

國立交通大學
資訊科學與工程研究所
碩士論文

以本體論為基礎之學習討論區主題分析器
Ontology-based Topic Analyzer for Learning Forum



研究生：蔡昂叡
指導教授：曾憲雄 博士

中華民國九十六年六月

以本體論為基礎之學習討論區主題分析器

學生：蔡昂叡

指導教授：曾憲雄 博士

國立交通大學資訊學院
資訊科學與工程研究所

摘 要

在線上學習領域中，討論區是一個提供給學習者討論在學習上所遇到問題的平台。學習者的學習歷程資訊被隱含紀錄在討論區中。在遠距教學來說，學生在討論區中的行為是老師了解學生學習歷程的一個重要的依據。因此在學生間討論文章的主題是非常值得去分析的。但在相關的研究中，像是自然語言處理，需要非常大量的訓練資料，這是在學習討論區中所沒有的。而且老師不只需要知道單一討論中在討論的主題，更需要知道主題的趨勢。因此採用資料探勘的方法來解決老師的需求。傳統的文字分群法中，特徵選擇是一個非常重要的議題，連帶的多維度更是造成了分群上的瓶頸。根據我們的觀察，討論區中的文章事實上不是完全沒有結構的文字，在不同的文體中隱含著結構的資訊。為了解決上述兩種在文字分群中會遭遇的問題，提出了以選擇不同部份的領域本體論為基礎引導分群法。所以我們提出了以本體論為基礎的主題分析器來解決老師了解學生間討論的主題。以本體論為基礎的主題分析器包含了預先定義好的規範，領域關鍵字本體論和討論區文件分類。使用這兩規範來描述文章的概念以及決定文章的種類。在以本體論為基礎的主題分析器中包含了三個行程。首先是決策表格分類器被用來把討論區中的文章做分類。再使用本體論為基礎的分群器依不同文章的類型做適性化的分群。最後用主題觀察器來突顯討論區中不同的主題。依據我們的實驗結果發現，從討論區中分析出來的主題趨勢能夠提供老師足夠的學生學習情況資訊來幫助老師教學。

關鍵字：本體論，主題分析，討論區探勘，文字探勘

Ontology-based Topic Analyzer for Learning Forum

Student: Ang-Jui Tsai

Advisor: Dr. Shian-Shyong Tseng

Department of Computer Science
Institute of Computer Science and Engineering
National Chiao Tung University

ABSTRACT

In the e-learning domain, forum is a platform provided for learners discussing the problem they encountered. The portfolio of learner is hidden in the discussion of the forum. Especially in the distance learning, the learners' behavior in the forum is important for teachers to realize their portfolio. Therefore, the topics of the discussions between learners are worth to analyze. However, the related researches such as NLP need a large amount of training data which is not enough in learning forum. Teachers need to know the trend of topics, not the topic in single document. Hence, data mining approaches are adopted to solve the needs of teachers. In traditional text clustering, feature selection is an important issue, and the high-dimension is also the bottleneck. According to our observation, the documents in the forum are semi-structured. There exists implicit structured information for specific types of documents. In order to solve the issues of feature selection and high-dimension, a guided clustering approach is proposed based on selecting different parts of global domain ontology. Therefore, an "Ontology-based Topic Analyzer (OTA)" is proposed in order to assist in teacher realizing the topic between learners' discussions. The OTA combines two predefined heuristics, Domain Keyword Ontology (DKO), and Forum Document Classes (FDCs). The heuristics are used to describe the documents concepts and determine the type of documents. Three processes are adopted in OTA. The first process is Decision Table based Classifier (DTC) used to classify the documents into different types. Next, Ontology-based Adaptive Clustering (OAC) guides clustering the documents depending on its types. Finally, Topic Viewer (TV) is used to view the trend of topics in the forum. The experimental results show that the trend of topics extracted from the learning forum can provide students' situation to assist teachers in teaching.

Keywords: Ontology, Topic Analysis, Forum Mining, Ontology-based Text Mining

誌謝

這篇論文的完成，首先要感謝我的指導教授，曾憲雄老師。在研究所兩年的歲月裡，無論是在學術研究或是為人處世方面，皆讓我受益匪淺，尤其是我學到了對一個知識領域的研究方法、邏輯思考及表達能力的訓練，這將使我終生受用不盡。同時也感謝我的口試委員，黃國禎教授、楊錦潭教授和洪宗貝教授，他們給予了我相當多的寶貴意見，讓本論文更有意義與價值。

再來要感謝的是蘇俊銘學長、翁瑞峰學長和林喚宇學長，在這段期間內，常常麻煩他們在百忙之中騰出時間與我討論並給我建議、想法，協助我修改論文。此外，我也從他們身上學習了不少生活態度及為人處事的方法，在此深表感激。同時也感謝實驗室的同窗夥伴們，芙民、嘉妮、信男、雨杰、東權、曉涵，在這兩年的時光裡，不管是學業上或是生活上，他們都陪伴著我渡過這段碩士生涯，很高興能夠交到可以這樣同甘共苦、互相扶持鼓勵的朋友。還有其他在身邊鼓勵我的朋友們，雖然無法在此一一提及，但我心裡真的非常感激有你們在我身邊。

女朋友在背後的默默支持更是我前進的動力，沒有的體諒、包容，相信這兩年的生活將是很不一樣的光景。

最後，謹以此為獻給我摯愛的雙親以及每一位支持與幫助我的人。

Table of Content

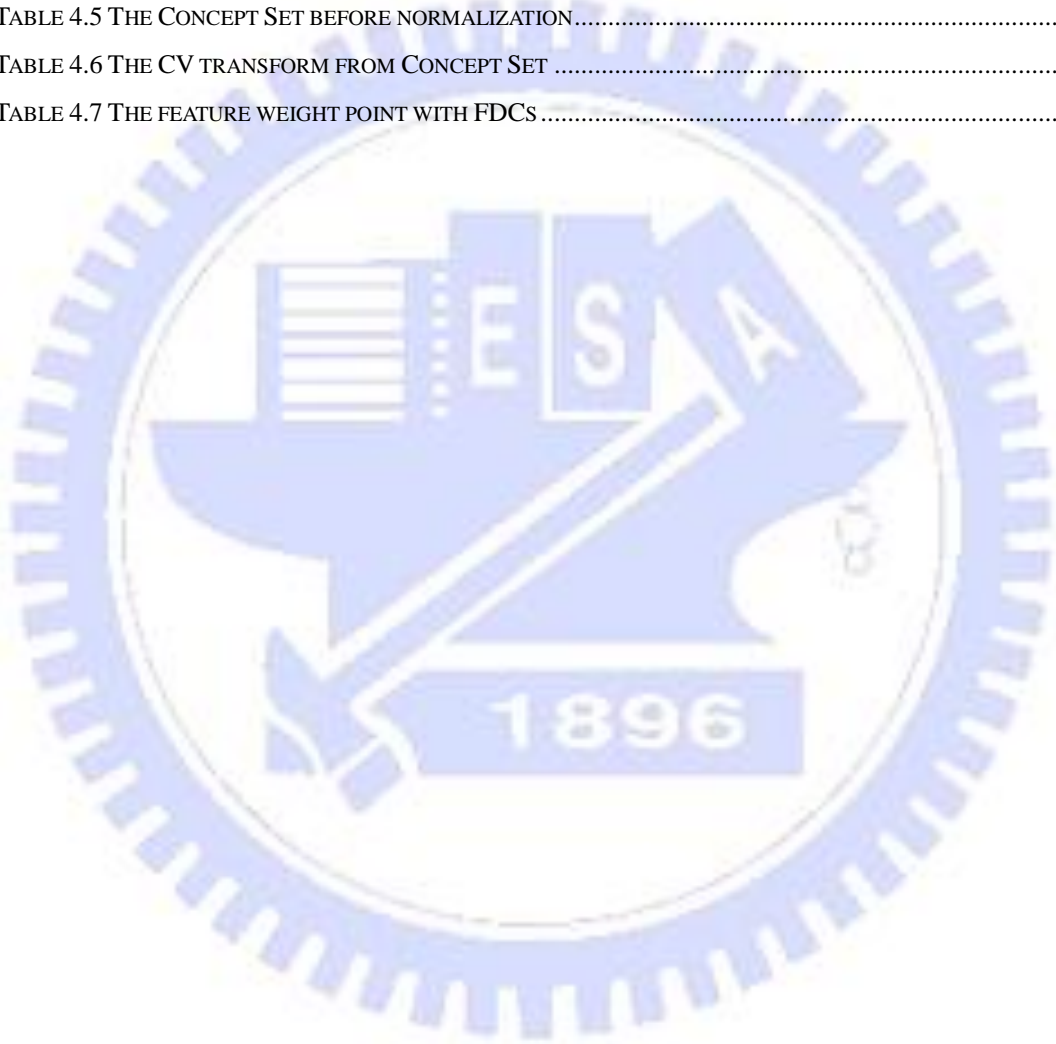
摘	要.....	I
ABSTRACT		II
TABLE OF CONTENT		IV
LIST OF FIGURES		V
LIST OF TABLES		VI
CHAPTER 1. INTRODUCTION		1
CHAPTER 2. RELATED WORK		4
2.1. Learning Portfolio Analysis		4
2.2. Discussion Forum Analysis		5
CHAPTER 3. ONTOLOGY-BASED TOPIC ANALYZER FOR LEARNING FORUM		7
3.1. The Ontology-based Topic Analyzer (OTA).....		8
3.2. Domain Keyword Ontology.....		10
3.3. Forum Document Representation		13
3.4. Forum Document Class.....		14
CHAPTER 4. THE FORUM TOPIC MINING WITH CLASSIFIER AND CLUSTERER		16
4.1. Phase I – Decision Table based Classifier		16
4.2. Phase II - Ontology-based Adaptive Clusterer.....		20
4.2.1. Forum Document Similarity Measurement		20
4.2.2. Ontology-based Adaptive Clusterer		23
4.3. Topic Viewer		25
CHAPTER 5. SYSTEM IMPLEMENTATION AND EXPERIMENT.....		27
5.1. System Implementation and Design of Experiment		27
5.2. Experiment result and explanation.....		28
CHAPTER 6. CONCLUSION		33
REFERENCE		34

