1.3 Objectives

The following sections present the main objective and the specific objectives of the thesis

1.3.1 Main Objective

The main objective is to build a model to add new features in MOODLE messaging system, learning management system, as filtering messages sent by students to check if the message contains taboo words. Also it allows students to receive automatic suggestions by the agent on the most relevant material, which corresponds with their questions depending on the material rank which is more suitable for students according to their question.

1.3.2 Specific Objective

The specific objectives of this research are as follows:

- Preprocess messages by using stop words, stemming, comparing, and tokenizing function.
- Design a function for filtering incoming messages according to taboo words, to keep instructors inbox clean from any unwanted messages.
- Design a function to send recommended material for students using automated messages to answer their questions according to the cosine model in vector space to rank material.
- Implement the proposed model.
- Evaluate the accuracy and the reliability of the model by matching between the request result of the system and the instructor result.

1.4 Scope and limitation

This work is expected to be developed under some constraints and limitations such as:

- Applying to six courses: Computer skills, Management information system, System analysis and design, Introduction to programming, Algebra, and software engineering.
- MOODLE's messaging system is the only e-learning system which this research deals with.
- Presenting the material course content by PowerPoint.
- The standard language for the material and the message is English language

- The standard format is only text format in the model.
- The message which is sent by student must contain question mark "?"
- Every message has only one question.
- Filtering taboo words depends on a list build on the model, can add any taboo word on it.
- The process of messages content taboo words, stop in the proposed model while the message forward to the academic affairs secretary.

1.5 Research Methodology

This research follows a methodology that consists of four stages as follows:

- ***Stating the theoretical foundation:*** In this stage, a deep reading, understanding and analyzing the recommended system, filtering message and the related works which were done in this field will be performed in order to form a clear orientation of the theoretical framework.
- ***Designing the model functions:*** In this stage, designing message filtering function, which contains taboo words, also designing a function for a recommended system of the student's questions are designed clearly.
- ***Building the model's functions:*** In this stage, the model is built using java language.
- ***Model evaluation:*** In this stage, the model evaluation done by using accuracy and reliability by comparing the result, which is done by the system and the instructor.

1.6 Thesis Overview

Chapter two is about the technical foundations, then chapter three is about the related work , after that chapter four defines in detail the system structure and the proposed approach , chapter five includes experiments, the used evaluation techniques and system results, finally chapter six presents conclusions and future works.

CHAPTER 2

TECHNICAL FOUNDATIONS

This chapter intends to give an overview about the technical foundations that are used for building FRMM model, starting from preprocessing the message as parsing, stopwords, stemming, then explaining the used method for information retrieval as vector space model and cosine similarity.

2.1 Pre-processing

To make FRMM model processes more efficient, the messages in the MOODLE and the PowerPoint files must be done in some preprocessing on the text as stemming, parsing and stopwords.

2.1.1 Parsing

The message will be pre-processed by using parsing which is “the process of analyzing a text, made of a sequence of tokens”, the token is a string of characters, categorized according to rules as a symbol (e.g., IDENTIFIER, NUMBER, COMMA), also there is tokenization which is “the process of forming tokens from an input stream of characters”[6].

Larger language features are built from the first six categories of tokens that can appear in Java programs. The EBNF writes one simple rule that captures this structure:

token \Leftarrow *identifier* | *keyword* | *separator* | *operator* | *literal* | *comment*

1. **Identifiers:** names the programmer chooses
2. **Keywords:** names which already exist in the programming language
3. **Separators** (also known as punctuators): punctuation characters and paired-delimiters
4. **Operators:** symbols that are operated in arguments and produced results.
5. **Literals** (specified by their type)
 - Numeric: Integer and Double
 - Logical: Boolean
 - Textual: Char and String
 - Reference: Null

6. Comments

- Line
 - Block
- Extended Backus–Naur Form (EBNF) is a code that expresses the grammar of a formal language, which describes an unordered list of EBNF rules. Each EBNF rule has three parts: a left-hand side (LHS), a right-hand side (RHS), and the character separating the two sides; reads this symbol as "is defined as". The LHS contains one (possibly hyphenated) word written in lower-case; it names the EBNF rule. The RHS supplies the definition associated with this name. It can include combinations of the four control Sequence, Choice, Option and Repetition [7].

2.1.2 Stop words

There are some words in the document that are frequently occurred but they are meaningless, we call them stop words, Words like IT, AND and TO can be found virtually in every sentence in English-based documents, these words make very poor index terms . Users are indeed unlikely to ask for documents with these terms. Moreover, these words make up a large fraction of the text of most documents. According to Francis and Kucera [8], the amount of information carried by these words is negligible. Consequently, it is usually worthwhile to ignore all stop word terms [9].

2.1.3 Stemming

Stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form—generally a written word form , for example, should identify the string "cats" (and possibly "catlike", "catty" etc.) as based on the root "cat", there are many Algorithms for stemming have been studied in computer science since the 1960s which differs in respect to performance and accuracy for example Porter algorithm , Lookup algorithm , Suffix-stripping algorithm , Lemmatization algorithm , Stochastic algorithm [10].

2.1.4 The Porter Stemming Algorithm

The Porter stemming algorithm (or ‘Porter stemmer’) is a process for removing the commoner morphological and inflexional endings from words in English. Its main use is as part of a term normalization process that is usually done when setting up Information Retrieval systems. It is widely used and became the de facto standard algorithm used for English stemming. The Porter Stemmer is a conflation Stemmer developed by Martin Porter at the University of Cambridge in 1980[11]. The Stemmer is based on the idea that the suffixes in the English language (approximately 1200) are mostly made up of a combination of smaller and simpler suffixes. This Stemmer is a linear step Stemmer. Specifically it has five steps applying rules within each step. Within each step, if a suffix rule matched to a word, then the conditions attached to that rule are tested on what would be the resulting stem, if that suffix was removed, in the way defined by the rule. For example such a condition may be, the number of vowel characters, which are followed by a consonant character in the stem (Measure), must be greater than one for the rule to be applied. . The Porter algorithm had become the standard for stemming English, and it hence provided a natural model for the processing of other languages.

The rules in the Porter algorithm are separated into five distinct phases numbered from 1 to 5. They are applied to the words in the text starting from phase 1 and moving on to phase 5. Further, they are applied sequentially one after the other as commands in a program. Thus, in the immediately following, we specify the Porter algorithm in a pseudo programming language whose commands take the form of rules for suffix substitution (as above). This pseudo language adopts the following (semi-formal) conventions:

- A consonant variable is represented by the symbol C which is used to refer to any letter other than a,e,i,o,u and other than the letter y preceded by a consonant;
- A vowel variable is represented by the symbol V which is used to refer to any letter which is not a consonant;
- A generic letter (consonant or vowel) is represented by the symbol L;
- The symbol #1 is used to refer to an empty string (i.e., one with no letters);
- Combinations of C, V, and L are used to define patterns;
- The symbol * is used to refer to zero or more repetitions of a given pattern;
- The symbol + is used to refer to one or more repetitions of a given pattern;

- Matched parenthesis are used to subordinate a sequence of variables to the operators * and +;
- A generic pattern is a combination of symbols, matched parenthesis, and the operators * and +;
- The substitution rules are treated as commands which are separated by a semicolon punctuation mark;
- The substitution rules are applied to the suffixes in the current word;
- A conditional if statement is expressed as ``if (pattern) rule" and the rule is executed only if the pattern in the condition matches the current word;
- A line which starts with a % is treated as a comment;
- Curled brackets are used to form compound commands;
- A selected rule with longest suffix" statement selects a single rule for execution among all the rules in a compound command. The rule selected is the one with the largest matching suffix.

Thus, the expression (C)* refers to a sequence of zero or more consonants while the expression ((V)*(C))* refers to a sequence of zero or more vowels followed by zero or more consonants which can appear zero or more times. It is important to distinguish the above from the sequence (V*C) which states that a sequence must be present, and that this sequence necessarily starts with a vowel, followed by a subsequence of zero or more letters, and finished by a consonant [12].

2.2 Information retrieval

The information retrieval (IR) process of data mining requires extraction of valid patterns and relationships in very large data sets automatically, a general definition of IR is, “Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information”, and the IR methods employed depend on the organization and type of data [13].

There are many IR strategies for retrieving relevant documents, each retrieval strategy incorporate a specific model for its document representation purposes, the models are categorized according to two dimensions: the mathematical basis and the properties of the model. [14]

- **First dimension: Mathematical Basis**
- Set-theoretic models represent documents as sets of words or phrases. Similarities are usually derived from set-theoretic operations on those sets. Common models are:
 - Standard Boolean model
 - Extended Boolean model
 - Fuzzy retrieval
- Algebraic models represent documents and queries usually as vectors, matrices, or tuples. The similarity of the query vector and document vector is represented as a scalar value.
 - Vector space model
 - Generalized vector space model
 - (Enhanced) Topic-based Vector Space Model
 - Extended Boolean model
 - Latent semantic indexing
- Probabilistic models treat the process of document retrieval as a probabilistic inference. Similarities are computed as probabilities that a document is relevant for a given query. Probabilistic theorems like the Bayes' theorem are often used in these models.
 - Binary Independence Model
 - Probabilistic relevance model
 - Uncertain inference
 - Language models
 - Divergence-from-randomness model
 - Latent Dirichlet allocation
- Feature-based retrieval models view documents as vectors of values of feature functions (or just features) and seek the best way to combine these features into a single relevance score, typically by learning to rank methods. Feature functions are arbitrary functions of document and query, and as such can easily incorporate almost any other retrieval model as just a yet another feature.[15].

▪ **Second dimension: model Properties**

- Models without term-interdependencies treat different terms/words as independent. This fact is usually represented in vector space models by the orthogonally assumption of term vectors or in probabilistic models by an independency assumption for term variables.
- Models with immanent term interdependencies allow a representation of interdependencies between terms. However the degree of the interdependency between two terms is defined by the model itself. It is usually directly or indirectly derived (e.g. by dimensional reduction) from the co-occurrence of those terms in the whole set of documents.
- Models with transcendent term interdependencies allow a representation of interdependencies between terms, but they do not allege how the interdependency between two terms is defined. They rely on an external source for the interdependency degree between two terms. (For example a human or sophisticated algorithms.)

2.2.1 Vector Space Model

Vector space model recognizes the limitations of using binary weights and builds partial matching framework (achieved through giving the query and index term non-binary weights.). The weights of these words are stored in the computer system and used to calculate the similarity between each document and user queries. Vector space model can order the similarity of selected files in descending order and solve the partial matching problem. The model makes the retrieval result set much more precise than those got by Boolean model [16]. VSM model has two major advantages. First of all, the weights calculation of key words in the VSM model has used a multiple function and similarity weight can fluctuate between 0 and 1, which makes the searching engine system used the VSM model, which has the capacity of partial matching queries and relevance ranking. Although the VSM model makes a huge improvement compared to the Boolean model, it has obvious shortcomings undeniable [17].

First, the VSM model takes too much emphasis on valuing the correlation and the calculation. But the correlation is difficult to accurately describe and measure in fact. It is more dependent on the users feeling of the result list. Cosine Law is not necessarily the best measurement method. Second, the value selection of vector dimension plays an important role to the computer arithmetic labor, because the correlation calculation has been transformed to the cosine operation of two multi-dimensional vectors in the VSM model. There is no perfect method to determine the dimension of the vectors

If the dimension is too small, the computation can be reduced, but the accuracy is low. On the other hand, increasing the dimension will derive high accuracy but enormous computation amount. Finally, the words are mutually independent between each other in the VSM model. In fact, it exists similarity relation, classification relation and correlation between words, of which the correlation is most complex. And the existence of polysemy also brings damage to the relevance measurement. Larger documents would be beneficial due to more words and they are more likely to be hunt than small ones. At the same time, they often use repeated words, so their word frequency is higher than the small document.

Although the vector space model has the above deficiencies, it is also considered to be a good measure model of relativity calculation.

The vector space model involves two main phases as follows:

1. Calculate the weight
2. Calculate the cosine similarity.

The weight W_{ij} is calculated using the term frequency/inverse document frequency (TFIDF) with normalized frequency as shown in equation (1).

$$w_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}} * \log\left(\frac{D}{d_i}\right) \rightarrow 1$$

Where $f_{i,j}$ denotes the frequency a term i occurs in document j . The $\max_z f_{z,j}$, is the maximum frequency among all the keyword z that appears in document j . The D is the total number of documents that can be recommended to the learners. The d_i is the number of documents that contains term i .