List of Figures

FIGURE 3.1 THE ONTOLOGY-BASED LEARNING FORUM TOPIC ANALYSIS.....	8
FIGURE 3.2 THE REPRESENTATION OF DOMAIN KEYWORD ONTOLOGY IN C++ PROGRAMMING DOMAIN	11
FIGURE 3.3 DOMAIN CONCEPT SET CONSTRUCTION FROM RAW DATA.....	12
FIGURE 4.1 THE EXAMPLE OF DOCUMENTS SIMILARITY MEASUREMENT.....	23
FIGURE 4.2 THE HIERARCHICAL ISODATA CLUSTERING.....	24
FIGURE 4.3 THE PIE CHART OF "HOW" CLUSTERING RESULT	25
FIGURE 4.4 THE LINE CHART ANALYZE BY TIME	26
FIGURE 4.5 THE BAR CHART ANALYZE BY AUTHORS.....	26
FIGURE 5.1 THE USER VIEW OF THE OTA	27
FIGURE 5.2 THE RESULT WITH POPULARITY IN "HOW"	28
FIGURE 5.3 THE TOPIC IN "HOW"	29
FIGURE 5.4 THE TOPICS IN "WHAT"	30
FIGURE 5.5 THE TOPICS IN "WHY"	30
FIGURE 5.6 THE TOPICS IN "DEBUG".....	31
FIGURE 5.7 THE DETAILS IN THE "NEW" TOPIC OF "HOW"	31
FIGURE 5.8 THE TOPICS IN "NEW TYPE".....	32

List of Tables

TABLE 3.1 THE RECORDS OF THE FORUM DOCUMENTS	14
TABLE 3.2 FORUM DOCUMENT CLASSES IN <i>C++ PROGRAMMING DOMAIN</i>	15
TABLE 4.1 THE EXAMPLE OF DECISION TABLE.....	18
TABLE 4.2 THE FORUM DOCUMENT TRANSFORMATION.....	19
TABLE 4.3 MAPPING TO THE DECISION TABLE OF “WHY”	19
TABLE 4.4 THE TYPE HAVE BEEN PREDICTED	19
TABLE 4.5 THE CONCEPT SET BEFORE NORMALIZATION.....	21
TABLE 4.6 THE CV TRANSFORM FROM CONCEPT SET	21
TABLE 4.7 THE FEATURE WEIGHT POINT WITH FDCS	24



Chapter 1. Introduction

With the rapid growth of Internet, the web forums are getting more and more popular in various application domains, such as product discussion, online diary, information sharing, learning discussion, etc. There are also many published open-source forum systems, such as phpbb2 [1] or WebCT [2][3] etc. The forum provides a platform for users to interact and discuss specific topics with each other. With the vision of Web 2.0, a lot of information and knowledge are contributed in the forum documents by the community.

In the e-Learning domain, the forum is the most popular platform formed by the learning community for collaborative learning [10]. Therefore, there is large amount of information hidden in the logs and published documents of the learning forum. For example, students usually discuss the questions of the course, or publish their learning experience about the course in the forum. The learning portfolios and attitudes of students can be extracted from the logs of accessing the forum documents. The forum provides valuable information for teachers to find out the topics which students are concerned and encountered during e-Learning. However, there are usually hundreds of documents in the forum and it is almost impossible for teachers to realize or reply the content of the documents one by one. Therefore, an analysis assistant tool is necessary for teacher.

Since the documents in the forum are semi-structured as plain text and difficult to analyze directly with traditional statistical tool such as SPSS. Hence, researches in the forum topic analyzing are proposed [27][12][24]. These researches only deal with the system logs in the learning forum such as posting number, browsing counts or posting time. The other researches are about forum content analysis [27] [12]. They analyzed the forum documents with specific writing style and formats. The analyzing results

are reported and concluded with the patterns found in the forum. To realize the discussion topics in the learning forum, the techniques of text mining and natural language processing are needed. However, these approaches need a large amount of training data to build the model. It is costly to apply the text mining approaches.

With our observation, there is some convention in the learning forum where different document styles can be obtained from the structure information or domain specific keywords. For example, some learners may post the documents with “How to...” as the title word to ask questions. The documents for sharing the information of his/her learning experience may have longer contents. Therefore, the community of each forum has their writing convention patterns. Therefore, in this paper, we aim to the problem of how to use the domain knowledge and conventions to design an efficient and low cost forum analyzing scheme for teachers.

In order to assist teachers for topic analysis in the learning forum, the following are some idea for analyzing the topics discussed in the forum. We can refer to the expertise of forum document categorization. Generally speaking, the administrator of forum firstly has his domain knowledge to identify the discussion concepts of documents in the forum. Cooperating the identified concepts with the conventions of each forum, the categorization can be heuristically done by setting the keyword patterns occurred in the document. Since the documents are categorized, the topics of each category can be analyzed by clustering documents with similar concepts.

With the ideas above, in this paper, we propose the **Ontology-based Topic Analyzer (OTA)** to provide an assistant tool for learning forum analysis. There are two predefined domain knowledge and three analyzing processes in the OTA. The domain expertise firstly defines the **Domain Keyword Ontology** which represents the domain concept keywords. Secondly, the **Forum Document Classes** are defined and the decision table is provided for using the keyword patterns to predict class. In the

OTA analyzing process, an adaptive clustering is used according to the domain expertise behavior. A decision table based Classifier is used to determine the document class just like the expert depending on the convention knowledge to classify the documents into different document classes. After the document classes are identified, the Domain Keyword Ontology is used for analyzing the topic in the forum just like the expert relying on the domain knowledge to categorize the similar discussion documents into the same topics.

In addition to the topic, the trends of the topics are also required for teachers who want to realize the topic discussed in the forum. For example, teachers may want to know the problem which students encountered or the issues that learners not understand are. Accordingly, a topic analysis result viewer should be provided for understanding the topic analysis.

According to the OTA, we analyze a learning forum “programming-club” [4] for more than 33 thousands documents reality. The analysis result shows some conditions of program learning. For example, learners always have some problem about the difference “string” and “char*”. According to the analysis result, teachers can enhance the course of weighting more on the difference between “string” and “char*”.

With the parts mentioned above, in this thesis, an Ontology-based Topic Analyzer including the Decision Table based Classifier, Ontology-based Adaptive Clusterer, and the Topic Viewer is proposed for giving the users a visualization report and charts for assisting teachers to realize the trend of the topic analysis result.

In Chapter 2, we briefly introduce the related researches about topic analysis and learning portfolio mining. In Chapter 3 the idea and the model of OTA are described. In Chapter 4, the details of OTA including DTC, OAC, and TV are described. In Chapter 5, the OTA implementation and experiment results are presented. Finally, the conclusion is given.

Chapter 2. Related Work

To assist teachers analyze the learning behavior and learning portfolio of students for teaching, several related researches including learning portfolio mining, and topic analysis are introduced.

During learning activity, learning behaviors of learners can be recorded in database, called *learning portfolio*, including learning path, preferred learning course, grade of course, and learning time, etc., in e-learning environment. Articles [13][9][16][37][34][38] have proved that the information of learning portfolio can help teachers analyze the learning behaviors of learners and discover the learning patterns as reasons of why learners got high or low grade.

2.1. Learning Portfolio Analysis

In order to make instructors realizing the learner's learning situation, many learning portfolio mining approaches have been proposed. Most of the mining approach is using some web log mining. [13][14] applied decision tree and data cube techniques to analyze the learning behaviors of students and discover the pedagogical rules on students' learning performance from web logs including the amount of reading article, posting article, asking question, login, etc. According to their proposed approach, teachers can easily observe learning processes and analyze the learning behaviors of students for pedagogical needs. However, although their proposed approaches can observe and analyze the learning behavior of students, they didn't apply education theory to model the learning characteristics of learners. For providing the personalized recommendation from historical browser behavior in e-learning system, [43] proposed a personalized recommendation approach which integrates user clustering and association-mining techniques. Based upon a specific

time interval, they divided the historical navigation sessions of each user into frames of sessions. However, these researches only considered limited log information that can be collected in specific learning management system. Therefore, it is not enough for instructor to understand the learner's learning situation. According to Groeling [17], facilitating discussion has the potential to improve the teaching and learning experiences in traditional classroom formats, as well as in distance learning.

2.2. Discussion Forum Analysis

In the domain of e-Learning, the forum is the most popular platform for students' social interaction especially in distance learning. Dringus et al (2006) [27] has proposed a new approach to give the score to students according to the post in the learning forum. Tharrenos BRATITSIS [11] has proposed a D.I.A.S. system collecting more information about the learners in order to analyze the interactive discussion in the learning forum. But the researches above only consider the system logs in the learning forum, such as posting articles, browsing documents, or posting counts. The information of these is too rough for teachers to realize the problems that students encountered during learning.

To support the analyzing of semi-structured forum plain texts, the techniques of text mining and Natural Language Processing (NLP) are required for topic analysis. However the technique of NLP needs much training data for precision. In the learning forum analysis, the topics trends of the discussion content are more important for teachers than the topic extracted from the documents with high cost.

According to the observation, teachers will have heuristic for topic analysis. Data Mining with predefined heuristics such as ontology-based text clustering can support the situation of less of training data [15][5][6]. Ontology-based clustering has been applied to text mining since 2000 AD. But the researches in the analysis of learning

forum haven't adopted the ontology-based text mining for discussion content analysis.

Ontology based text clustering in topic analysis

In the recent research, topic analysis adopts the well known approach called ontology based text clustering [36][7][8][21][33]. In our paper, we also adopt this kind of clustering approach for find the topic of the forum documents. Ontology-based text mining will be proposed at 2000 Ontology-based Text Clustering [5]. It uses keyword ontology to present the keywords relation to measure the similarity, but didn't consider about the hierarchy of the keyword in the domain ontology.

As mentioned former, there exist three issues to extract the students' behaviors form learning forum documents:

- How to model the discussion subjects of the unstructured forum document?
- How to extract the trend of the topics since there are huge amounts of various content subjects?
- How to visualize the analysis results for teachers to easily understand the behaviors of students?

Chapter 3. Ontology-based Topic Analyzer for Learning

Forum

As mentioned above, many valuable discussion topics are hidden in the learning forum documents. The information within the learning forum is useful for teachers to know the learning problems of students. However, teachers have to read the documents one after another to realize the discussion topics of students.

Therefore, we propose a novel learning forum analyzer, which can efficiently assist teacher analyzing the topic of the forum documents with the ideas as follows.

- In the forum, the documents are unformatted. Hence, concepts are used to be the description of the documents in this paper. Therefore, we define the **Domain Keyword Ontology** to extract the concepts of the forum documents. Domain experts can dynamically edit the ontology in order to scope and index the concerned concept keywords.
- Since there are various document types in a forum, we propose a novel approach call “**Adaptive Clustering**”, which combines the Decision Table based Classifier and the adaptive Clusterer. The forum documents are firstly heuristically classified by some keyword patterns. Next, within different classes, different clustering parameters are applied since they have different concerned concepts.
- After the clustering analysis, an **Interactive** and **Visualized Tool** are used to show the global view of the topic statistical pie chart and detail topic documents list by interactively click the chart.

3.1. The Ontology-based Topic Analyzer (OTA)

According to the ideas we mentioned above, we propose a novel approach called “**Ontology-based Topic Analyzer (OTA)**”. As shown in Figure 3.1.1, the concept identifier firstly represents the documents as a set of concepts and formatted as a concept vector for further topic analysis. After the concept identifier preprocessing, there are three processes in OTA which includes *Decision Table based Classifier*, *Ontology-based Adaptive Clusterer* and *Topic Viewer*.

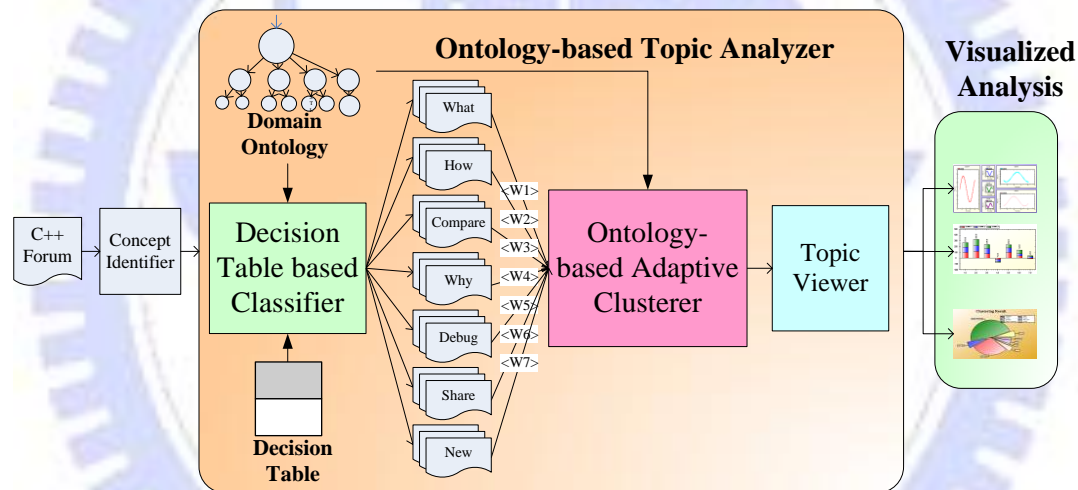


Figure 3.1 The Ontology-based Learning Forum Topic Analysis

The input of the OTA is the structured Domain Concept Set (DCS) which is defined latter. The output of the OTA is the visualization chart reports, the main processes of the OTA are described as follows.

Decision Table based Classifier:

As mentioned in chapter 2, most of the topic analysis systems only focus on single documents type, such as News. Actually, in the forum documents, there are

many classes of document types with different key points. Therefore, we use the **Decision Table based Classifier** to predict the **Forum Document Class** which we predefined, in order to describe the topic in the forum documents in different focused points. The detail of Forum Document Classes definition and the processes in different classes will be discussed later.

Ontology-based Adaptive Clusterer:

According to the FDC output by the Decision Table based Classifier, in the Ontology-based Adaptive Clusterer, we can choose a suitable classifier for each FDC to clustering each of the forum documents into the clusters which have the similar concepts. With the clustering algorithm of ISODATA, the topics result of the learning forum will be clustered by the text fields in the forum documents. The details of the clustering record will be discussed later. According to the definition of topic, we can conclude that the documents in the same clusters are most likely discussing about the same topic. Then we can explain the topic by the Cluster Center.

Topic Viewer:

After OAC clustering, the result will be passed to the Topic Viewer in order to generate the human understandable report and charts of global view, such as a pie chart of topic amount distribution analysis and the bar chart with time slide or location and the report summary by the topics in the each FDCs.

3.2. Domain Keyword Ontology

Since we want to describe the topics in forum according to **Domain Concept (DC)** of the forum documents, the **Domain Keyword Ontology (DKO)** is proposed. Next, the text of the forum document is transformed into **Domain Concept Set (DCS)** using the DKO. The formal definitions of DKO and DCS are described as follows.

Definition 1: Domain Keyword Ontology (DKO)

Domain Keyword Ontology (DKO):= <N, R>, where

- $N := \{n_1, n_2, n_3, \dots, n_k \mid n_i \in [\text{Concept}] \parallel [\text{Keyword Set}]\}$
Each node in the DKO is Concept or Keyword Set. $N := \{n_1, n_2, n_3, \dots, n_k \mid n_i \in [\text{Concept}(C)] \parallel [\text{KeywordSet}(KS)]\}$
 $N := \{n_1, n_2, n_3, \dots, n_k \mid n_i \in [\text{Concept}(C)] \parallel [\text{KeywordSet}(KS)]\}$
- $C := \langle \text{concept name (CN)} \rangle$ Each internal node in DKO is called “Concept” containing the Concept description
- $KS := \{\text{Main keyword, belong to Concept, } k_1, k_2, k_3, \dots, k_n\}$ KS only appear at leaves node, where k_1, k_2, \dots, k_n are the keywords of synonyms in the Keyword Set.

Example 1: Domain Keyword Ontology in C++ programming domain.

In the example of “C++ programming” domain, as shown in Example 1, “C++ programming” has many sub concepts such as “Object Oriented”, “Library”, etc. Each of the concepts has some sub concepts or a **Keyword Set** to describe the concept. Such as the concept “Interface” has the **Keyword Set** named “Interface” and the **Keyword Set** “Interface” has synonyms keyword “interface”, “介面” in the set

Domain Keyword Ontology (DKO)

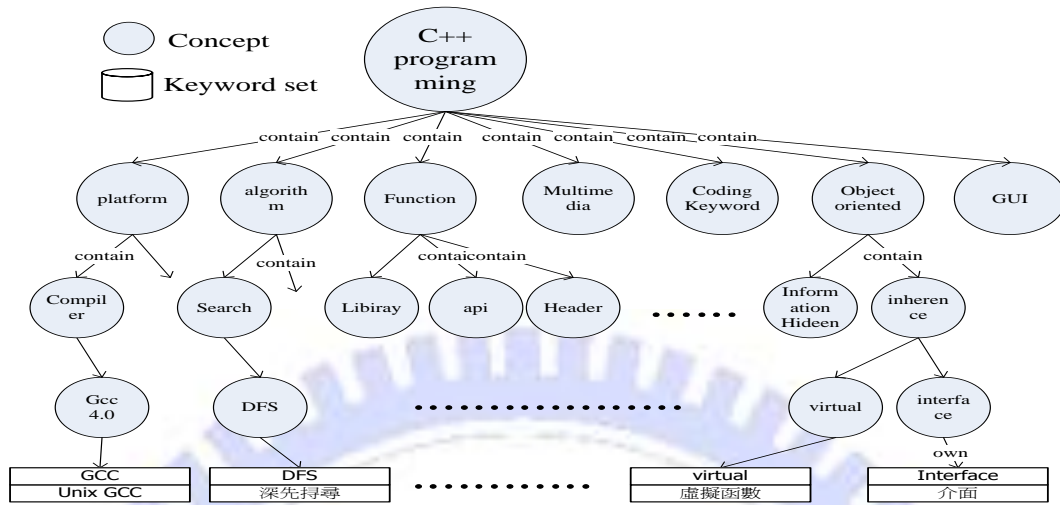


Figure 3.2 The representation of Domain Keyword Ontology in C++ programming domain

In fact, each forum document has different viewpoints of the domain for different features, for example in “C++ programming” domain, some documents in the forum are talking about how to use the tools on such platform, then actually, the forum document are considering about the issue of platform. Besides using **Domain Keyword Ontology (DKO)**, we also define the **Domain Feature (DF)**, and the DFs are represented as the Vector Space Model, called **Domain Concept Set (DCS)**. By using the DKO the formal definition of **Domain Concept Set (DCS)** is described as follow.

Definition 2: Domain Concept Set (DCS)

$$DCS = \{DFS_1, DFS_2, DFS_3, \dots, DFS_s \mid DFV_k \in DFV\}, \text{ where}$$

- $FV := \{CV, FW\}$
- $CV := \{C_1, C_2, C_3, \dots, C_l \mid C_k \text{ is the Concept in the leaves of DKO}\}$
 l is the leaves of the DKO, S is the number of the sub-tree of DKO
- $FW :=$ The feature weight

As shown in Definition 2, a Doman Concept Set (DCS) is composed of Domain

Feature (DF). Each of the DF is a feature in the domain, such like “*Object-Oriented*” or “*Platform*” is a feature of the “*C++ Programming*”, consists of the sub-tree of DKO. User can modify the weight of the feature according to the different points. The predefined seven features in the “*C++ programming*” domain include “*Platform*”, “*Algorithm*”, “*Function*”, “*Coding Keyword*”, “*GUI Programming*”, “*Object Oriented*”, “*Multimedia*” by the given Domain Keyword Ontology from definition mentioned above.

The following is the DCS example of transforming the raw data into DCS.