The weight obtained from equation (1) is used to calculate the similarity value between the two items. The cosine similarity value is defined as follows

$$\cos(\vec{Wc}, \vec{Ws}) = \frac{(\vec{Wc} \cdot \vec{Ws})}{\|\vec{Wc}\| \|\vec{Ws}\|}$$

Where \vec{Wc} and \vec{Ws} are treated as a vector of content based profile of user c and content of document s . Both $\|\vec{Wc}\|$ and $\|\vec{Ws}\|$ are the magnitude of the vector \vec{Wc} and \vec{Ws} .

2.3 Evaluation Method

The selection methods which use to evaluate the system were measure the accuracy and the reliability .The accuracy is the degree of quantity measurements closeness that quantities actual (true) value. The system's reliability is the probability, over a given period of time, that the system will correctly deliver services as expected by the user[18].

2.3.1 Independent T-Test

The independent t-test, also called the two sample t-test or student's t-test, is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups[19].

2.3.2 Cronbach's alpha

Cronbach's alpha was developed by Lee Cronbach [20] in 1951 to provide a measure of the internal consistency of a test or scale; it is expressed as a number between 0 and 1. Internal consistency describes the extent to which all the items in a test measure the same concept or construct and hence it is connected to the inter-relatedness of the items within the test. Mathematically, reliability is defined as the proportion of the variability in the responses to the survey that is the result of differences in the respondents.

The basic equation for alpha used in the system will be:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum V_i}{V_{test}} \right)$$

- n = number of questions
- V_i = variance of scores on each question
- V_{test} = total variance of overall scores (not %'s) on the entire test

2.4 Conclusion

Through this chapter, the technical foundations of the model was explained, start from the preprocess process on the messaging as parsing, stopwords, stemming, also the algorithm of information retrieval which used to rank the material .The FRMM model used stopwords ,stemming for reach to the main words in the message and increase the performance of the system ,also it use vector space model to retrieve the highest rank of the PowerPoint files which related to the question . The selection methods which use to evaluate the system were measure the accuracy and the reliability

CHAPTER 3

RELATED WORK

This chapter introduces the presence of related works in the e-learning environment, learning management system and filter message techniques. As known there is more than one way to filtering the message according to various criteria. Also there are many types of recommender system in many field. After that this chapter introduces some related work in information retrieval.

3.1 MOODLE as Learning Management System

The MOODLE Architecture and comparative study of MOODLE proposed in [21], thus the paper discuss comparisons between different virtual learning management systems and present some authentication plug-in that MOODLE supports. The open source learning management, LMS MOODLE has been adopted by many people and organizations around the world because it offers a tightly integrated set of tools said to be designed from a social constructive perspective. MOODLE is composed from independent modules; plug-ins. In order to ensure better understanding of a whole MOODLE architecture, these modules will be presented in groups according to their purpose or use. From this perspective, there are six groups of modules are Communication modules , Productivity modules , Student involvement modules , Administration modules , Course delivery modules and Curriculum design modules

It is comparative study between MOODLE and other e-learning systems as (Esire2 learn8.1, KEWL, ANGELe7.1, e College, The Blackboard, Clarolin, Dokeos, OLAT, and Sakai) and this was based on two kinds of comparison. The first phase was based on the Architecture of MOODLE and other e-learning system, and the second phase was based on the technical aspects of MOODLE and other e-learning system. According to the previous research there were 32 features in Learning Management System, were 30 one available in the MOODLE .So, we found the MOODLE suitable for the work of our research.

Communication modules in MOODLE as message, chat and forum used to communicate with users.

The messaging system is one of communication module in MOODLE; user through it can send messages for other users. There is a history record for messages, which sent to/from a person.

3.2 Filter Message Techniques

Many researchers used the filter techniques, which have several different methods and uses to be suitable, their work, this section display some of these techniques.

A new methods for efficient spam detection and filtering in [22]. The architecture of the method deployed the filtering at IM client, server and gateway sides to eliminate spam more effectively, also it is integrates spam filtering techniques, including White/Black List, challenge-response based filtering, content-based technique, and collaborative feedback based filtering. More importantly, also based on filtering based on limiting IM sending rate and server/gateway side filtering .Also when the server/gateway receives a message, it will compare the sender's current sending rate with the global sending rate threshold and the receiver's personalized sending rate threshold , the third layer is Content based Filtering , Similar to e-mail spam, the collaborative user feedback based filtering is an effective mechanism to collect spam information, such as Black List or known spam database. This information can be distributed throughout the network and help other IM users filter out the spam messages.

Another technique based on intelligent spam filtering techniques [23] in the SMS (Short Message Service) text paradigm, in the context of mobile text message spam. It compares some of the popular spam filtering techniques on a publically available SMS spam corpus, to identify the methods that work best in the SMS text context. It depends on collection of SMS with over 5000 separate text messages of which about 15% are spam messages, and the SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. They use WEKA use of collection data mining algorithm and apply them to compare the effectiveness of each in an attempt to identify the ones that perform better in the SMS paradigm, from the experiment Bayesian technique was one of the best approaches towards spam filtration to be considered optimization for better performance in the SMS text context. But this paper depend on the data set and data mining technique but our work using lists to compare terms after stemming process done for detect if it is related or not related to the course .

There are large number of messages posted on the forum of MOODLE [24] , which does not receive adequate feedback from the tutor in sufficient ,this paper proposes an approach for solving this problem using the multi agent system paradigm , the system is composed by intelligent agents. The writers named their model MAS which is composed of MOODLE forum, MySQL Database, AIML interpreter, the AIML knowledge base, and intelligent agents. The MAS is composed of six intelligent agents. Tutor agent is the principal agent of the system; its function is to interact with other agents to acquire information's. AIML interpreter is responsible for reading a question, finding the answer in its knowledge base and returning a response. Agents solve questions about matters discussed at the forum, and they use perception to recommend the implementation of activities that the student has not done. The partial results indicate that the approach of combining AIML and MAS is promising too

Using filtering system to filter and monitor junk text message significantly [25] improves the quality of text-message, and create a healthy environment for legal provider. It also provides mobile carriers a platform to manage text message, and boost the healthy market development of text message.The Filter Monitoring system is composed by four parts: communication sub-system the communication interface of filter monitor system, content-filter sub-system to filter illegal content via filtering algorithm, and save them to database for query, group message monitoring sub-system, it identifies mobile numbers that send group message and then punishes these numbers according to punishment criteria to blacklist, and maintenance management sub-system is a maintenance platform for this system.

3.3 Recommender system

The paper [26] proposed architecture for intelligent E-Learning system that helps learners in classifying the web massive contents through the use of an automated assistant (a recommender system) according to the learner interest in general domain. It is based on a novel feature reduction technique, the architecture of the system was designed in three layers presentation, logic and data layer.

The presentation layer is the first one which is graphical user interface for the user, the second layer is the logic layer contains three modules: tutor module which uses

information from the student model to determine what aspects of the domain knowledge should be presented to learner. Also it contains planner which generates the plans for teaching each item of subject matter, it monitors the student's actions and adapts the system responses to the student. The student (profiler) module which consists of two parts. The first part is responsible for modeling the individual characteristics of the student. The second part called profiler which builds a model of user interest based on the pages he visits, and uses a clustering algorithm to build the profile, and recommender module which analyzes the content of web pages using the user profile to indicate whether it is interesting or not by comparing its content against the information stored in the user profile, then let the GUI annotate if the website interest for the user. The final layer is Data Layer is where information stored in databases, and user (learner) profiles. Three-Tier client server architecture (presentation, logic and data) is represented to handle the implementation in distributed environments.

In [27] the writers here implemented a model of intelligent Question /Answer system depending on previous knowledge using natural language which is converted into data in order to be processed by computer. The paper uses natural language processing technology to segment text and calculate weight, then matches it in database, the system returns the data matched successfully to user. Otherwise the system will submit questions to administrator by email. After administrator solves the problems, it is saved on database.

3.4 Information retrieval

Most IR systems compute a numeric score on how well each object in the database matches the query, and rank the objects according to this value. The models of information retrieval are the Boolean model, the Statistical model, which includes the vector space and the probabilistic retrieval model, and the Linguistic and Knowledge-based models

Here in [28] a method for reformulating the Recommender Systems problem in an Information Retrieval one. They have a dataset of users who give ratings for some movies; they hide some values from the dataset, and try to predict them again using its remaining portion (the so-called "leave-n-out approach"). In order to use an Information

Retrieval algorithm, they reformulate this Recommender Systems problem in this way: a user corresponds to a document, a movie corresponds to a term, the active user (whose rating they want to predict) plays the role of the query, and the ratings are used as weights, in place of the weighting schema of the original IR algorithm. The output is the ranking list of the documents ("users") relevant for the query ("active user"). They use the ratings of these users, weighted according to the rank, to predict the rating of the active user. They carry out the comparison by means of a typical metric, namely the accuracy of the predictions returned by the algorithm, and they compare this to the real ratings from users. In first tests, the writers use two different Information Retrieval algorithms: LSPR, a recently proposed model based on Discrete Fourier Transform, and a simple vector space model.

This paper[29] proposed a recommender system for e-learning that is able to recommend similar materials as presentation slides and word documents based on the similarity of content items (using Vector Space Model) and good learners' average rating strategy, it used to improve the learners learning process.

In [30] paper they integrated the information retrieval method and document clustering. The method is known as Latent Semantic Index (LSI) approach which used Singular Vector Decomposition (SVD) or Principle Component Analysis. The aim of this method is to reduce the matrix dimension by finding the pattern in document collection with refers to concurrent of the terms . Each method is implemented to weight of term-document in vector space model (VSM) for document clustering using fuzzy c-means algorithm. Besides reduction of term-document matrix, this research also uses the cosine similarity measurement as replacement of Euclidean distance to involve in fuzzy c-means . And as a result, the performance of the proposed method is better than the existing method with f-measure around 0.91 and entropy around 0.51.

In [31] paper, aim to compute on-line automatic recommendations to an active learner based on recent navigation history, as well as exploiting similarities and dissimilarities among user preferences and among the contents of the learning resources. First, they start by mining learner profiles using Web usage mining techniques and content-based profiles using information retrieval techniques. Then, they use these profiles to compute relevant

links to recommend for an active learner by applying a number of different recommendation strategies.

The Vector Space Model is one of the most common information retrieval, the cosine of the angle or the Euclidean distance between the query vector and each document vector is commonly used to measure similarity for query matching. In [32] paper presents a modified vector space model for measuring similarity between the query and the document. This paper presents a modified vector space model for measuring similarity between the query and the document when responding to a multi-term query . More weight is assigned to the keywords based on the adjacency between the terms in the documents. When a document contains the adjacency terms, its vector will typically move closer to the query vector to show stronger relevancy between query and the document.

In [33] paper discusses a new similarity measure between documents on a vector space model from the view point of distance metric learning, The documents are represented by points in the vector space by using the information of frequencies of words appearing in each document . The similarity measure between two different documents is useful to recognize the relationship and can be applied to classification or clustering of the data. Many words, which appear in documents, have correlation to one another depending on the sentence structures, topics and subjects. This paper presents a new effective method to acquire a distance measure on the document vector space based on an extended cosine measure. In this paper, the vector space model is adopted to represent the document data. The similarity measure is naturally defined in the vector space model and each document is expressed as a vector whose elements are basically meaning the information of word frequencies

In [34] paper, they build a meta-searches Soft Bot Search, that retrieves the web pages from the internet and groups them according to their theme , then coming the next step which is Pre-processing involves removing stop words and then stemming the remaining words . After preprocessing, a document matrix is created, and words appearing too rarely or too frequently using vector space model. Then the word frequency scores are converted to TFIDF scores, they use threshold to reject some terms which occurring in less than 2% and more than 30%, Instead of using binary values representing presence

or absence of words, we use the Term Frequency – Inverse Document Frequency or TFIDF, then used PCA to reduce the feature vector dimension while retaining most of the information by constructing a linear transformation matrix. Then using clustering method to discover groups such that documents in one group are similar to each other using K-Means Clustering and Suffix Tree Clustering.