Example 2: Domain Concept Set Construction from Raw Data

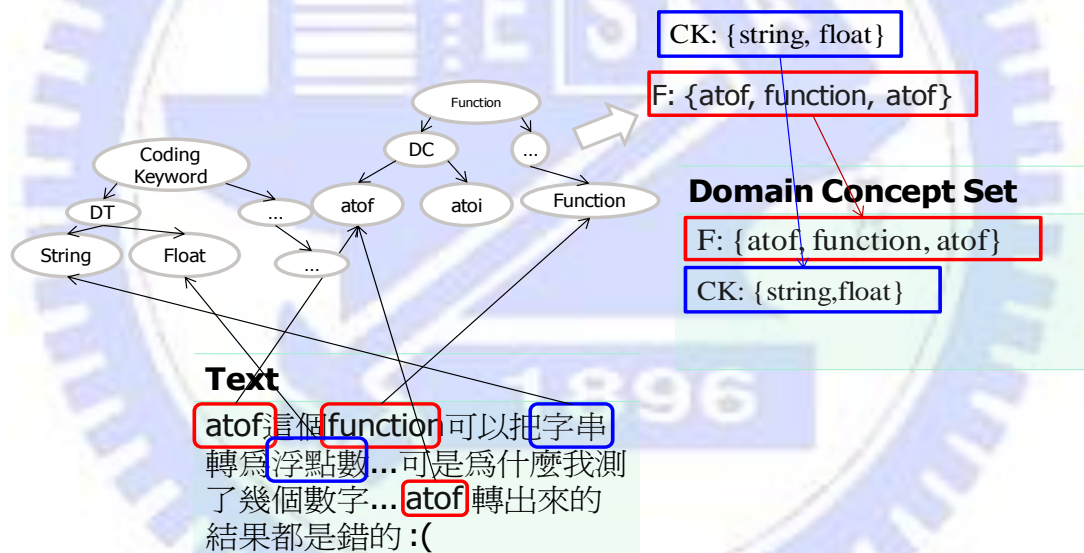


Figure 3.3 Domain Concept Set Construction from raw data

In this example, from the text of original table, we use the keyword set in the DKO to map the keywords to the concepts by the concept identifier. The text in the left-hand-side contains these keywords which is “atof, function, 字串, 浮點數, atof” all of the keywords are identified and combine to the DCS in the right-hand-side table. With the concept identifier, the text can be identified to the concept and generate the Domain Concept Set.

3.3. Forum Document Representation

As mentioned above, text can be described by the DKO and represented with Domain Concept Set., Therefore, we can consider the structure of forum document as “Title”, “Body”, and “Reply”. And in the topic analysis, teachers always have their own view points to analyze the topics. Hence these attributes include “Time”, “Location”, “Author”, and “Popularity” have been defined. The seven parts compose a forum document. The formal definition of Forum Document as follows.

Definition 3: Forum Document

Forum Document (FD) = {Attribute, Data Type, Value}

Attribute: = <T, B, R, WC, Time, L>

Data Type: = <string, string*, int, Time, double>

The attributes “Title”, “Body”, and “Reply” are the core parts of the forum document. The values in these fields are used to represent the content of the forum document. Others are the dimensions used to be analyzed.

Example 3: The forum document representation

The example shows the representation of a forum document. The text parts “Title”, “Body”, and “Reply” are represented in the Domain Concept Set format.

Table 3.1 The records of the forum documents

Title	Body	Reply	Time	Location	Author	Popularity
String	String	String*[]	Time	String	String	Double
F: <atof>	F: <atof, function, atof>	1.<...>	4/8/2007	台北	Thorx	21.46
	DT: <string,float>	2.<...>	3:00:46			
	Others: <>	3.<...>	PM			
CK:<string>	CK: <string,terminated>	1.<...>	6/8/2007	台北	阿龍	8.38
	P: <compiler>	2.<...>	9:03:10			
	Others: <>	3.<...>	PM			
OO:	F: <define, macro >	1.<...>	6/5/2007	台北	ericlin	188.62
<namespace>	OO: < namespace >	2.<...>	11:17:59			
F: <#define>	Others: <>	3.<...>	PM			

3.4. Forum Document Class

Although we have defined the *Domain Keyword Ontology* to describe the concepts of forum document, the representation of the forum has also been defined above. However, different document classes have different focused features. For example, a forum document is talking about the “*The usage of VS.net 2005*”; the topic of the forum document is about the feature which is “*Platform*”. Actually, human are more interested in the topics of the documents which begin with question style. In the “*C++ programming*” domain, we predefine these classes including “What” “How” “Compare”, the three classes are question style class. “What” class is talking about concept meaning, like “what is the window API?” The class is always talking about the feature “Platform”, “Function”, and “Object-Oriented”. “How” class is talking about something usage, such as “How to use JPEG Library?” and the topic should focus on the usage of “JPEG Library”. In this class, topic should focus on the feature of “Coding Keyword”, “Function”, and “Multimedia”. “Compare” class is talking about two things comparison, for example “The difference between ‘string’ and ‘char*’ “. Besides, in this domain, there are two more different classes, “Share” and

“Debug”. “Share” is talking about the learning experience or the new information of some concepts. The discussions of “Debug” are always about the problems of coding, this kind of forum document is always composed by a paragraph of code and some sentences of the problem description. According to the features mentioned above, we have predefined the **Forum Document Class (FDC)**. Example 4 is the predefined Forum Document Classes in our “*C++ programming*” domain.

Example 4: Forum Document Classes in *C++ programming* domain

Table 3.2 Forum Document Classes in *C++ programming* domain

Name	Describe	Example
What	Describe the definition of concept , usually used in Library	The document Title is What is STL
How	Describe how to use about the concept, ex. Coding example	The document Title is How to use STL Map
Compare	Describe the relation between two feature, ex. Platform and Library	The document Title is Where is the different between string and char*
Debug	Bug in the Code.	The document body is “<Code>” + description
Share	Describe the concept or experiment of learning concept.	Template learning experiment
New Type	Can't be analysis	

Chapter 4. The Forum Topic Mining with Classifier and Clusterer

In this chapter, there are three phases in the OTA including **Decision Table based Classifier**, **Ontology-based Adaptive Clusterer**, and **Topic Viewer**. With our idea of clustering after classification, DTC is used to predict the classification and OAC is used to cluster all of the documents in the forum by different classification. Finally, the Topic Viewer is used to visualize the charts and report by “Time”, “Location”, “Author”, and “Popularity”.

4.1. Phase I – Decision Table based Classifier

The DTC is used to predict the Forum Document Class of each document in the learning forum. As mentioned above, different Forum Document Classes will bring about different result of the topic analysis. Hence the process of predicting the FDC of the document in the forum is needed for topic analysis. By the definition, we have predefined the seven classes in the “*C++ programming domain*”. The formal definition of Decision Table is described as follows.

Definition 4: Decision Table based Classifier

Decision Table based Classifier: = (Class, Feature), where

■ Class := {“How”, “What”, “Compare”, “Why”, “Debug”, “Share”, “New Type”}

■ Feature: = {Title, Body, Reply} , where

Title, Body, Reply are the text parts of the Forum document

◆ Title, Body, Reply := Expression, where

● Expression := {C, K, WC, Operator}

- C: Concept identified by Concept Identifier
- K: Keyword identified by Concept Identifier.
- WC: Words Count.
- Operator: &, -, ∈, W#[],

By the definition above, Decision Table contains two parts which are “Class”, “Feature”, “Concept”, “Keyword”, and “Words Count”. There are seven classes that have been defined in Definition 3 and listed in the Table 3.2. “Feature” is the text parts of the forum document representation including “Title”, “Body”, and “Reply” that have been mentioned in Chapter 3. Concept and Keyword are identified by the Concept Identifier before. Words Count is used to counting the words in Feature. With the Decision Table based Classifier, we can predict the Forum Document Class of the forum document.

As shown in Table 4.1, if forum document contains the keyword “What” and the following concept belongs to “Library”, then the Forum Document Class of the document in the forum is “What”. As the definition of the Decision Table in the DTC, not only text information, but also the numerical value such as words count of each text and the relative position of each keyword are considered. If the title of the forum document contains the concept belonging to “C++” and followed by the keyword “Share” and the words count of the text in “Body” of the document is more than three hundred, then the Forum Document Class of the document in the forum is “Share”. Besides these five Forum Document Classes, a Classes “New Type” is also used in order for the extension. A Table editor will be provided for editing the Decision for extension.

Table 4.1 The Example of Decision Table

Class Name	Title	Body	Reply
What	W20[K ∈ “What” C ∈ “Library”]	*	*
	W20[C ∈ “Library” K ∈ “是什麼”]	*	*
	W20[C ∈ “Algorithm” K ∈ “What”]	*	*
How	W20[C ∈ “Library” K ∈ “How”]	*	C ∈ “Coding Keyword”
	W20[K ∈ “How” C ∈ “Algorithm”]	*	C ∈ “Algorithm”
Why	K “Why”	C ∈ “C++” & -(C ∈ “Coding Keyword”)	C ∈ “C++”
	*	K ∈ “Why” C ∈ “Algorithm”	*
	*	W10[K ∈ “Why” C ∈ “Function”]	
Compare	W10[C “Algorithm” C “Algorithm”]	*	WC > 30
	W10[C “Platform” C “Platform”]	*	*
	*	*	C “Platform” C “Platform”
Debug	*	C “Coding Keyword”	*
	K “Debug”	*	*
Share	W20[C All K “Share”]	WC > 300	*
New	*	*	*

C: concept, K: keyword*: Don’t Care, w#: window and size, -: not

Example 5: Data transformation

The example shows that how the Decision Table based Classifier is used to predict the Forum Document Class of the document in the forum. As shown in the example, the “Body” of the forum document contains the concept belonging to “Function” and contains the keyword “Why” which matches one of the records of

“Why” in the Decision Table. Consequently the forum document class will be determined by the matched rule. The forum document class is “Why”.

Table 4.2 The forum document transformation

Title	Body	Reply
F: <Atof>	F: <atof, function, atof> CK: <string,float> K: <爲什麼> Others: <>	1. CK: <float, int, float > 2. CK: <precision> 3.



Table 4.3 Mapping to the decision table of “Why”

Why	W10[K∈“Why” C∈“Function”]
-----	---------------------------



Table 4.4 The type have been predicted

Title	Body	Reply	Type
F: <Atof>	F: <atof, function, atof> CK: <string, float> Others: <>	1. CK: <float, int, float > 2. CK: <precision> 3. ...	Why

4.2. Phase II - Ontology-based Adaptive Clusterer

The process followed by the Decision Table based Classifier process is OAC-Ontology-based Clusterer. In this chapter, we will introduce about the Forum document similarity measurement and the clustering algorithm. With the idea of our approach, clustering after classification, in the OTA, we can choose the most suitable clusterer by the result of Decision Table based Classifier. And using the DKO to calculate Forum Document similarity to determine the Topic Cluster, distance based clustering has been chosen for clustering the forum documents.