3.5 Conclusion

According to what has been reviewed from papers in this chapter, there are many studies have done a comparison between the learning management systems, as shown [21] that MOODLE have 30 features available from 32 features for Learning Management System, so the MOODLE is more suitable as Learning Management System, also the chapter presents many types of used techniques on filtering messages as spam [22], viruses and unwanted message [25]. The filtering process is based on black and white lists, or depends on the data set and data mining technique for spam message in e-mails [23]. Also there are some techniques for filtering messages which are posted in the MOODLE forum [24], which does not receive adequate feedback from the tutor in sufficient way. In this research we will use the technique list to filter the message, whether if it contains taboo words. There are many usages for Recommender system as recommend similar materials, presentation slides and word documents based on information retrieval algorithms [26], and question /answer system depending on previous knowledge using mother language [27].

Many papers talk about the importance of information retrieval algorithm in retrieving the material and ranking them according to the most related fields [28]. One of the algorithm is vector space model [32] which computes a continuous degree of similarity between queries and documents, and also can make partial comparing.

CHAPTER 4

FILTERING FEATURES AND RECOMMENDER MODEL FOR MOODLE MESSAGING SYSTEM (FRMM)

This chapter presents FRMM model "Filtering Features and Recommender model For Messaging System of MOODLE".

FRMM model was described using flowcharts and figures. Various stages have been performed to achieve the objective of the model, starting with inserting the message by students ,then FRMM model makes preprocessing for the message including parsing , stop words and stemming process , after that FRMM searches for taboo words in the message , if any of the taboo words were founded in the message, the FRMM directly forward the message for the Academic Affairs secretary and will be deleted from the instructor's account. If FRMM find a question mark "?" that means this message contains a question, and needs an answer , if the question mark is not found, that means there is no question in it, and the instructor received it .

4.1 The Proposed Model FRMM

As known the MOODLE consist of many modules as user interface, course interface, grouping interface as shown in figure 4.1

Through user interface the users can deal with all functions in MOODLE, the interface is what user can see and click on in the web browser, and it provides ways to access, understand and change the database, one of the user interface is course Interface module uses to manage all the courses in MOODLE.

Every course has its group modules as communication module for emails, chats and messages where the work of FRMM, courses module (Recourse) for upload files, put a link in the course page and activity module to build forum.

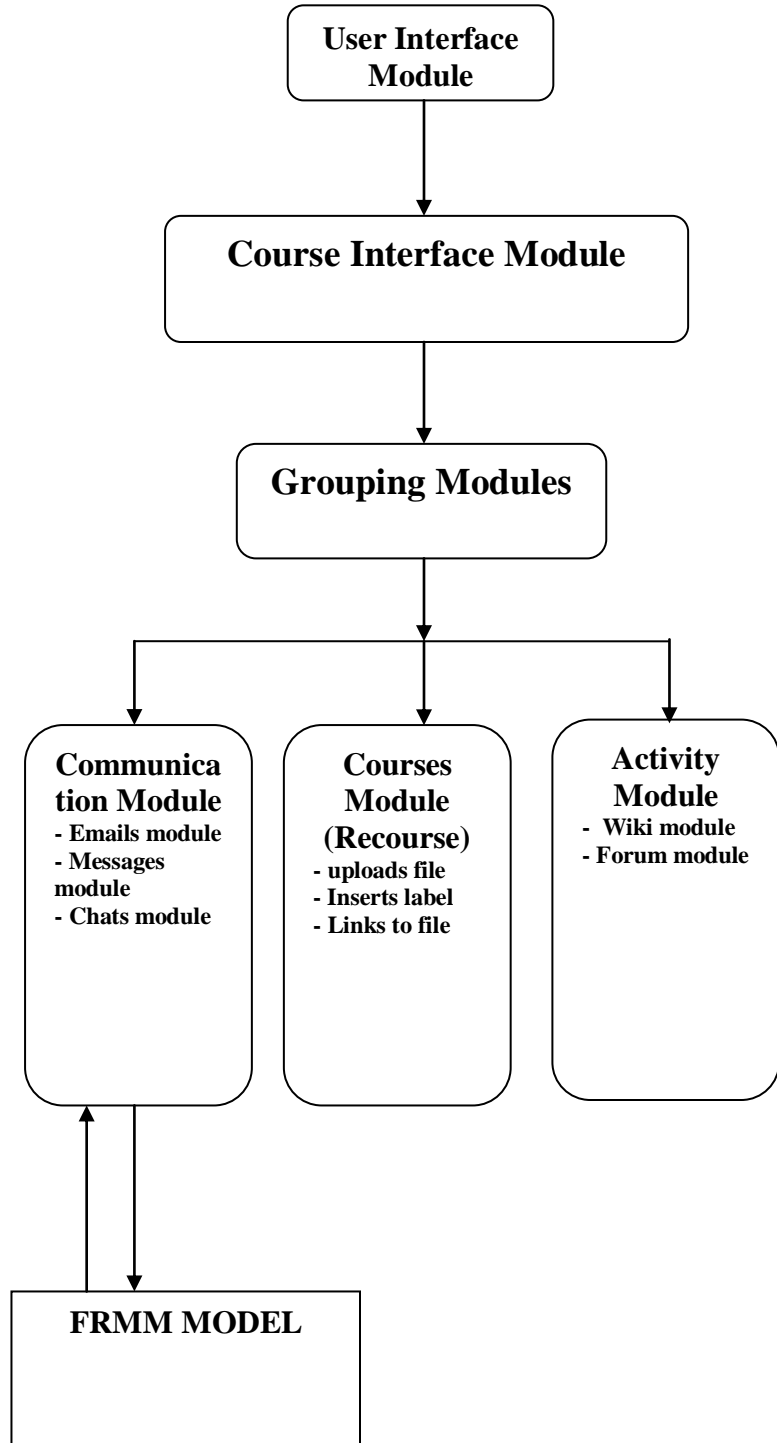


Figure 4. 1 : Explains the structure of the MOODLE and where the FRMM MODEL integrated Model

- **Model Design:** The FFRM system structure is shown in Figures 4.2 which explained in the following steps :
 1. The Model deals with message system MOODLE system: it uses only messaging system in which student sends message to instructor that contains a question mark "?" .
 2. The agent programmed using Java language, it has retrieve the message from the MOODLE Database, then processing it, after that it restore it on the MOODLE database.
 3. Pre-process the message: Stop words, Stemming, token.
 4. After filtering messages, if they contain taboo words or not, then the system check if there is a question mark "?", if the system answers the student question by resending the related material about the student question as an example.
 5. If the student sends a question about "what is the software process?" The system send the chapter about the question requested information according to a special rank using cosine model in vector space.
 6. The result from the model have been arranged, using vector space model, which sends the highest arrange of the material to the student.

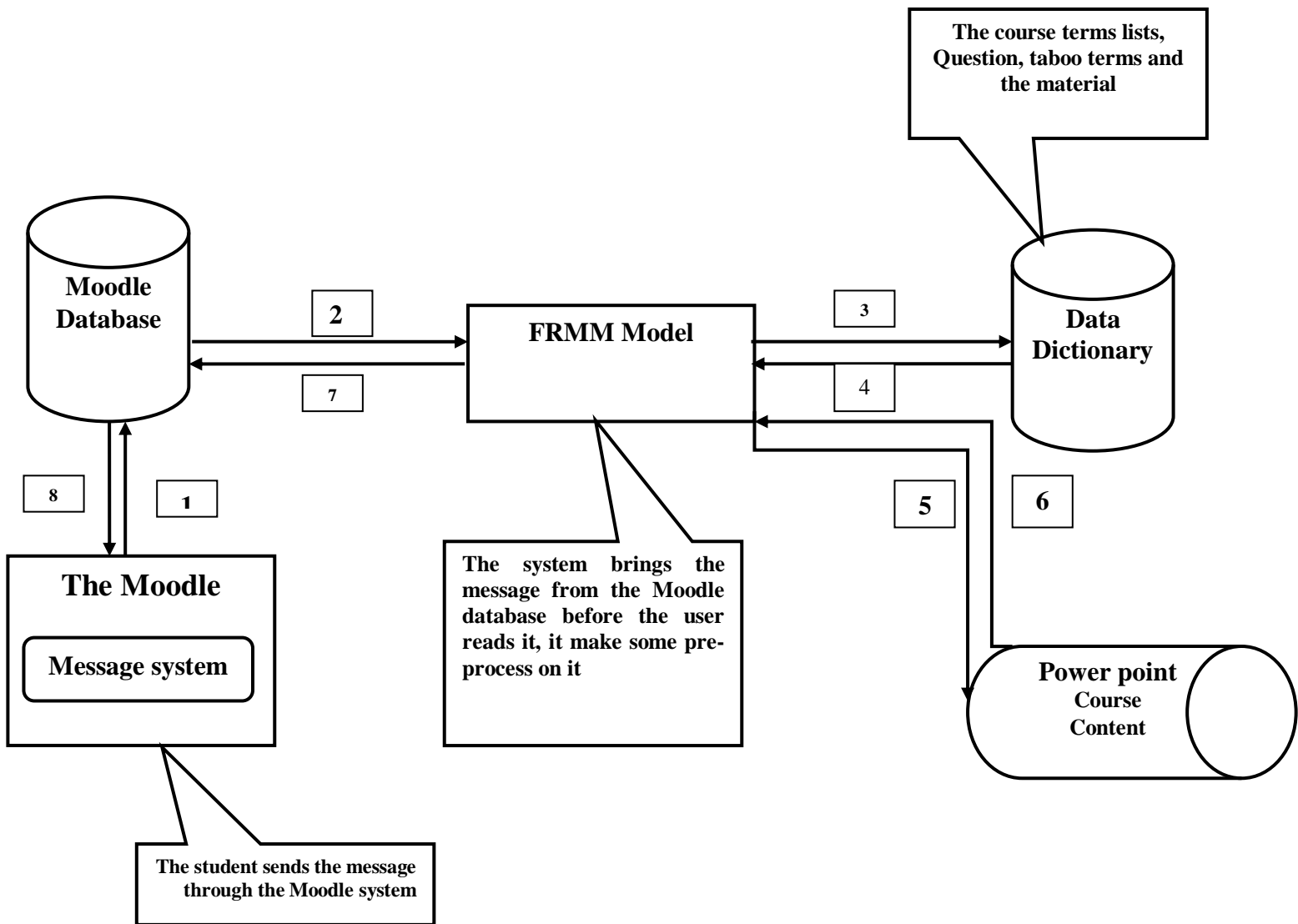


Figure 4. 2 : Explains the model of the System

Table 1 : Summaries the process of the model

No. Of Operation	Description
1	The student sent a message using MOODLE then it will be stored in the database.
2	FRMM Model retrieves the message from the database then pre-process the message
3	The main terms will be send to the lists for searching, where this term will find " Taboo words" ,then search for question mark "?",after that it will compare the materials in MOODLE to its suitable material
4	Engine sends the chapter's name that the terms founded ,store in the Database
5	The MOODLE will retrieve the request according to a special rank and sends it to the student.

- **Model Implementation:** Using Java programming language, more detail see Appendix C.
- **The method includes the following general steps as shown in figure4.2:**
 1. Students write their question about any course using messaging system in the MOODLE.
 2. The system pre-process the question using tokenize, stemming and matching string.
 3. Check if there are taboo words in the message.
 4. If there is a question on the message about the course, the system sends related material about the question after ranking them using vector space method.

As shown in model table1, it has 5 processes to achieve its work , which starts with sending the message by student , process number 2 the model makes some process on the message which we can name pre-process to list the important words in the message as stemming to stem the words to their root , parsing process and stopwords, process number 3 which is the main process in the model , it searches for taboo words if there is, it resend the message to the Academic affair secretary , if not the model starts searching for question mark "?" if the question mark is not founded that means ,there is no question and it will be sent to the instructor , process number 4 the model searches for the highest ranking in PowerPoint files related to the student question and send it on the database , the last process is retrieved by the chapter's name and is sent it to the student.

4.2 The flowchart for FRMM

Figure 4.3 explain the flowchart of FRMM model, starting with inserting the message by students using interface of the messaging system, then FRMM model makes preprocessing for the message after retrieve it from the database including parsing, stop words and stemming process. After that FRMM searches for taboo words in the message, if any of the taboo words were founded in the message, the FRMM directly forward the message for the Academic Affairs secretary and will be deleted from the instructor's account.

FRMM searches for question mark "?", If it is founded that means this message contains a question, and needs an answer, if the question mark is not founded, that means there is no question in it, and the instructor received the message.

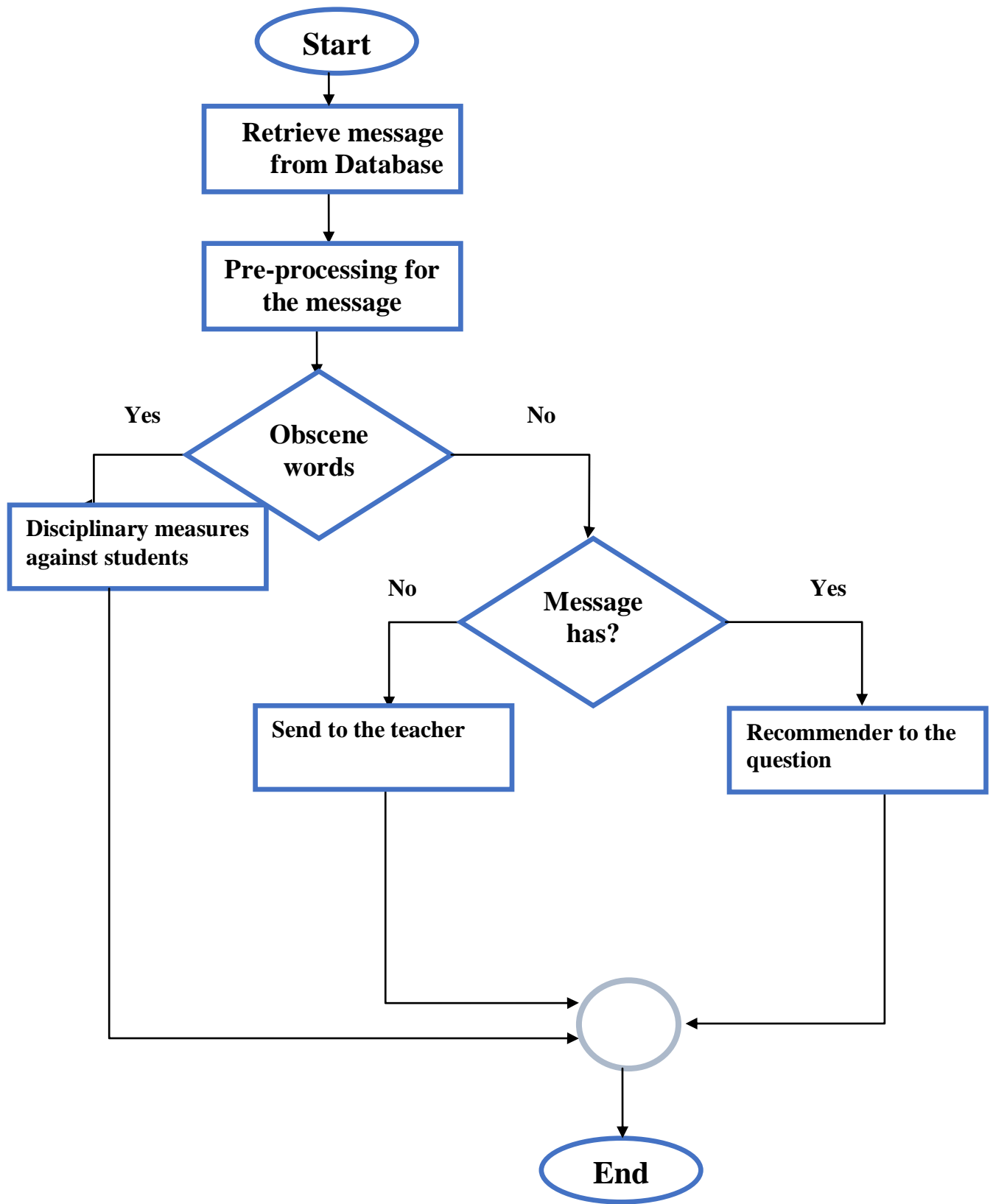


Figure 4. 3 : The proposed FRMM flow chart

4.3 Input the data to the system

This step is considered to be the system inputs, that the system retrieves the message from the database, where the student sends it through the MOODLE, there is no limitation on the size of the included text or number of sentences in the message.

4.4 Message Pre-processing

- 1.Token the purpose of tokenization is therefore to break down the text into tokens (or terms), which are small units of meaningful text on the message.
- 2.Stop words are words, in which are filtered out before starting processing the message.
- 3.Stemming which is returning the words to their stem.

The main words only will kept in the message, this will increase the engine speed.

4.5 Taboo words

Each word in the message will be compared with the pattern of Taboo word in the engine, as shown in **Figure 4.4**.

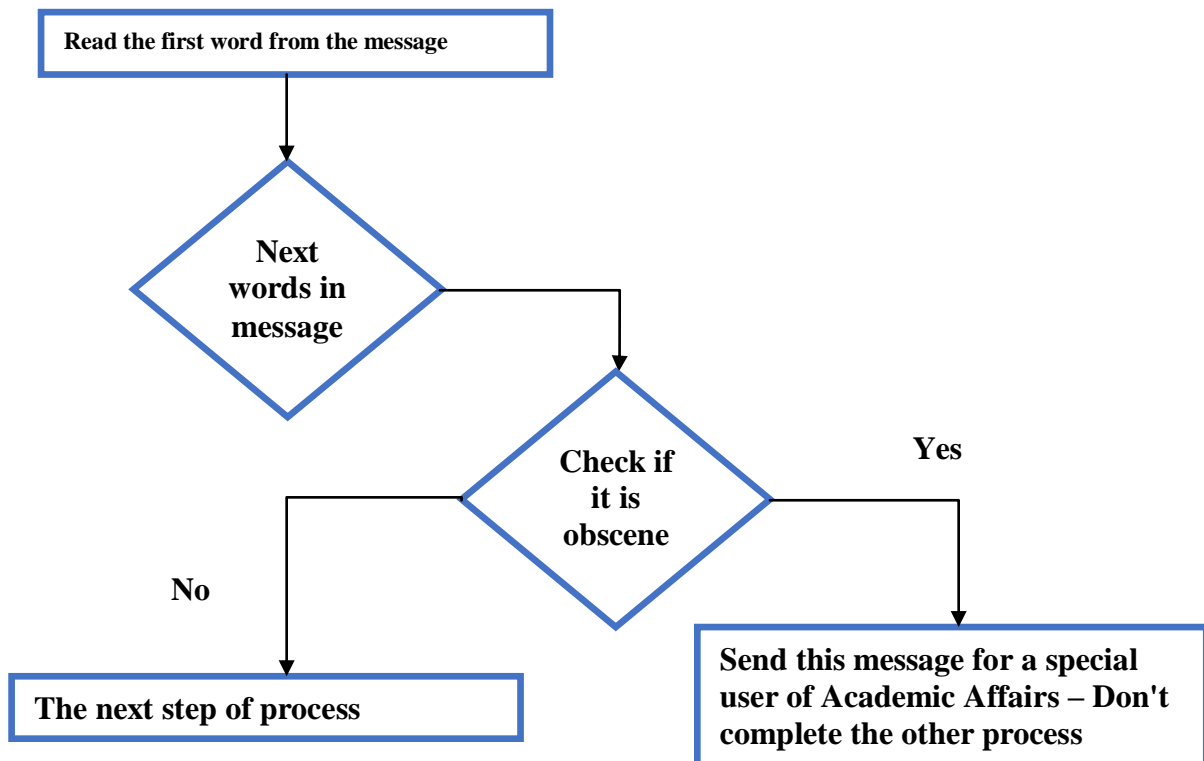


Figure 4. 4 : Explain the process of searching taboo word on the message Search for Question Mark"?"

This part was to search for the Question mark "?" in the message, if the question finds the message is sent to the agent as shown in **Figure 4.5** it recommends a chapter to read , if it doesn't find question mark "?" that means the message is not a question, so the system will send it to the teacher to read it .

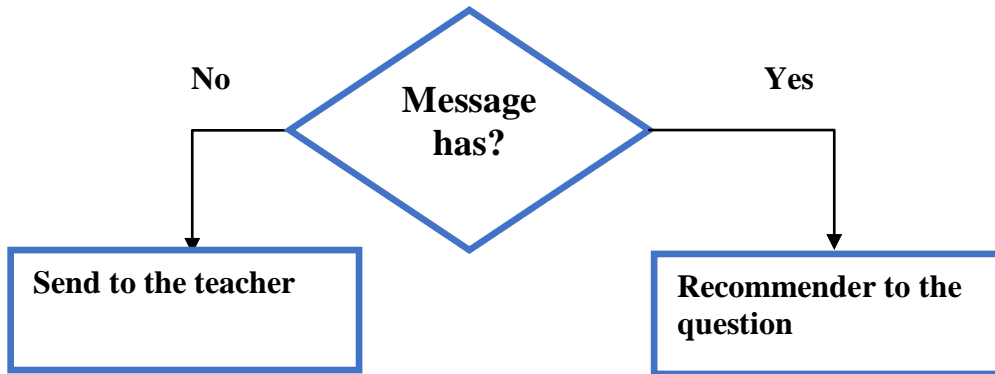


Figure 4. 5 : Explain the process of searching about the Question mark "?"

4.5 The Agent for recommender

The main keywords in the message will be searched for in the main keywords list, which creates an automatic results for each chapter of the course material, after using tokenize and stop words to delete all common words as the , they , a , as, is , so ,and.

4.6 Examples on how FFRM works:

1. Starting from the student who sends a message to the instructor, there are 3 types of messages which can be sent by the student:
 - a. The message which contains taboo words as shown in figure 4.6, the student sends to the instructor "you are donkey".



Figure 4. 6 : An example shows a message which contains taboo word

- b. The message which contains question mark "?" and doesn't contain taboo words as shown in figure 4.7, the message was "what we mean by reusing array variable?"

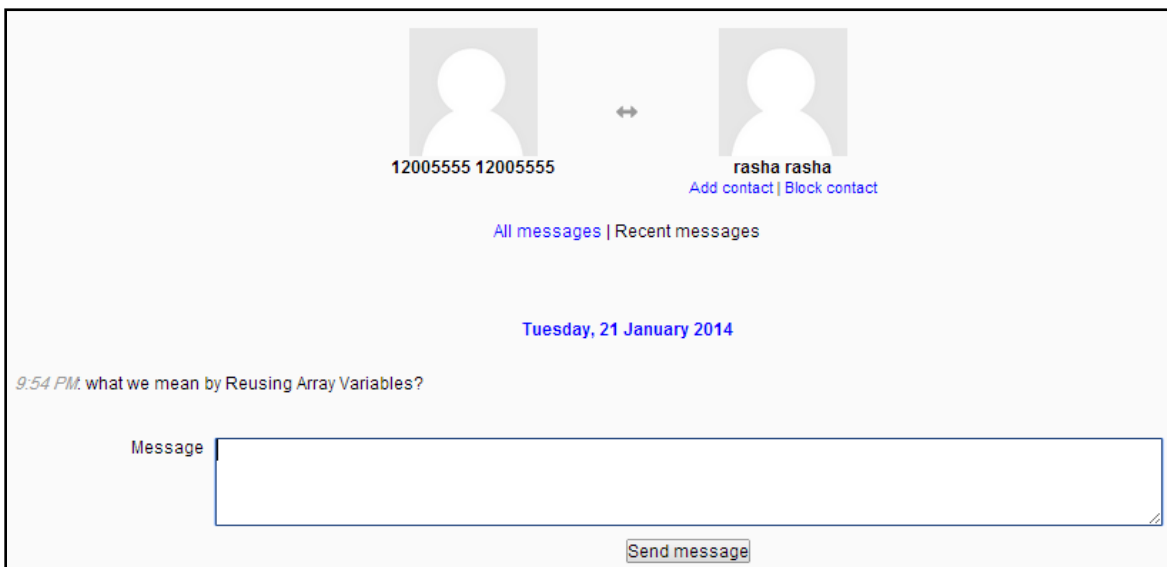


Figure 4. 7 : an example shows a message which contains question mark "?"