Firstly, we will talk about the clusterable Vector construction, with the data processing, the document concept set will be transformed into concept vector. Depend on our similarity function; the ISODATA will be adopted as the algorithm of clustering.

4.2.1. Forum Document Similarity Measurement

In this session, we will talk about how to calculate the similarity between forum documents. As mentioned above, the Ontology-based Adaptive Clusterer will choose the suitable clusterer for different FDCs. In the following, we will use example to describe the clusterer of different FDCs.

Before similarity measurement, we have to normalize the input Document Concept Vector of the forum documents. All the DCSs have to be normalized by the frequency of the concept to the Concept Vector (CV).

As shown in Example 4, the input of the OAC is DCS. DCS has been defined above. The concepts in the DCS will be transformed into the value by frequency of usage. For example, the DCS of title “<atof>” will be transformed into “<0,0,1,0,0...>” the length of the vector are the leaves of the DKO. When all of the DCS of the forum document be transformed, followed by the step is similarity

measurement.

Example 6: Concept Set Normalization from Concept Set to CV

This example shows how to normalize the Concept Set to Normalized Concept Vector. The normalization is depending on the frequency of the concept appearance value. As shown in Table 4.2.1 and Table 4.2.2, the Concept Set in the “Body” contains the concept “atof” two times, the concept “function” appears one time. Therefore, the value of “atof” field in the Table 4.2.2 becomes 0.66 and the value of “function” in the Table 4.2.2 becomes 0.33.

Concept Set

Table 4.5 The Concept Set before normalization

Title	Body	Reply	Type
F: <Atof>	F: <atof, function, atof> DT: <字串, 浮點數> Others: <>	1. CK: <float, int, float > 2. CK: <precision> 1. ...	Why

CV

Table 4.6 The CV transform from Concept Set

Title * 0.4	Body * 0.25	Reply * 0.35
F: <1,0,0...> * 0.5	DT: <0,0,0.5,0,0,0.5> 0.1 F: <0,0.33,0.66,...> * 0.5	1. DT: <0.5,0.25,0.25,0,0...> * 0.1 2.

After constructing the CV, the distance measurement is described below. In this thesis, the Euclidean Distance function is adopted to measure the distance between CVs. The following is the definition of the distance measurement function.

Definition 5: the similarity measurement of CVs

Similarity = F (“Title”) * W_1 + F (“Body”) * W_2 + F (“Reply”) * W_3 , where

Title, Body Reply = CV, CV normalized DCS.

DCS have been defined in Definition 2.

W_1, W_2, W_3 are the weights of each parts of the forum document representation described in Definition 3.

$F(CV) = \sum S(DFS)$,

$S(DFS) = \sum ED(LCV_i)$, where

$LVC_i = \begin{cases} \text{if } i \text{ is leaves in DKO} & LCV_i = CV \\ \text{if } i \text{ is internal in DKO} & LCV_i = \text{average value of child node} \end{cases}$

ED function is Euclidean Distance Function $ED(\text{Vector}) = \sqrt{\sum_{i=0}^n (\text{Vector}_i^2)}$

Example 7: Document Similarity Measurement

In this example, the similarity measurement of the Concept Vector is shown. The Document A and Document B have the similar concepts because the two concepts are in the same sub-tree of DKO. By the definition mentioned above, the similarity of AB, AC, and BC are calculated below. Thus, we have the similarity of AB is the smallest.

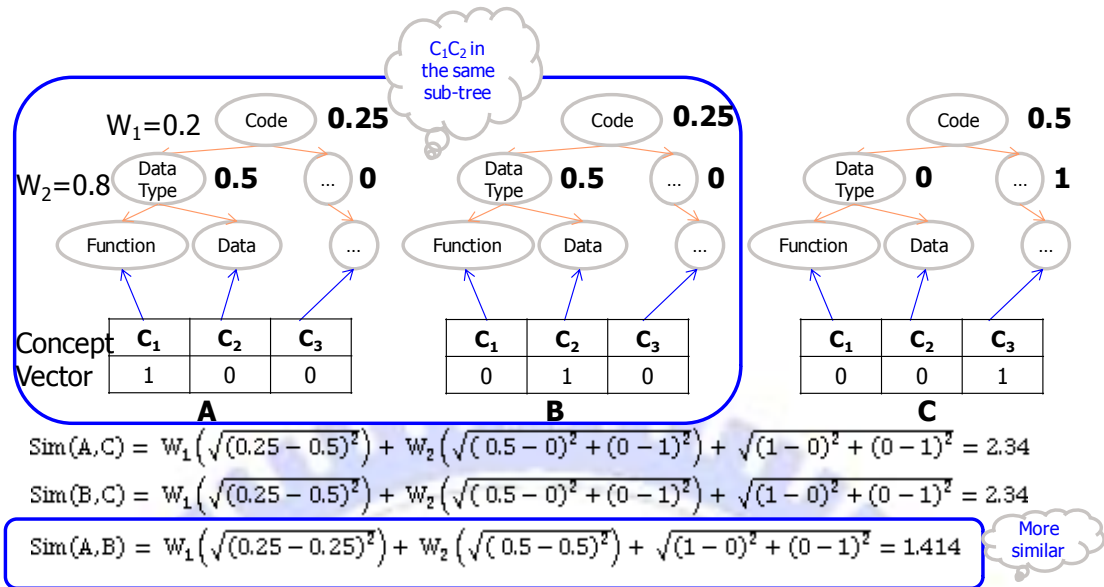


Figure 4.1 The example of documents similarity measurement

4.2.2. Ontology-based Adaptive Clusterer

With the result from the Decision Table based Classifier, all of the forum documents will have been predicted the FDCs. As mentioned above, the different FDCs have different feature weights to cluster. Therefore, in the OAC we can choose an appropriate clusterer in order to analyze the topic by different class and different feature weight. In the predefined six FDCs, “What”, “how”, “compare” the three classes have the feature that the reply of the forum document are more important than the original body of the forum document. And the forum documents in the “what” class are always talking about the concept of “Function”, “Platform”, and “Object Oriented”. So in the “what” class, the OAC will adaptive the weighting more on the three features, and weighting more on the reply then body. As shown in Table 4.7, the others feature and weight points are described.

Table 4.7 The feature weight point with FDCs

Class	Feature Weight Point
What	“Function” “platform” “Object Oriented” in “Reply”
How	“Coding Keyword” “Function” “Platform” in “Reply”
Compare	“Platform” “Function” in “Reply”
Why	“Function” “object Oriented” in “Reply”
Debug	All but “Coding Keyword” in “body”
Share	All in “body”
New	Average

After setting the feature weight of the clusterer, we adopt ISODATA to clustering. ISODATA with cluster initial number: 10, min cluster size: 3, max cluster size: 30,

Besides, the clustering result from above, we consider the time issue of the clustering. As shown in figure 4.2.

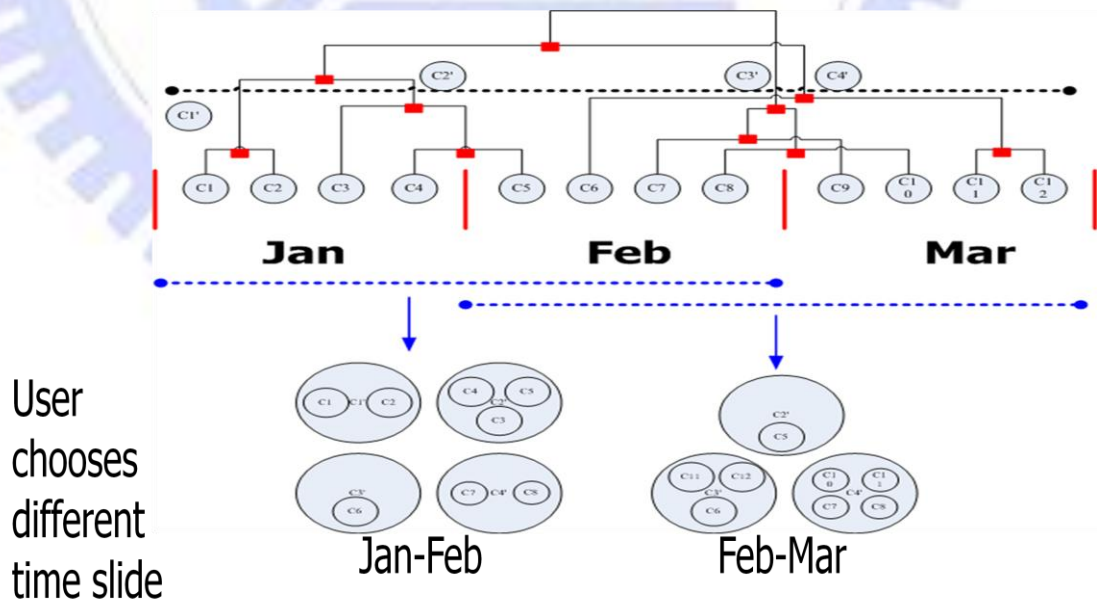


Figure 4.2 The hierarchical ISODATA clustering

4.3. Topic Viewer

The clustering result Topic Clusterers will be passed to TV, which is used to visualize the topic cluster and analyze with different dimensions including “Author”, “Time”, “Popularity”, and “Location”. In the section, we will discuss about the interpretation of the report and charts.

Analyzing with Popularity

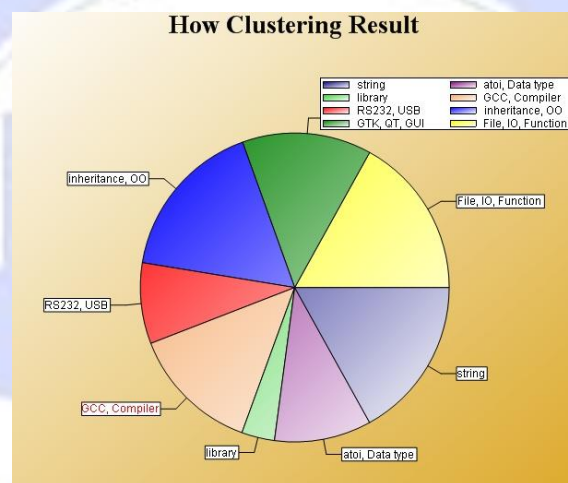


Figure 4.3 The pie chart of "how" clustering result

This pie chart can show the topic discuss popularity. In the graph, we can see the topics discussed in the forum, and the popularity of each topic. In Figure 4.4, we can see there are eight topics in the “how” class whose discussion rates are also shown in the graph.