- c. The message which hasn't question mark and hasn't taboo words as shown in figure4.8 the message was " Dear Teacher I would like to say I will be absent next lecturer "

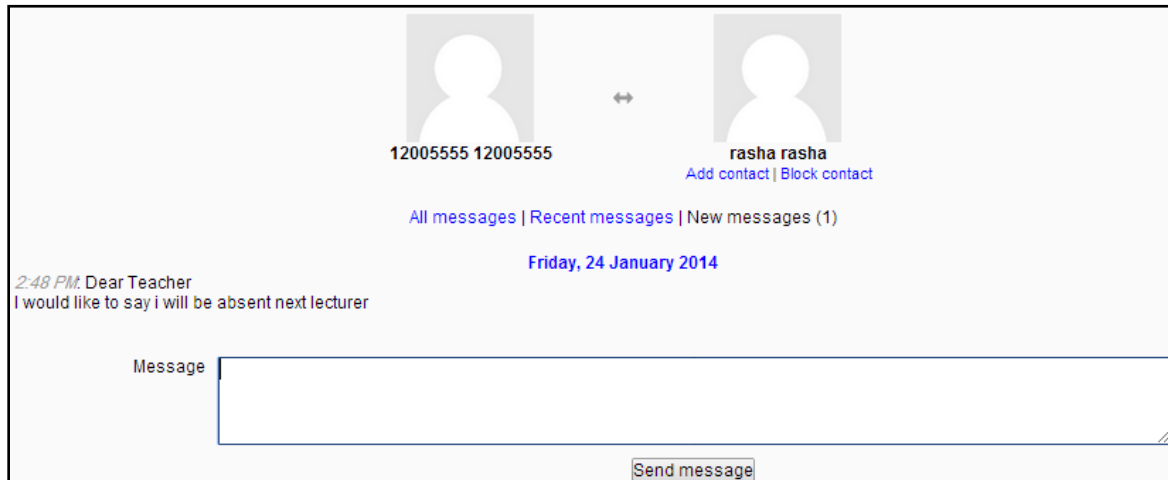


Figure 4. 8 : An example explains a message that doesn't contain taboo word or question mark?"

2. If the message contains taboo words, FRMM will forward this message for the Academic affair secretary, and disciplinary measures will be taken against students such as verbal or written warnings. Also the message will not be received by the instructor .Figure 4.9 shows a new message received by the academic affair secretary, and the message format, as shown that the student which have ID 120111006 have a problem with the instructor.

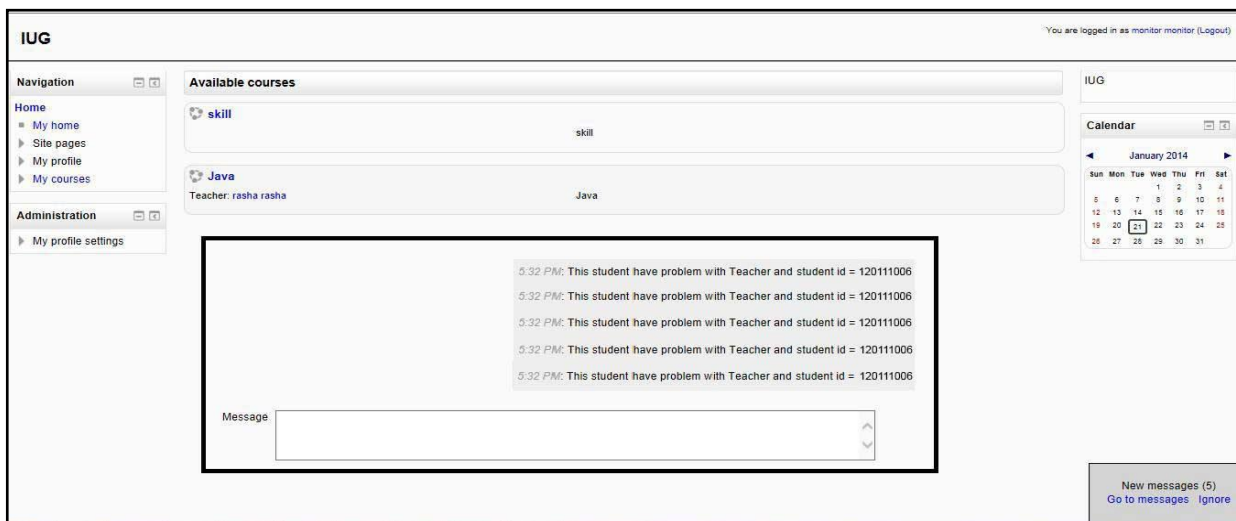


Figure 4. 9 : shows a new message which is forward to the academic affair secretary and it is formatted

3. If the message do not contain taboo words and contains question mark "?", that means the student asks about something he is not understanding. So FRMM makes his process to reach to the suitable answer. As noted in figure 4.10 according to the question in figure 4.7 the system makes its process and the rank of all PowerPoint files and determines the highest rank.

```
Doc2 ->file name :array.ppt
Doc3 ->file name :class.ppt
Doc4 ->file name :polymorphism.ppt
Doc5 ->file name :recursion.ppt
Doc2 ->file name array.ppt ___ 0.21368693698332852
Doc3 ->file name class.ppt ___ 0.17560213071331215
Doc4 ->file name polymorphism.ppt ___ 0.1083875890069111
Doc5 ->file name recursion.ppt ___ 0.08780866895102216
check ? -->
? is found
10
BUILD SUCCESSFUL (total time: 5 seconds)
```

Figure 4. 10 : show the rank for all PowerPoint files in the system

Figure4.11 show the formatted of the message that will send for the student from FRMM , Here the system send the Array file to the student to read it according to his question shown in figure 4.7 and figure 4.10 which shown that the array file is the highest rank.

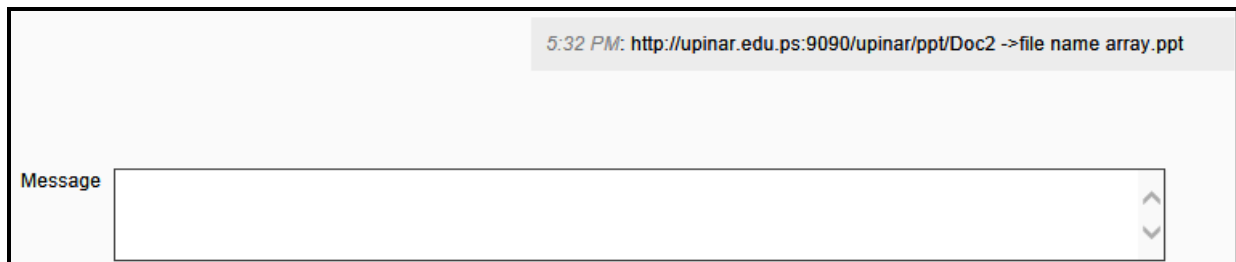


Figure 4. 11 : show the chapter's name which send to the student

4. If the message do not contain taboo word or question mark that means this message doesn't need any action, and it sent to the instructor.

4.7 Conclusion

To achieve the objectives of FRMM there are many stages must be performed on the message. Firstly preprocess operation done on the message. Secondly then search for any taboo words and search for question mark. Thirdly search for the highest rank of PowerPoint file based on the student answer. In this chapter flowchart, figures have been used to explains the structure of FRMM and explain an examples for how the system work

CHAPTER 5

EXPERIMENTAL RESULTS AND EVALUATION

This chapter discusses the experiments which were carried out to evaluate FRMM model. The chapter includes three sections: Section 5.1 presents the implementation, Section 5.2 present the system's user, section 5.3 briefly describes the experimental environment and states the programming language and tools used to develop the proposed system. Finally, Sections 5.4 talks about the material nature, section 5.4 gives the evaluation, finally section5.5 it present and discuss experimental results.

5.1 Implementation

To implement the FRMM, we used java programming language because it's easy to deploy our work with the MOODLE , for more detail see Appendix C.

5.2 The users of the system

For this research, the domain of experiment is for the students who registered programming 1, Computer skills, Software engineers, Management information system, System analysis and Algebra courses at the first semester 2013-2014 in Information Technology College at University of Palestine. All students who registered those courses whether they are males and females, they use the system to send message to their teachers.

5.3 Experimenting Setup

This section describes the setting of experimenting and evaluating our proposed approach. It is a primary purpose, to empirically evaluate our proposed model. Our approach is expected to recommend a power point presentation for every question from students and check the message which has taboo words, and sends it to the system monitor. The following describes experimental and environmental tools.

5.3.1 Experimental Environment and Tools

The experiments took place in a server with the following specification Intel® Xeon® E5645 , 6 core, 2.40 GHz processor , 6GB of RAM , we need some specific software to implement our model :

- Net Beans 6.9 with integrated JDK (Java Development Kit) 1.7 java environment: Net Beans program helps us to develop, build, compile, validate and execute our algorithm and to build jar file, which will be included in the MOODLE.
- MySQL: to build the MOODLE Database.
- Dreamweaver: to update the MOODLE code pages which are written in php.

5.3.2 Experiment al Procedures

The objective of experiments to measure the system's performance, by comparing the system's performance, and its response to students' questions with the teacher recommendations regarding to the same questions.

1. First the student sends a message to his teacher from the MOODLE system using the course page.
2. The system recommends the student with the needed chapter to read.
3. We send the 50 message, which are sent through the system for a teacher who are teaching the same courses to recommender which chapter for every question.
4. Manually we compare the differences of the answers from the system and the teacher.

5.4 The Material's Nature

The experiment applicant can obtain the material of PowerPoint type only and also these files must not contain any picture because the system doesn't support retrieving image, which clarifies the difference between the material in FRMM system and the teacher.

The teacher files are with images, but the FRMM doesn't include images, but most of these images are founded as supporter for science files.

5.5 System Evaluation:

This section shows the system's accuracy and the reliability evaluation, the accuracy is the degree of quantity measurements closeness that quantity's actual (true) value, and the system's reliability is the probability, over a given period of time that the system will correctly deliver services as expected by the user.

5.5.1 Accuracy Calculation:

As shown in table 2 the accuracy is various from course to another, also from taboo words experiment, all messages which contains taboo words is not sent to the instructor but they resend it to the academic affairs secretary account (for more details about the message see appendix A).

1. Accuracy of FRMM on computer skill course = $(44 * 100) / 45 = 97.7$
 2. Accuracy of FRMM on Management Information system course = $(44 * 100) / 45 = 97.7$
 3. Accuracy of FRMM on Software Engineer course = $(44 * 100) / 45 = 97.7$
 4. Accuracy of FRMM on System analysis and Design course = $(42 * 100) / 45 = 93$
 5. Accuracy of FRMM on Algebra course = $(45 * 100) / 45 = 100$
 6. Accuracy of FRMM on Introduction to Programming course = $(42 * 100) / 45 = 93$
- **The Accuracy for the whole system is = $(44 * 3 + 42 * 2 + 45) * 100 / 270 = 96.666$**

Table 2 : the figure shows the system's accuracy

No.	Course Name	Correct Recommender 45 message	Accuracy %
1.	Computer skill	44	97.7
2.	Management Information System	44	97.7
3.	Software Engineer	44	97.7
4.	System analysis and Design	42	93
5.	Algebra	45	100
6.	Introduction to Programming	42	93

Also the system's accuracy which filters taboo word message from by 100%, according to the experiment that was done in the system, all messages that contain taboo word were sent to the academic affair secretary, see appendix A

- **The Accuracy for the whole system in taboo words experiment = $(45/45) * 100 = 100\%$
If we take it into consideration to list taboo words.**

5.5.2 Using SPSS to calculate the reliability:

The system is evaluated by using SPSS T-independent test and Cronbach's Alpha method, the system's reliability is 92.55%, and the error in the system is 7.45%.

- T-independent test:

The method inputs were the answers from the system and the teacher's answers.

Figure 5. 1 : It shows the T-independent test method

- Cronbach's Alpha method :

Descriptive Statistics									
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Correct	6	42	45	43.50	1.225	-.490	.845	-1.467	1.741
Valid N (listwise)	6								

Valid = 92.55 %
Missing = 7.45 %

Cronbach's alpha is the most common measure of the internal consistency ("reliability").

Figure 5. 2 : It shows the result

- The system's reliability was 92.55

5.6 The Result Analysis

The experimental for FRMM done on six courses , Computer skill , Management information system , System analyses and design , Introduction to programming , Algebra , Software engineer . Every course has it is result according to it is natural material.

5.6.1 Computer Skill:

The computer skill material consists of 8 chapters , every chapter is presented by a Power Point file , as we know the computer skill topics are separated , the first chapter is about Computers and Digital Basics , the second chapter talks about computer hardware , the third chapter talks about Software , the fourth chapter about Operating Systems and File Management , the fifth chapter is about LANs and WLANs , the sixth chapter is about Internet ,the seventh chapter is about Web and E-mail , the last chapter talks about Digital Media.

As noted from the analysis of computer skill material nature , there are no interference between the topics , so the terms which send with query , have no probability ,that means the terms will not appear in more than one chapter , as an example Database only is existed in chapter 5.

5.6.2 Management Information System

The material in Management Information system consists of 15 chapters, every chapter talks about topics, the first chapter talks about managing the digital. the second chapter is about information system in the enterprise , the third chapter is about information systems , organizations management and strategy ,the fourth chapter talks about digital firm , electronic business and electronic commerce , the fifth chapter is about ethical and social issues , the sixth chapter talk about how managing the hardware and software , the seventh chapter is about database , the eight chapter is about communication and network , the ninth chapter is about new information technology infrastructure , the tenth chapter is about how managing digital firm knowledge , the eleventh chapter is about decision making in digital firm , the twelfth chapter is about redesigning the organization to use

information system , the thirteenth chapter is about the system's business value, the fourteenth chapter talks about security and control , the fifteenth chapter is about managing international information system.

As noted from the analysis of the natural of management information system every chapter is separated topic, but the only unmatched query is one answer, it was question about normalization, the stem of normalization is normal, term normal founded in two places chapter 7 that is the correct answer and the another place on chapter 13, Therefore, the result appeared like this.

5.6.3 System Analysis and Design Course

The material in system analysis and design consists of 16 chapters ,the first chapter is about the systems, roles, and development methodologies , the second chapter is about understanding and modeling organizational Systems , the third chapter explains the project management , the fourth chapter is about how gathering information by using interactive methods , the fifth chapter is about the information gathering using unobtrusive methods ,the sixth chapter talks about Agile Modeling and Prototyping , the seventh chapter is Using Dataflow Diagrams , the eight chapter is Analyzing Systems Using Data Dictionaries ,the ninth chapter is Process Specifications and Structured Decisions ,the tenth chapter is Object-Oriented Systems Analysis and Design Using UML , the eleventh chapter is about Designing Effective Output , the twelfth chapter is about Designing Effective Input , the thirteenth chapter is about Designing Databases , the fourteenth chapter is about The Human-Computer Interaction , the fifteenth chapter is about Designing Accurate Data Entry Procedures , the sixteenth chapter is about Quality Assurance and Implementation.

Some terms are repeated in more than one chapter because some topics appear as points in chapters and they are detailed in other's chapters.

5.6.4 Introduction to Programming Course

The material of programming course consists of 9 chapters , through the experiments the agent failed to give the perfect recommender chapter for 3 questions , because of the

overlapping in topics in programming , the first chapter is about the function , the second chapter is about array, the third chapter about classes and objects, the fourth chapter is about inheritance , the fifth chapter is about the list, the sixth chapter is about polymorphism , the seventh chapter is about queues , the eight chapter is about recursion , the ninth chapter is about stacks .

As known the curriculum of programming course, the topics of the course are tied together, so the terms can be founded on more than one chapter , for example method is founded in more than one chapter .

5.6.5 Algebra Course

The course consists of 20 chapters, every chapter has separated topic in algebra, so there is no overlap between topics .As noted all results which were done by the agent was matching with results which were done by the teacher.

5.6.6 Software Engineer Course

The material consists of 9 chapters , the first chapter is about introduction for all the course so many terms will be repeated through this chapter , the second chapter is about Socio-technical Systems , the third chapter is about the critical systems, the fourth chapter is about software processes which was also explained in the first chapter , the fifth chapter is about project management , the sixth chapter is about software requirements which also was explained in the first and the second chapter as an introduction , the seventh chapter is about the requirements engineering processes , the eight chapter is about system models which also was mentioned in chapter one , that made mismatching between the agent and the teacher result about the query "I don't not understand the software process model " as noted the process model was found in chapter 1 and 8 and the 2 correct but the agent chooses chapter 8 and the teacher chooses chapter 1 because it is the base , the ninth chapter is about interface design.

5.7 Conclusion

The accuracy is various from course to another according to the courses material nature, if the topics is related to each other and the topic may be founded in more than one chapter , the accuracy of the experiment is low, if the topics are separated, the accuracy as noted will be high . Also from the experiment about taboo words, all messages which contain taboo words are not sent to the instructor, but they are forward to the academic affairs secretary account. The whole system accuracy is 96.66 % , that means only 3.34 % was the error in the system.

The system is evaluated by using SPSS T-independent test and Cronbach's Alpha method, the system's reliability is 92.55%, and the error in the system is 7.45% .

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

This thesis solves a problem faced messaging system users in the MOODLE in an inefficient way from the student which may contain taboo words, so the system filters this message and prevents the instructor from reading it but on the other hand the system sends it to the monitor . Also if a message has a question about the course the system can recommend student which material he/she can read, There are two advantages for the system : speed to reply and help the instructor.

The model is applied in the cosine model in information retrieving to improve the system's performance.

The system is tested by comparing it in six courses, for each course there were 45 messages that compare the result from the system and the instructor.

The accuracy varies from course to another according to the course material nature, if the topics are related to each other and the topic may be founded in more than one chapter, the accuracy of the experiment is low, and if the topics are separated the accuracy as noted will be high . Also from the experiment about taboo words all messages which contain taboo words are not sent to the instructor but they are resent to the academic affairs secretary account.

As a result the courses can be divided for two types either related topics for each other or separated topics , the evaluation of the system increases when the topics are separated and unrelated .Also the system is evaluated by using SPSS T-independent test and Cronbach's Alpha method , the system's reliability is 92.55% , and the error in the system is 7.45%.

Finally the accuracy of the system is 96.66% for the whole system, and 3.34% is the error in the system and the reliability of the system was 92.55% using Cronbach's Alpha method that means the model success is to improve messaging system in MOODLE as a first recommender system in the messaging system.

6.2 Future Work

For the future work ,the model can support not only text material , also it can support another type of material as image and video , using multimedia retrieval algorithm , also it can link between students levels education according to their questions , for more performance and accuracy , using ontology instead of database , and using semantic Characteristics.

Future research will be directed in the following direction:

- Use various type of material as video , web pages
- Link between the behavior of the students from the questions they ask
- Use the ontology as an alternative for data structure.

REFERENCES

- [1] M. J. Rosenberg, *E-Learning: Strategies for Delivering Knowledge in the Digital Age*, New York: NY: McGraw-Hill Companies, 2001.
- [2] Tor Hjeltmes. *QUIS: Quality, Cost Effectiveness and Cost Efficiency in E-Learning*, 2005.
- [3] Rosalina Babo, Ana Azevedo, "A Comparison of Two Learning management Systems: Moodle vs Blackboard," in *18th Annual Conference of the National Advisory Committee on Computing Qualifications*, 2005.
- [4] Babacar Ngom, Hervé Guillermet and Ibrahima Niang, "Enhancing Moodle for offline Learning in a degraded connectivity environment," in *Multimedia Computing and Systems (ICMCS)*, 2012.
- [5] Cansu Aydin, Guzin Tirkes, "Open source learning management systems in e-learning and Moodle," in *IEEE EDUCON*, 2010.
- [6] James Gosling ,Bill Joy ,Guy Steele, Gilad Bracha ,Alex Buckley, *The Java™ Language Specification* , Java SE 7 edition, 2011.
- [7] Richard E. Pattis, Carnegie Mellon University (2005), "Extended Backus-Naur Form or BackusNormal Form (EBNF)
- [8] W. Francis, *Frequency Analysis of English Usage: Lexicon and Grammar*, Houghton Miin, 1982.
- [9] Van Rijsbergen, C. J, *Information Retrieval*, 2nd edition, Butterworth-Heinemann, 1979
- [10] J. B. Lovins, *Development of a Stemming Algorithm". Mechanical Translation and Computational Linguistics*, 1968.
- [11] P. Willett, *The Porter stemming algorithm: then and now. Program: electronic library and information systems*, Emerald Group Publishing Limited, 2006.
- [12] Berthier Ribeiro-Neto, Ricardo Baeza-Yates, *Modern Information Retrieval*, ACM, 1999.

- [13] T. A. R. Bradman, *The Process of Knowledge Discovery in Databases: A Human-Centered Approach,* " *Advances in Knowledge Discovery and Data Mining, Menlo Park, The AAAI Press/the MIT Press, 1996.*
- [14] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, *From Data Mining to Knowledge Discovery: An Overview,* " *Advances in Knowledge Discovery and Data Mining, 1996.*
- [15] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze,, *Introduction to Information Retrieval, 2008: Cambridge University.*
- [16] C. S. Yang , A. Wong ,G. Salton, *A vector space model for automatic indexing, ACM, 1975.*
- [17] P. Peter, D. Turney, *Vector Space Models of Semantics, Journal of Artificial Intelligence Research 37 , 2010.*
- [18] M.L .Mursidi, N.A.A.Husin, Raihana Reh, "Reliability Analysis for Pilot Survey in Integrated Survey Management System," in *5th Malaysian Conference in Software Engineering (MySEC) , 2011.*
- [19] M. George D., *A simple guide and reference. 11.0 update (4th ed.), 2013.*
- [20] LEE CRONBACH, "COEFFICIENT ALPHA AND THE INTERNAL STRUCTURE OF TESTS," 1951 .
- [21] Babacar Ngom, Guillemet H. , Niang I. *Enhancing Moodle for offline Learning in a degraded connectivity environment, IEEE, 2012.*
- [22] Zhijun Liu, Weili Lin , Na Li , Lee D. *Detecting and Filtering Instant Messaging Spam – A Global and Personalized Approach, Secure Network Protocols, 1st IEEE ICNP Workshop, 2005.*
- [23] B. I. Kuruvilla Mathew, "Intelligent Spam Classification for Mobile Text Message," in *International Conference on Computer Science and Network Technology, 2011.*
- [24] Márcio Alencar, Netto, J.M. "Improving Cooperation in Virtual Learning Environments Using Multi-Agent Systems and AIML," in *41st ASEE/IEEE Frontiers in Education Conference 2011, 2011.*

- [25] G. Z. Haifeng Ke, "A Text-Message Filtering System Design based on ACE Platform," in *2nd International Conference on Computer Engineering and Technology*, 2010.
- [26] Samir Abdel Rahman, Abdallah Gomah, Amr Badr, Ibrahim Farag "An Auto- Recommender Based Intelligent E-Learning System," in *IJCSNS International Journal of Computer Science and Network Security*, 2011.
- [27] Sujuan Wang , Yan Zou , Qin Wang , Shouning Qu, "Research and Design of Intelligent Question Answering System," in *International Conference on Computer and Electrical Engineering*, 2008.
- [28] F. R. Alberto Costa, "Recommender Systems by means of Information Retrieval," in *WIMS '11 Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, 2011.
- [29] Nor Aniza Abdullah, Khairil Imran Bin Ghauth, "Building an E-Learning Recommender System using Vector Space Model and Good Learners Average Rating," in *Ninth IEEE International Conference* , 2009.
- [30] Baharum Baharudin ,Lailil Muflikhah, "Document Clustering using Concept Space and Cosine Similarity Measurement," in *Computer Technology and Development*, 2009. *ICCTD '09. International Conference*, 2009.
- [31] M. Jemni and O. M. Nasraoui, "Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval," in *Advanced Learning Technologies*, 2008. *ICALT '08. Eighth IEEE International Conference on*, 2008.
- [32] L. Wang, "Relevance Weighting of Multi-Term Queries for Vector Space Model," in *Computational Intelligence and Data Mining*, 2009. *CIDM '09. IEEE Symposium on*, 2009.
- [33] K. Mikawa, T. Ishida , M. Goto, "A Proposal of Extended Cosine Measure for Distance Metric Learning in Text Classification," in *Systems, Man, and Cybernetics (SMC)*, 2011 *IEEE International Conference* , 2011 .
- [34] A. S. Agarwal, "Soft Bot Search : Soft Bot Approach to information retrieval from Web," in *12th Annual IEEE International Conference*, 2008.