Analyzing with Time

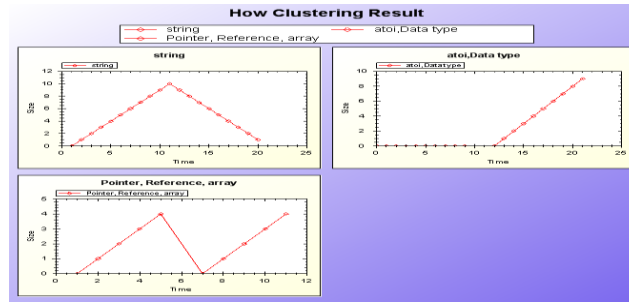


Figure 4.4 The Line chart analyze by time

This line chart can show the topic scope with time. In the graph, we can know the when the topics are discussed frequently, and when the topics start. In Figure 4.5, we can see there are three topic in the “how” class, with topics lifecycle.

Analyzing with Author

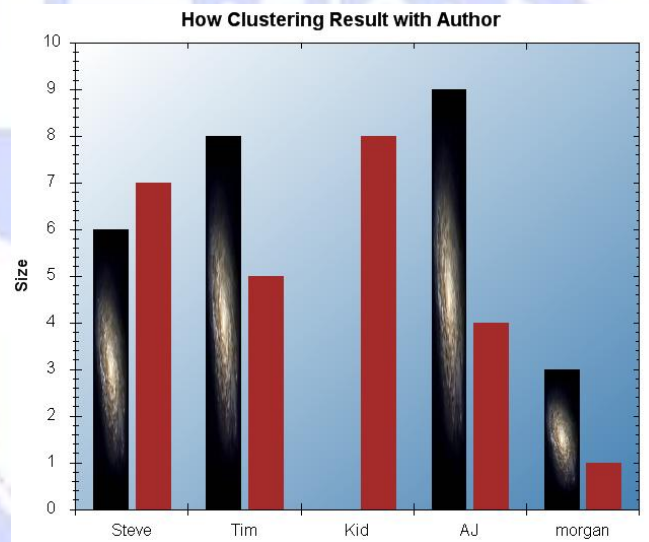


Figure 4.5 The Bar chart analyze by authors

This bar chart can show the author discussion content in each topic. In the graph, we can see the comparison of the interesting in topics. In Figure 4.6, we can see there are two topics in the “how” class and the discussion that posted by the each author.

Chapter 5. System Implementation and Experiment

For evaluating the OTA system, the OTA system has been implemented. In this chapter, the implementation and the system function will be discussed.

5.1. System Implementation and Design of Experiment

OTA system is implemented by the programming language C# on .net platform, which consists of user view and crawler. User view is shown in Figure 5.1.

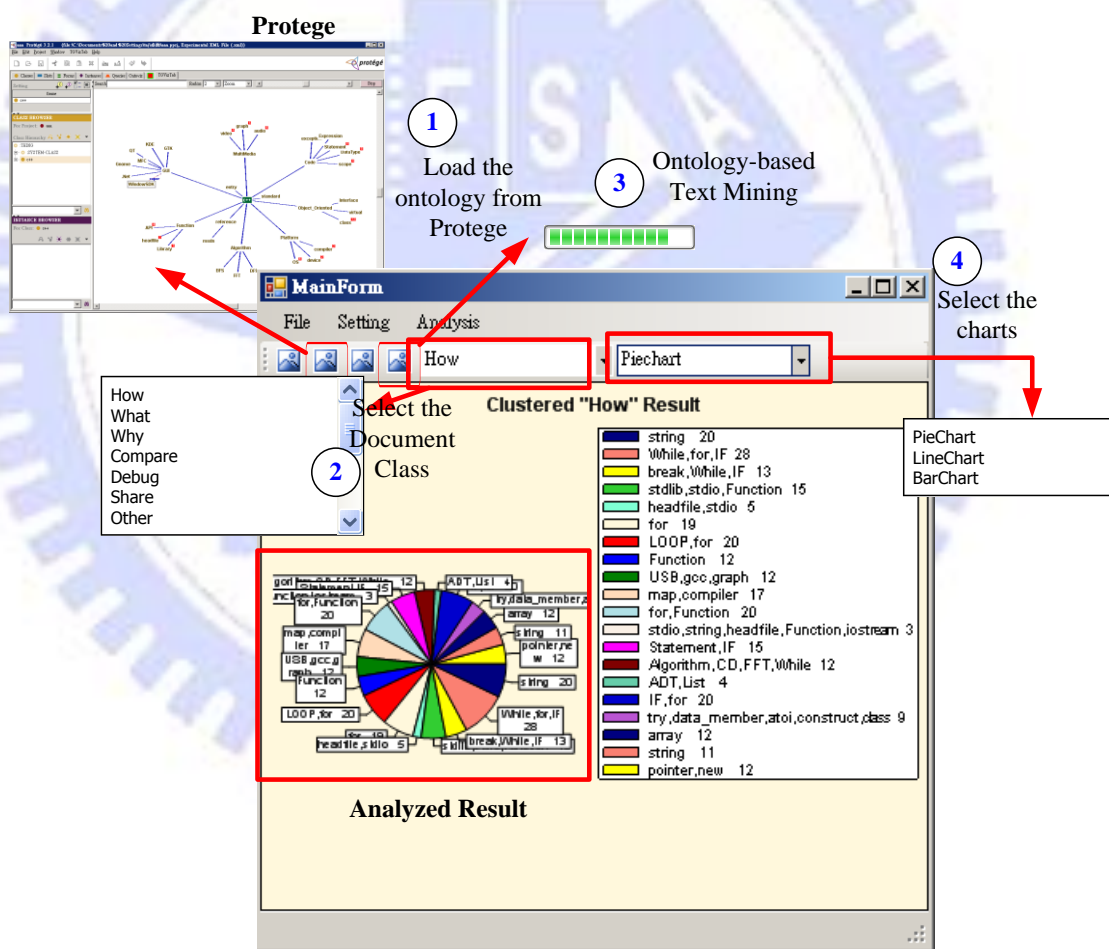


Figure 5.1 The user view of the OTA

The crawler in OTA is implemented to crawl from programming learning forum [“http://www.programmer-club.com/pc2020v5/Forum/ForumN.asp?board_pc2020=c”](http://www.programmer-club.com/pc2020v5/Forum/ForumN.asp?board_pc2020=c)

[107] every day automatically. Until now, there are more than 33 thousands forum documents that have been crawled and stored into the SQL database.

The forum is built from 2000 AD to now, and builds for programming teaching and learning. The users in the learning forum are students and the senior programmers. The documents in the learning forum are always asking some problems about the programming or the homework or learning issue in the courses. Some enthusiastic people will post documents with new and important information about the programming domain. With these documents in the learning forum, the topics of the programming domain will be showed up by the OTA system.

5.2. Experiment result and explanation

In the section, we will discuss about the experiment result from the learning forum that have been mentioned above, and describe the result of the statistical charts.

Result 1: The Result of ‘‘How’’ Document Class

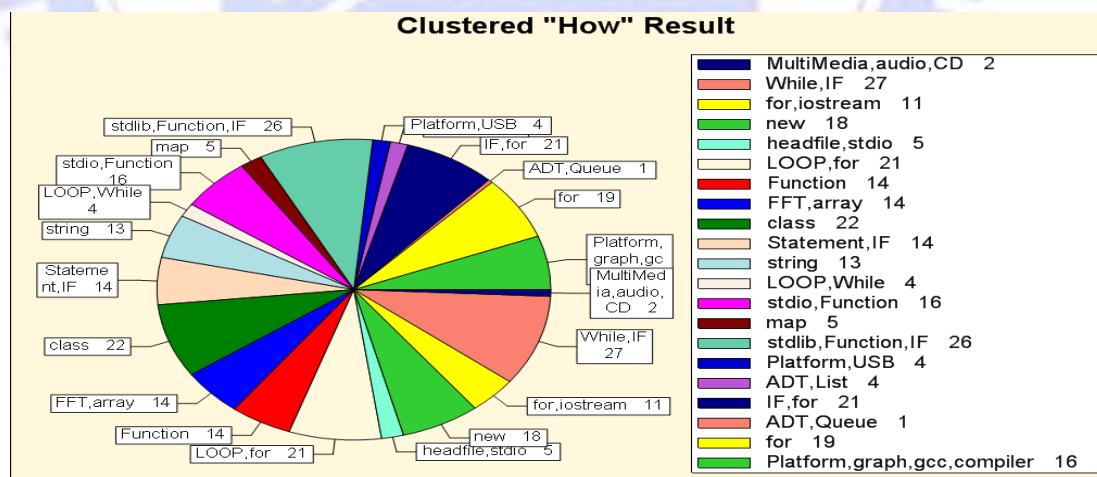


Figure 5.2 The result with popularity in ‘‘How’’

From the result we will see that in the ‘‘Statement’’ parts, about the ‘‘Loop’’ learning, there are more problems for learners in the ‘‘for’’ loop then in the ‘‘while’’

loop in the C++ domain. It is reasonable that “for” loop are more hard than “while” loop because the syntax in the “for” loop are more complex. In the “Data Structure” parts, about the part “Abstract Data Type” learning, there are more problems in the “Queue” and “List”, but there is no “Tree” or “Stack” in the “How” document classes. There may be two reasons for the situation. One of the reasons is that “Stack” and “Tree” are easier than the other two Data Structures. The other is that “Stack” and “Tree” have not been teach at the time or “Stack” and “Tree” are not in the “How” document class. Because the learning issue of the two kinds of the Data Structure and are not focused on how to use it but how to debug it.

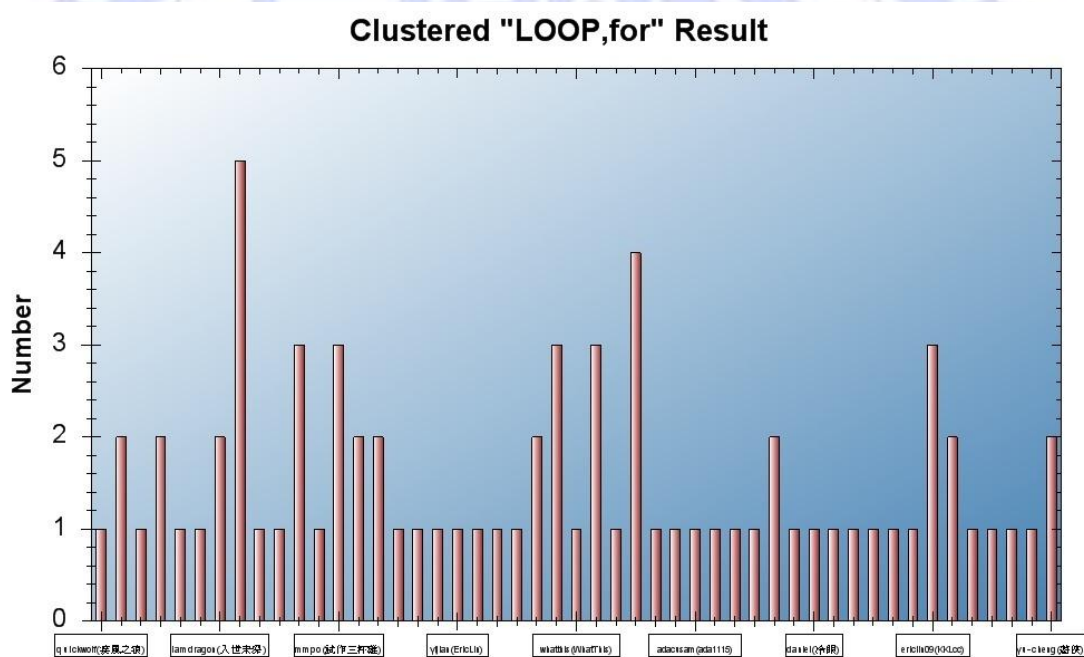


Figure 5.3 The topic in “How”

The Figure 5.3 of the result are the authors published documents number in the topic “Loop, for”. As the topic shown, we can find that the topic documents are talking about the “for loop”, and with the bar chart, we can find that the learner “tamdragon” post in the topic. Or the learner has more problems in the topic. It can conclude that the learner is more interested in the topic than other learners.

Result 2: The result of “What” document class

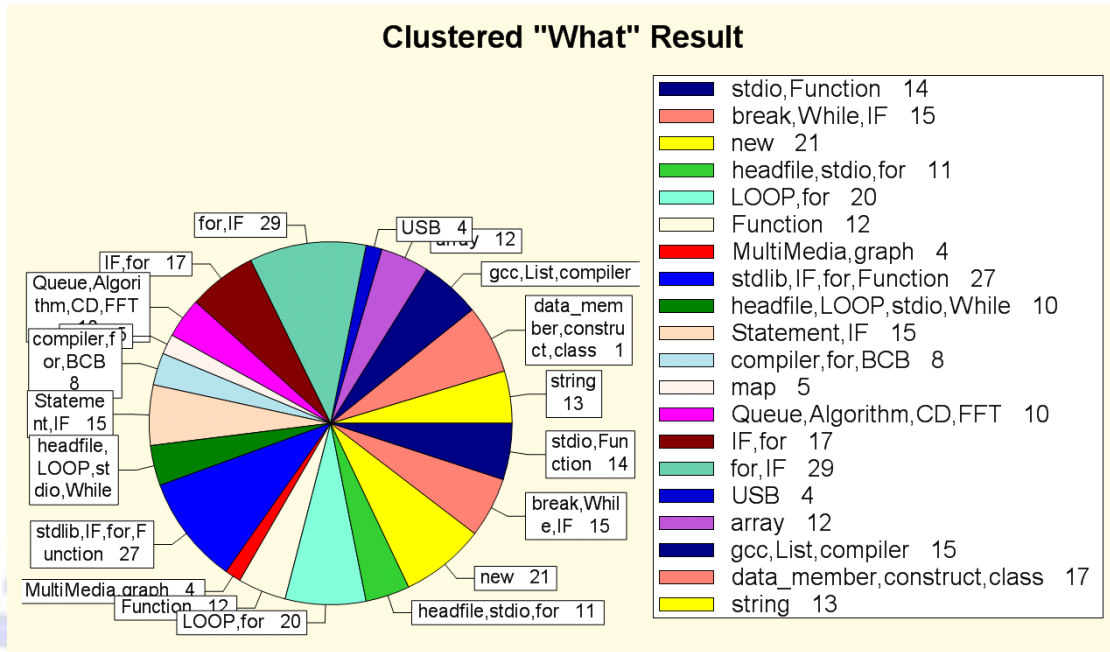


Figure 5.4 the topics in “What”

In Figure 5.4, we can find out that the compositions in “What” are similar with “How”. Actually, in the programming domain, after understanding a concept, learner is always interesting about how to use it.

Result 3: The result of “Why” document class

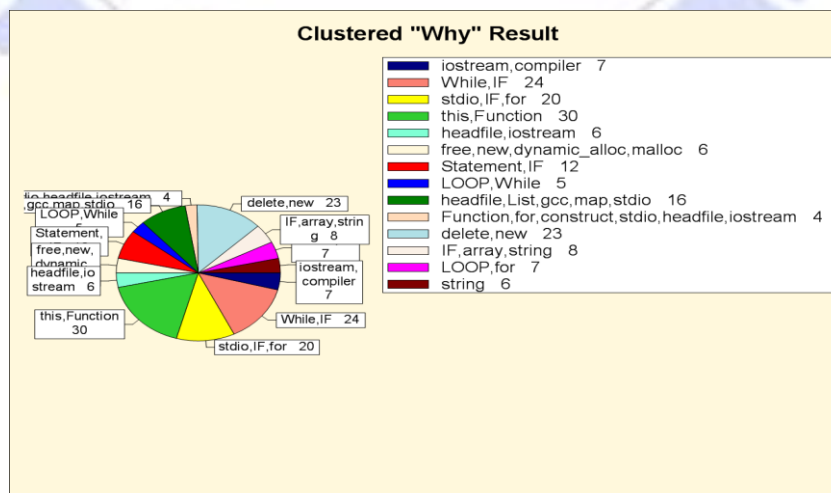


Figure 5.5 the topics in “Why”

Result 4: The result of “Debug” Document Class

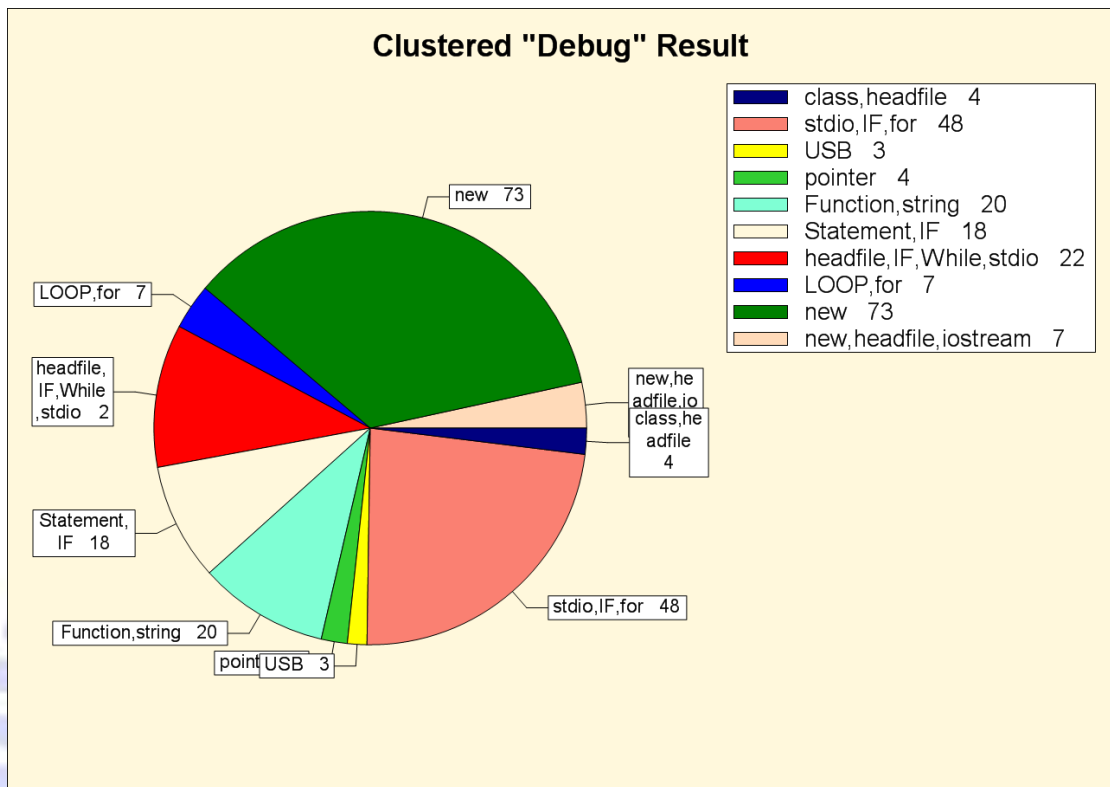


Figure 5.6 the topics in “Debug”

We can find that the topics in “Debug” are much different with the topics above. The topic “New” are focused on. “Topic” new is talking about the dynamic allocation. It is difficult for senior programming learners. Following figure is the detail in the “New” topic.



Figure 5.7 The details in the “New” topic of “how”

In the URLs list, we can find out that most of the documents in “New” are talking about the construction and destruction of dynamic memory allocation. Teachers may enhance the course on dynamic memory allocation.