Appendix A: Students' Questions

In this section we introduce students' questions which the system recommends for them, the first column shows the questions for every course, the second column shows recommender the system form, and the third column shows the recommender from the teacher.

Table 3 : The table shows 45 questions for programming 1 course sent by students

Example From programming 1 course			
NO.	Questions	FRMM Material	Teacher Material
1.	Array Definition	array	array
2.	Declaring and Allocating Arrays	array	array
3.	Reference Parameters	array	array
4.	Linear Search	array	array
5.	•Class constructor	class	class
6.	Static Class Members	class	class
7.	what do we mean by Reusing Array Variables?	array	array
8.	What are the Multidimensional Arrays?	array	array
9.	Why overload a method?	Polymorphism	Polymorphism
10.	Shadowing	Polymorphism	Polymorphism
11.	How to override a method?	Polymorphism	Polymorphism
12.	Calling an overridden method	Polymorphism	Polymorphism
13.	The equals method	Polymorphism	Polymorphism
14.	Class method declaration general form	class	class
15.	Creating class objects	class	class
16.	Executing Methods in Object/Circle	class	class
17.	How to define a class?	class	class
18.	How to create objects?	class	class
19.	How to add data fields and methods to classes?	class	class
20.	How to access data fields and methods to classes?	class	class
21.	Super classes and Subclasses	inheritance	inheritance
22.	Creating and Using Interfaces	inheritance	inheritance
23.	Dynamic Method Binding	inheritance	inheritance
24.	Composition vs. Inheritance	inheritance	inheritance
25.	Inner Class Definitions	inheritance	inheritance
26.	-Pass-by-value	array	array
27.	Sorting Arrays	array	array
28.	-Binary searching	inheritance	array
29.	-Allocate array	array	array
30.	-Pass-by-reference	inheritance	array
31.	Sorting Arrays	array	array
32.	-Bubble sort	inheritance	array
33.	Multiple-Subscripted Arrays	array	array
34.	Fibonacci Series	inheritance	array

35.	the first-in first-out FIFO principle	queue	queue
36.	Lists in general	list	list
37.	List	list	list
38.	Add To End(E)	list	list
39.	Efficiency of linked lists	list	list
40.	Array Size through Input	array	array
41.	Array Length	array	array
42.	Initializing Array of Arrays	array	array
43.	What are Multidimensional Arrays	array	array
44.	Unless specified, entire arrays are passed _____ and individual array elements are passed _____.	array	array
45.	If grade has the value of 60, what will the following code display? if (grade >= 60) cout << "Passed"	function	function

Table 4 : The table shows 45 questions for Computer Skill course sent by students

Example From Computer Skill course			
NO.	Questions	FRMM Material	Teacher Material
1.	second phase of the digital revolution	ch1	ch1
2.	Application Software	ch3	ch3
3.	what is Utility Software?	ch3	ch3
4.	what is Spreadsheet Software?	ch3	ch3
5.	what is Database ?	ch3	ch3
6.	Explain to me SQL	ch3	ch3
7.	Explain to me operating system	ch4	ch4
8.	I don't not understand Multitasking	ch3	ch4
9.	list Operating System Categories	ch4	ch4
10.	file-naming conditions	ch4	ch4
11.	How can manipulate files and folders?	ch4	ch4
12.	Can we store files in the root directory	ch4	ch4
13.	list Network Classifications	ch5	ch5
14.	Wired Network Basics	ch5	ch5
15.	LAN Advantages	ch5	ch5
16.	LAN jacking	ch5	ch5
17.	Cable Internet Service	ch6	ch6
18.	Internet Infrastructure	ch6	ch6
19.	What is the difference between Upstream the downstream speed?	ch6	ch6
20.	Advantages Voice over IP address	ch6	ch6
21.	what is Web ?	ch7	ch7
22.	Web Browsers	ch7	ch7
23.	E-commerce Basics	ch7	ch7

24.	Fake Web site	ch7	ch7
25.	Second phase of digital revolution	ch1	ch1
26.	Personal Computers, Servers, Mainframes, and Supercomputers	ch1	ch1
27.	What are personal digital assistant?	ch1	ch1
28.	Programs and Instruction Sets	ch1	ch1
29.	Software Categories	ch3	ch3
30.	Groupware	ch3	ch3
31.	What is desktop widget?	ch3	ch3
32.	device driver	ch3	ch3
33.	Document Production Software	ch3	ch3
34.	What is database ?	ch3	ch3
35.	What the Graphics Software?	ch3	ch3
36.	Explain Vertical market software	ch3	ch3
37.	What is the software suite ?	ch3	ch3
38.	What is the operating system ?	ch4	ch4
39.	Operating system's multiprocessing	ch4	ch4
40.	Operating System Activities	ch4	ch4
41.	Operating System Categories	ch4	ch4
42.	Today's Operating Systems	ch4	ch4
43.	Mac operating system	ch4	ch4
44.	How can I name a file?	ch4	ch4
45.	How can I give file a name?	ch4	ch4

Table 5 : The shown table 45 questions for Algebra course sent by students.

Example From Algebra course			
NO.	Questions	FRMM Material	Teacher Material
1.	1.The Set of Real Numbers	1	1
2.	Basic Real Number Properties	1	1
3.	Factoring	2	2
4.	Greatest Common Factor (GCF)	2	2
5.	How can Factoring out the GCF?	2	2
6.	Factoring a Difference of Two Squares	2	2
7.	What are Factoring Strategy	2	2
8.	How can I do adding and Subtracting for Polynomials?	1	1
9.	Domain of a Rational Expression	3	3
10.	Fundamental Principle of Rational Expressions	3	3
11.	How can I Simplify Rational Expressions?	3	3
12.	How can I calculate Least Common Denominator (LCD)?	3	3
13.	Can I Add Rational Expressions?	3	3
14.	Rational Exponents	6	6
15.	Define Equations	6	6
16.	What is Linear Equation?	6	6
17.	Solving linear equations	6	6

18.	Inconsistent Equation	6	6
19.	define Inequality	6	6
20.	what are the Inequality Signs?	6	6
21.	Interval notation	6	6
22.	Open Interval	6	6
23.	Subtraction Property for Inequalities	6	6
24.	Strategy for Solving a Linear Inequality	6	6
25.	Solving a Compound Inequality	6	6
26.	Solving an Absolute Value Inequality	6	6
27.	Rational Exponents and Roots	6	6
28.	Strategy for Solving a Linear Equation	6	6
29.	How to Solve Linear Equation?	6	6
30.	Inequality	6	6
31.	Factoring out the GCF	2	2
32.	Determine if the function is even, odd or neither?	10	10
33.	Factoring Trinomials of the Form	2	2
34.	Polynomials and non Polynomials	1	1
35.	Degree of a Term	1	1
36.	Factoring a Difference of Two Squares	2	2
37.	Adding and Subtracting Polynomials	1	1
38.	How can the numbers will be excluded from the domain?	3	3
39.	Fundamental Principle of Rational Expressions	3	3
40.	Least Common Denominator (LCD)	3	3
41.	Multiplying Rational Expressions	3	3
42.	Closed Interval	6	6
43.	Strategy for Solving a Linear Inequality	6	6
44.	Solving a Compound Inequality	6	6
45.	Solving an Absolute Value Inequality	6	6

Table 6 : The table shown 45 questions for Management Information System course sent by students.

Example From Management Information System course			
NO.	Questions	FRMM Material	Teacher Material
1.	How Intranets Support Electronic Business	4	4
2.	Capital budgeting	13	13
3.	Central Processing Unit (CPU)	6	6
4.	Parallel Processing	6	6
5.	Classifying Computers	6	6
6.	Midrange computers	6	6
7.	define Midrange computers	6	6
8.	define Distributed processing	6	6
9.	Network Computers and Peer-to-Peer Computing	6	6
10.	The Major Types of Software	6	6
11.	Language translation and utility	6	6

12.	Programming languages	6	6
13.	Enterprise software	6	6
14.	Total Cost of Ownership (TCO) of Technology Assets	6	6
15.	what is Database	7	7
16.	define entity	7	7
17.	Traditional File Environment problem	7	7
18.	define Database Management System	7	7
19.	Data definition language	7	7
20.	Data dictionary	7	7
21.	Types of Databases	7	7
22.	Hierarchical DBMS	7	7
23.	Disadvantages of Hierarchical DBMS	7	7
24.	Basic SQL Commands	7	7
25.	Define Normalization	13	7
26.	Difference of Distributed and Centralized database	7	7
27.	Data warehouse	7	7
28.	define Data mart	7	7
29.	Data mining	7	7
30.	Database server	7	7
31.	define Telecommunications	8	8
32.	Telecommunications System Components	8	8
33.	Functions of Telecommunications Systems	8	8
34.	differences between Analog and Digital	8	8
35.	Communications Channels	8	8
36.	Network Topologies	8	8
37.	Local Area Networks	8	8
38.	distance learning and E-Learning	8	8
39.	Knowledge Work Systems	2	2
40.	Controlling Risk Factors	13	13
41.	Sociotechnical Design	13	13
42.	Knowledge Work Systems (KWS):	2	2
43.	Decision Support System	2	2
44.	Executive support system	2	2
45.	Major functions of systems:	2	2

Table 7 : The table shown 45 questions for System Analysis and Design course sent by students

Example From System Analysis and Design course			
NO.	Questions	FRMM Material	Teacher Material
1.	Systems Analysis and Design Needs	1	1
2.	Roles of the Systems Analyst	1	1
3.	Qualities of the Systems Analyst	1	1
4.	Systems Development Life Cycle (SDLC)	1	1
5.	What the Analyzing System Need?	1	1
6.	Use Case Modeling	2	2
7.	Four Steps Used to Create Use Cases	2	2
8.	Why we use Case Diagrams , and how helpful they are?	2	2
9.	How do you to an Interview?	2	2
10.	explain types of Question	2	2
11.	explain Open-Ended Questions	4	4
12.	Advantages of Open-Ended Questions	4	4
13.	Disadvantages of Open-Ended Questions	4	4
14.	Closed Interview Questions	4	4
15.	Benefits of Closed Interview Questions	4	4
16.	Disadvantages of Closed Interview Questions	4	4
17.	Closing the Interview	4	4
18.	How to write Interview Report?	4	4
19.	Why we use questionnaires ?	4	4
20.	two different forms of measurement scales	6	4
21.	Problems with Scales	4	4
22.	Designing the Questionnaire	4	4
23.	How make good sample?	4	5
24.	Agile Modeling	6	6
25.	types of Prototyping	6	6
26.	explain Patched-Up Prototype	6	6
27.	explain First-of-a-Series Prototype	6	6
28.	explain Selected Features Prototype	6	6
29.	•Two main problems with the SDLC	6	6
30.	Prototyping Disadvantages	6	6
31.	Advantages of Prototyping	6	6
32.	Prototyping Using COTS Software	6	6
33.	what are RAD Phases?	6	6
34.	misunderstand Requirements Planning Phase	6	6
35.	Comparing RAD to the SDLC	6	6
36.	The Basic Principles of Agile Modeling	6	6
37.	when we use Rapid Application Development (RAD)?	6	6
38.	when we use RAD?	4	6
39.	when we use RAD Four Resource Control Variables of Agile Modeling?	6	6
40.	Four Core Agile Practices	6	6

41.	what is Data Flow Diagram?	7	7
42.	The Four Basic Symbols Used in Data Flow Diagrams	7	7
43.	Why we use data store in Data Flow Diagrams ?	7	7
44.	Steps in Developing Data Flow Diagrams	7	7
45.	explain Data Flow Diagrams	7	7

Table 8 : The table shown 45 questions for Software engineer course sent by students