Result 5: The result of the “New Type” documents

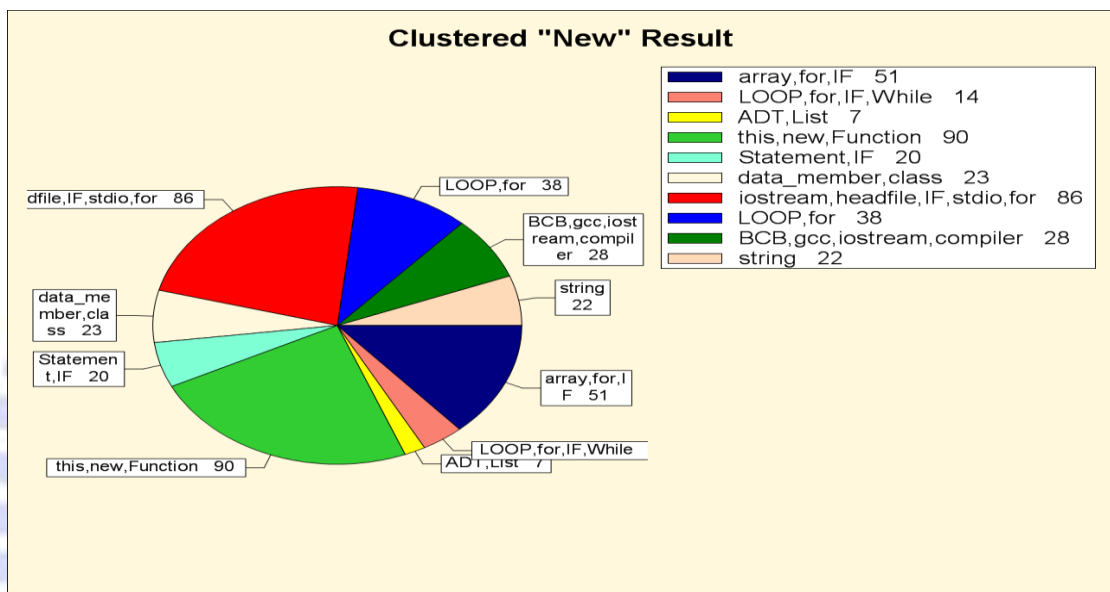


Figure 5.8 The topics in “New Type”

In the experiment, we have realized the different features in document classes by the Result 1~5. And the learning strategy that for different learner need can be proposed to teachers. We can also find the problems or the issues of learner that are popular. The trend of topics is showed to teachers.

Chapter 6. Conclusion

According to the discussion above, an Ontology-based Topic Analyzer has been proposed for assisting teachers to realize the topics in the learning forum and give the topic trend to teachers.

The OTA includes the DKO to give the forum document the description of concepts and the adaptive clustering for topic extraction by different document classes, and the TV to view the topic trend with various dimensions including “Time”, “Author”, “Location”, and “Popularity”.

With these parts, in this thesis, an Ontology-based Topic Analyzer including the Decision Table based Classifier, Ontology-based Adaptive Clusterer, and the Topic Viewer is implemented for giving the users a visualization report and charts for assisting the teachers to realize the global view of the topic analysis result.

In the experiment, we have proved that the approach of OTA for the learning forum is useful for teachers to realize the topic and the students’ behavior.

In the near future, OTA will be ported to the different open sources forum systems, such as phpBB [1] in order for popularity. And Decision Table editor will be provided for Decision Table refining. Moreover, we will extend the ontology formats to the famous formats like OWL and RDF. Since the ontology construction and Decision Table editing will be easier, the user feedback model will also be provided for refining the Table of Classifier and the Domain Keyword Ontology for more precision topic analysis. The mechanism of analysis reporting will be improved for summarization by the results of statistic.

Reference

- [1] phpBB <http://www.phpbb.com/>
- [2] Web Wiz Guide: The Web Development Site. <http://www.webwizguide.info/default.asp>
- [3] WebCT Learning System. <http://www.webct.com/>
- [4] Programming Club
http://www.programmer-club.com/pc2020v5/Forum/ForumN.asp?board_pc2020=c
- [5] A. Hotho, A. Maedche, and S. Staab. Ontology-based text clustering. In *Proceedings of the IJCAI-2001 Workshop "Text Learning: Beyond Supervision"*, August, Seattle, USA, 2001.
- [6] A. Hotho, A. Maedche, and S. Staab. Ontology-based text document clustering. *Künstliche Intelligenz*, pages 48–54, 04 2002.
- [7] A. Hotho, A. Maedche, S. Staab, V. Zacharias. On Knowledgeable Supervised Text Mining. *Proceedings of Text Mining Workshop*, Springer, 2002.
- [8] A. Hotho, S. Staab, G. Stumme. Text Clustering Based on Background Knowledge. Technical Report 425, University of Karlsruhe, Institute AIFB, 76128 Karlsruhe, Germany, April 2003
- [9] Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. Paper presented at the 11th International Conference on Data Engineering (ICDE), March 6-10, 1995, Taipei, Taiwan.
- [10] Barros, B., Verdejo M.F., Read, T., Mizoguchi, R., (2002), Applications of a collaborative learning ontology, *MICAI'2002 Advances in Artificial Intelligence*. In C. A. Coello, A. de Albornoz, L. E. Sucar, O. C. Battistutti, (Eds.), Lecture Notes in *Computer Science* 2313, Springer-Verlag pp. 301-310
- [11] Bratitsis T., Dimitracopoulou A., « Data recording and usage interaction analysis in asynchronous discussions: The D.I.A.S. System », AIED Workshops (AIED'05), July 2005
- [12] Content analysis of online discussion on a senior-high-school discussion forum of a virtual physics laboratory

- [13] Chang, C. K., Chen, G. D., & Ou, K. L. (1998). Student portfolio analysis by data cube technology for decision Support of web based classroom teacher. *Journal of Educational Computing Research*, 19 (3), 307-328.
- [14] Chen, G. D., Liu, C. C., Ou, K. L., & Liu, B. J. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research*,
- [15] Corich S., Kinshuk, Hunt L. (2004): *Assessing Discussion Forum Participation: In Search of Quality*. International Journal of Instructional Technology and Distance Learning. TEIR Center, Duquesne University, Pittsburgh
- [16] Dewhurst, D. G., Macleod, H. A., & Norris, T. A. M. (2000). Independent student learning aided by computers: an acceptable alternative to lectures? *Computers & Education*, 35, 223-241.
- [17] Groeling, T. (1999). Virtual Discussion: Web-based Discussion Forums in Political Science. Paper presented at the 1999 National convention of the American Political Science Association, Atlanta, Georgia.
- [18] Gunawardena, C. N., Lowe, C., Carabajal, K. (2000). Evaluating online learning: Models and methods. In *Society for information technology & teacher education international conference: Proceedings of SITE 2000 (11th, San Diego, California, February 8–12, 2000) (Vols. 1–3, pp. 1677–1684)*.
- [19] Henri, F. (1982). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing* (pp. 117–136). Berlin: Springer-Verlag.
- [20] He, Y., & Hui, S. C. (2002). Mining a Web citation database for author co-citation analysis. *Information Processing & Management*, 38(4), 491–508.
- [21] H-C. Chang, C-C. Hsu. Using topic keyword clusters for automatic document clustering. *The Second Inter-national Workshop on Knowledge Discovery and Ontologies at ECML 2005, Vol.1, 419- 424*.

- [22] Jarvela, S., & Hakkinen, P. (2003). The levels of web-based discussions: using perspective-taking theory as an analytical tool. In H. v. Oostendorp (Ed.), *Cognition in a digital world* (pp. 77–95). Mahwah, NJ: Lawrence Erlbaum Associates.
- [23] Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussions. *The American Journal of Distance Education*, 17(1), 25–43.
- [24] K. Sade, Y. Levo, S. Kivity “A Content Analysis of Web-Based Expert Forum on Asthma and Allergy”. *Journal of Allergy and Clinical Immunology*, Volume 117, Issue 2, Pages S54-S54
- [25] Koschmann, T. D. (1994). Toward a theory of computer support for collaborative learning. *Journal of Learning Sciences*, 3(3), 219–225.
- [26] Kirkley, S. E., Savery, J.R. & Grabner-Hagen, M. M. (1998). Electronic teaching: Extending classroom dialogue and assistance through e-mail communication. In C. J. Bonk & K. S. King (Eds.), *Electronic Collaborators: Learner-Centered Technologies for Literacy, Apprenticeship, and Discourse* (pp. 209-232). Mahwah, NJ: LEA, Publishers
- [27] Laurie P. Dringus *, Timothy Ellis Using data mining as a strategy for assessing asynchronous discussion forums *Computers & Education* 45 (2005) 141–160
- [28] L.Na.Luo, Z.Wan Li Y.Fu-Yu Z.Jing-Bo (2006/9) "Using Ontology Semantics to Improve Text Documents Clustering" *Journal of Southeast University* vols.22-3
- [29] Learning Technology Standards Committee of the IEEE Computer Society (2001). IEEE P1484.1/D8, 2001-04-06 Draft standard for learning technology – Learning technology system architecture. IEEE Standards Department: Piscataway, NJ.
- [30] Li, H., Yamanishi, K. (2001). Mining from open answers in questionnaire data. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, August 26–29, 2001, San Francisco, California (pp.443–449).

- [31] Luan, J. (2002). Data mining and its applications in higher education. *New Directions for Institutional Research* (Spring), 17–36.
- [32] Moore, M. G. (1989). Three types of interaction. *The American Journal of Distance Education*, 3(2), 1–6.
- [33] N. Jardine, C. J. Van Rijsbergen. The Use of Hierarchical Clustering in Information Retrieval. *Information Storage and Retrieval*, Vol.7, No.5, 1971, 217–240.
- [34] Quinlan, J. (1986). Induction of decision trees. *Machine learning*, 1, 81-106.
- [35] Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance education courses. *The American Journal of Distance Education*, 17(2), 77–98.
- [36] S. Bloehdorn, P. Cimiano, A. Hotho, S. Staab. An Ontology-based framework for text mining. *LDV Fo-rum – GLDV Journal for computational linguistics and language technology*, 2005, Vol.20, No.1, 87-112,.
- [37] Smith, E. S. (2001). The relationship between learning style and cognitive style. *Personality and Individual Differences*, 30, 609-616.
- [38] Shashaani, L., & Khalili, A. (2001). Gender and Computers: similarities and differences in iranian college students' attitudes toward computers. *Computers & Education*, 37, 363-375.
- [39] Schrire, S. (2003). *Journal of Instruction Delivery Systems*, 17(1), 6–12.
- [40] Tsantis, L., & Castellani, J. (2001). Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform. *Journal of Special Education Technology*, 16(4), 39–52.
- [41] Wentling, T. L., Johnson, S. D. (1999). The design and development of an evaluation system for online instruction. In *Web '99 World conference on the WWW and internet proceedings* (Honolulu, Hawaii, October 24–30, 1999) (pp. 2–7).

- [42] Williams, C. B., & Murphy, T. (2002). Discussion groups: How initial parameters influence classroom performance. *Educause Quarterly*, 25(4), 21–29.
- [43] Wang, F. H., & Shao, H. M. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications*, 27, 365–377.