Example From Software engineer course			
NO.	Questions	FRMM Material	Teacher Material
1.	What is Software?	ch1	ch1
2.	What is software engineering?	ch1	ch1
3.	Define the software engineering	ch1	ch1
4.	List the software engineering process	ch1	ch1
5.	What is a software process model?	ch1	ch1
6.	talk about software process model	ch1	ch1
7.	I don't not understand the software process model	ch8	ch1
8.	how can I calculate the costs of software engineering	ch1	ch1
9.	how know the software is good?	ch1	ch1
10.	not understanding Maintainability	ch3	ch3
11.	explain the Socio-technical system characteristics	ch3	ch3,2
12.	what are the Emergent System properties?	ch2	ch2
13.	misunderstand Emergent System properties	ch2	ch2
14.	explain Functional system components	ch2	ch2
15.	I don't not understand the System procurement process	ch2	ch2
16.	Critical Systems Types	ch3	ch3
17.	Principal dimensions of dependability are ?	ch3	ch3
18.	Can we improve the system reliability?	ch3	ch3
19.	explaining system's Security issues	ch3	ch3
20.	when Denial of service happen	ch3	ch3
21.	The project scheduling process	ch5	ch5
22.	explaining management process risks	ch5	ch5
23.	I don't not understand non-functional	ch6	ch6
24.	what is the non-functional classifications	ch6	ch6
25.	what are the classifications of the non-functional	ch6	ch6
26.	explaining user requirements	ch6	ch6
27.	software requirements document	ch6	ch6
28.	explaining Behavioral model	ch8	ch8
29.	Spiral model of the software process	ch4	ch4
30.	what is the step to design Interface	ch9	ch9
31.	what is the Patterns	ch9	ch9
32.	what is the System modeling	ch8	ch8
33.	for what is the use Context models	ch8	ch8
34.	why using Data-processing models (DFD)	ch8	ch8
35.	why using Entity-relation-attribute model	ch8	ch8

36.	why using Object models	ch8	ch8
37.	Define interaction modes	ch9	ch9
38.	Interface Analysis	ch9	ch9
39.	—Different models presented in the system from different perspectives	ch8	ch8
40.	I don't not understand software engineering?	ch1	ch1
41.	Engineering Requirements	ch6	ch6
42.	Types of requirement (level of description)	ch6	ch6
43.	6.1 Functional and non-functional requirements	ch6	ch6
44.	Domain requirements	ch6	ch6
45.	External requirements	ch6	ch6

Table 9 : Examples for message which contains taboo words

Example For Taboo Words		
NO.	Questions	Action
1.	You are donkey	The message is sent to the Academic Affairs Secretary
2.	Idiot teacher	The message is sent to the Academic Affairs Secretary
3.	A Boob teacher	The message is sent to the Academic Affairs Secretary
4.	You are booby teacher	The message is sent to the Academic Affairs Secretary
5.	You are goof teacher	The message is sent to the Academic Affairs Secretary
6.	You are goon teacher	The message is sent to the Academic Affairs Secretary

Appendix B: About University Of Palestine

Palestine University is one of the Palestinian academic higher education institutions that was established to serve the Palestinian people in the inside and the outside in particular and the Arab students in general. This University carries on its shoulders a great message which is providing a high and modernistic level of the University education through creating an educational environment supported by an integral modern and electronic techniques with permissions and academic systems with world specifications , it also provides support and help for students to ensure a high level of creation and distinction and it is concerned of the scientific and knowledgeable research with the other civilizations and sciences of the world in order to stabilize the values of good citizenship, corporation and respecting others to accomplish the welfare and human happiness.

Palestine University was established by an approval of the late president Yasser Arafat Abu Ammar in 2003, and started to practice its role in the Palestinian society alongside with its components of the local Universities on the early of March 2005 at Gaza city, through preparing the files for the approval on the colleges and programs.

By the approval of his highness the president Mahmoud Abbas a 30 acres land has been allocated in Alzahraa city at the center of Gaza city, by the time the University got the land, it started to establish its buildings and settle in it. And after the continued work in preparing the University programs according to the needs of the higher education ministry and the national organization for approval, quality and variety of the higher education institutions, the University got the initial license then the general approval at 15/7/2007 to the final approval for its colleges which are:

1. Applied Engineering and Urban planning
2. Information Technology
3. Money and Business Administration
4. Law and Judicial Practice
5. Media and Communication
6. Education
7. Dental and Oral Surgery
8. Diploma

▪ **Vision and Objects**

Palestine University is a scientific institution that stands on the basis of integration and true partnership with its counterparts the educational institutions in the operation of building and developing the society of knowledge through the following:

Building and preparing generations of scientific innovation that are intellectually and technically qualified specifically in an educational and creative environment and highly qualified programs that combines between theoretical and applied sciences by using the information's techniques in its latest applications in a way that will work on transforming the university and students to a creative and producing person that is participating in the developing operation and its demands.

The University is seeking to be a scientific and cultural center and a scientific and civilized bridge between the Palestinian and Arab environment on the one hand and the International environment from the other hand through focusing and being interested to transforming and localizing the International Technology in collaboration with the International and original Universities and institution in order to serve the comprehensive and sustainable development.

The University is committed to all its academic, administrative and student elements by practicing and accomplishing the following civilized values:

1. The moral values for the Arab and Islamic civilization.
2. The values of the complete quality and distinction.
3. The values of collective work and hard work and responsibility.
4. Promoting critical, creative and modernistic thinking.
5. Promoting lifelong self-learning.
6. Promoting knowledge and special and professional creativity in the technology technique.

▪ **Message**

Palestine University is a Palestinian high education institution which aims to prepare cadres that are scientifically and professionally qualified and able to provide the society needs through preparing the universal environment according to the quality demands to keep up with the latest sciences and techniques and promoting the role of the scientific research and knowledgeable development to contribute in supporting the continuous development efforts. We also seek for contributing to draft the future map in the form of the original principles and values of our civilization.

▪ **Strategic Objects**

1. Promoting the organizational work in the University Administration to accomplish the University vision and its message.
2. Preparing the Universal environment, which satisfy the students expectations in order to reach the creative thinking and distinction levels.
3. Connect the University programs with the Market needs and develop it to suit these needs.
4. Applying the complete quality administration at the administrative, technical and financial services according to University standard.
5. Achieving the quality and variety standards for the Academic programs according to the local and International standards.
6. Caring for scientific research and keeping up with the latest technologies and using it in the educational and applied process to contribute building a creative and distinct student who contribute in the service of his society and his nation and directing him to serve the development cases and problems in the society, and to have the University as a bridge for transforming and localizing and strengthen the sciences and technologies and to promote the scientific and cultural openness.
7. Promoting the University's relation with its social and institutional environment and strengthen its links and its scientific and cultural relations.

Appendix C : Java Code

- **Cosine_similarity**

```
static double cosine_similarity(Map<String, Integer> v1, Map<String, Integer> v2) {
    Set<String> both = new HashSet(v1.keySet());
    both.removeAll(v2.keySet());
    double sclar = 0, norm1 = 0, norm2 = 0;
    /* We need to perform cosine similarity only on words that
    * exist in both lists */
    for (String k : both) {
        sclar += v1.get(k) * v2.get(k);
        norm1 += v1.get(k) * v1.get(k);
        norm2 += v2.get(k) * v2.get(k);
    }
    return (double) (sclar / Math.sqrt(norm1 * norm2));
}
```

- **Token code for tokenize**

```
public class Token {

    private String name;
    private Type type;

    public Token(String name, Type type) {
        setName(name);
        setType(type);
    }
    public Token(String name) {
        setName(name);
    }

    public void setName(String name) {
        this.name = name;
    }

    public String getName() {
        return name;
    }

    public void setType(Type type) {
        this.type = type;
    }

    public Type getType() {
        return type;
    }
}
```

- **The main code of the implementation**

```

public static void insert_delet(int id) throws Exception {
    try {

        Class.forName("com.mysql.jdbc.Driver");
        connection =
DriverManager.getConnection("jdbc:mysql://localhost:3306/upinaredups_moodle", "root", "");
        // print();
        // int id=Integer.parseInt(JOptionPane.showInputDialog("Enter msg ID:"));
        int useridfrom = Integer.parseInt(get_useridfrom(id));
        int useridto = Integer.parseInt(get_useridto(id));
        String subject = get_subject(id);
        String fullmessage = getmsg(id);
        int fullmessageformat = Integer.parseInt(get_fullmessageformat(id));
        String fullmessagehtml = get_fullmessagehtml(id);
        String smallmessage = get_smallmessage(id);
        String notification_string = get_notification(id);
        int notification;
        if (notification_string.equals("true")) {
            notification = 1; // true
        } else {
            notification = 0; // false
        }
        String contexturl = get_contexturl(id);
        String contexturlname = get_contexturlname(id);
        double timecreated = Double.parseDouble(get_timecreated(id));
        delete(id);
        insert(id, useridfrom, useridto, subject, fullmessage, fullmessageformat, fullmessagehtml,
smallmessage, 1, contexturl, contexturlname, timecreated);
        stopwords stopword = new stopwords();
        System.out.println(" \n After ok  stopwords stopword = new stopwords(); \n");
        StringTokenizer st = new StringTokenizer(smallmessage.toLowerCase());
        System.out.println("After ok  StringTokenizer st = new
StringTokenizer(smallmessage.toLowerCase()); \n");
        // generate Pattern
        Statement pstat = connection.createStatement();
        Vector<String> strV = new Vector<String>();
        String pattern = "";
        //generate the word pattern
        ResultSet prs = pstat.executeQuery("select * from words_list");
        while (prs.next()) {
            pattern += prs.getString("word") + "|";
        }
        if (pattern.length() > 0 && pattern.charAt(pattern.length() - 1) == '|') {
            pattern = pattern.substring(0, pattern.length() - 1);
        }
        System.err.println("---->Pattern >" + pattern);
        Pattern p = Pattern.compile(pattern/"donkey|dog|shit|boob|idiot|asshole"*/);
        Matcher m = p.matcher(fullmessage);
    }
}

```

```

if (!m.find()) {
    System.err.println("not ashitty text");
    Statement stat = connection.createStatement();
    Vector<Document> DV = new Vector<Document>();
    ResultSet drs = stat.executeQuery("select * from files_upload");

    while (drs.next()) {
        System.err.println("Doc" + drs.getString("id") + " ->file name :" +
drs.getString("file"));
        DV.add(new Document("Doc" + drs.getString("id") + " ->file name " +
drs.getString("file"), stopwords.cleanText(new
QuickButCruddyTextExtractor("C:/xampp/htdocs/moodle2/ppt/" +
drs.getString("file")).getTextAsString().toLowerCase(), true));
    }
    Document dArray[] = DV.toArray(new Document[DV.size()]); //{ d1, d2, d3, d4, d5, d6,
d7};
    VectorModel vmTestArrayOfDocuments = new VectorModel(dArray, false);
    SortedMap sm =
vmTestArrayOfDocuments.getDocuments(fullmessage.toLowerCase()/*" function array function
recursion recursion"*/, true);
    ///test print result
    for (Iterator it = sm.keySet().iterator(); it.hasNext();) {
        Document myd = (Document) it.next();
        System.err.println(myd.getName() + " ___ " + myd.getCos());
    }
    Iterator it = sm.keySet().iterator();
    Pattern pP = Pattern.compile(".?");
    System.err.println("check ? -->");
    Matcher mM = pP.matcher(fullmessage);
    if (mM.find()) {
        System.out.println("? is found");
        if (it.hasNext()) {
            Document myd = (Document) it.next();
            insertsend(useridto, useridfrom, subject, fullmessage, fullmessageformat,
fullmessagehtml, "http://upinar.edu.ps:9090/upinar/ppt/" + myd.getName(), 1, contexturl,
contexturlname, timecreated);
        }
    }

    } else {
        insertsend(useridto, 5, subject, fullmessage, fullmessageformat, fullmessagehtml, "This
student have problem with Teacher and student id = " + useridfrom, 1, contexturl,
contexturlname, timecreated);
    }
} // end try
catch (Exception e) {
    // JOptionPane.showMessageDialog(null, e);
    System.out.println(e);
    System.exit(1);
}
}
}

```