

ACADEMIE DE MONTPELLIER
UNIVERSITE MONTPELLIER II
—SCIENCES ET TECHNIQUES DU LANGUEDOC—

MEMOIRE DE STAGE DE MASTER

SPECIALITE : **RECHERCHE EN INFORMATIQUE**
MENTION : **INFORMATIQUE, MATHEMATIQUES, STATISTIQUES**

Effectué au LRMM/INFO

—

Sous la direction de Danièle Hérin

**A Personalized Recommendation System
based on the students' trails**

Par

Zhang Huaxi

Soutenu le 28 août 2006

Acknowledgment	i
Abstract	ii
Chapter 1 Introduction.....	1
1.1 Introduction.....	1
1.2 Background.....	1
1.3 Current Approaches	2
1.4 Comparison the E-commerce with the E-learning.....	2
1.5 Our Approach.....	3
Part I Basic Theories Statement	8
Problem Statement	8
Chapter 2 Semantic Design: Ontology Models	10
2.1 Domain Model	13
2.2 Course Model.....	13
2.3 Documents Model.....	16
Chapter 3 Pedagogical Model.....	18
3.1 The Felder-Silverman Model.....	18
3.2 Application of the Felder-Silverman Model for E-learning.....	20
Chapter 4 Personality Model of the Learner	23
4.1 Learning Styles of the Learners	24
4.2 Knowledge State of the Learners	25
4.3 Basic Information of the Learners	25
Chapter 5 Trail Design.....	26
5.1 Introduction.....	26
5.2 The Basic Theory of the Learning Process: A Cognitive Model	26
5.3 The Activities of Constructing the Ontology by the Students	27
5.4 The Trail Design	29
Part II Analysis of the trails and the recommendation.....	37
Chapter 6 Data Preparation.....	39
6.1 Mapping the Individual Ontology.....	42
6.2 Preprocessing	44
6.3 Support Filter	54
6.4 The Data Integration	55
Chapter 7 Pattern discovery from the trails of the learners	56
7.1 Statistical Frequent Analysis.....	56
7.2 Clustering Approaches	62
Chapter 8 Recommendations for the Learner.....	66
8.1 The KNN Based Approach	66
8.2 The Course Recommendation.....	67
8.3 The Concept Recommendation.....	68
8.4 The Document Recommendation	69
Chapter 9 Conclusion and Future Work	71
Reference	72

Vocabulary	73
Definition	73
Figure	74
Table.....	75
Annex.....	i
Annex 1 Index of Learning Styles Questionnaire.....	i
Annex 2 The result of the ILS example	viii

Acknowledgment

During my work about six months, I must thank to my teacher Danièle Hérin. She is given me many supports and encouragements. In her guide, I enter the e-learning domain. For my report, she teaches me how to write the paper more formally. Thus, I must say thank you to her in first.

Here, I must thank to the Monsieur Roland Ducournau. Because of him, I choose this topic as my stage.

The next, I must think my friend Zhang Lei. He helps me to insist my idea and go to the end. In addition, he gives me many good ideas.

Abstract

We propose a model and some algorithms of the e-learning personalization system based on the web usage mining. The source is the learner trails and we define under the models and algorithms to discovery the usage profiles and generate the recommendation to the new learners. However, according to the different requirements of the e-learning and the e-commerce, we integrate the semantic knowledge from the preprocessing phase to the recommendation phase. Importantly, we add the pedagogical model to better comprehend the learner behaviors and to generate the better recommendations.

Chapter 1 Introduction

1.1 Introduction

The main objective of this paper is to propose a framework, which uses the data mining, semantic ontology and learning styles to analyze the learner trails and then automatically generate the personal recommendations to the learner.

What is the personalization? Personalization is to personalize the learning process of a learner by dynamically tailoring the looks, the feels and the contents of the course to the learner's needs and interests. We add two things important: the trail of the learners and the pedagogical information. So our work is to personalize the learning process (learning trail) of the learner by dynamically tailoring the looks, the feels and the contents of the course according to the learner's needs (knowledge) and interests (learning styles).

The goal of our work is to well understand the learner behaviors and give the best suitable recommendations to make the learner more comfortable and effectively learn. Therefore, we consider the semantic side of the learning materials, the pedagogical side of the learners then use the data mining techniques to find the profiles. Then in the online phase, we use the learner profiles and the usage profiles to generate the recommendations. We propose three main components: the data preparation, pattern discovery and the online recommendation.

1.2 Background

The problem is how to serve dynamically the customized content (pages, concept) to users based on their profiles, learning styles and the learning requirements.

In the web, there are a huge number of documents. There are many methods to search these documents and supply the learners to learn. But the problem is how to search the documents suitable the learner preferences. While the more important problem is how to organize these documents logically to explain the topic like a course; how to order these documents to meet the learners learning styles?

For this problem, there are many related works. They use the semantic design to the learning materials to supply the personalization. They just pay attention to the information of the learning materials, but ignore the learner personal information. For example, if there are two students with the same specialty and they want to learn the same topic, obviously the ancient system will give the same recommendations to these learners. But if the learner A likes the documents using the images to explain; the learner B just like the text documents, how we can we find the suitable document to the

learners? Just pay attention to the semantic information cannot find these differences of the learners and well comprehend the learner behaviors.

Another side, all of these recommendations generate from the information of the content (including the content, structure and context) without the usage information of the materials. When there is a bad organized and bad explained document and nobody accesses it for a long time, the system will continue to give the recommendation of this document to learners. Thus without the usage information, the system miss so much useful information and the mistakes in the materials.

Thus the problem is about how to comprehend the learner behaviors, and generate the personal recommendations based on the semantic meaning of the materials, the personal pedagogical information of the learners and the usage information.

1.3 Current Approaches

In the e-commerce, there are many works to use the web mining techniques to find the user interests and give the recommendation. The Bamshad Mobasher [4] works using the web usage mining techniques to personalization. The Bettina Berendt works for the semantic web mining to e business. They develop fast and there are many good algorithms to comprehend the user behaviors and they use the web mining techniques to generate automatically the recommendations.

1.4 Comparison the E-commerce with the E-learning

In the e-commerce, they can well comprehend the users using the web mining techniques, while in the e-learning, because of the similar between he users and the learners. We think that we can use the techniques to understand well the learner learning behaviors also. Thus, we consider importing the e-commerce model in the e-learning are. But importantly, there are some differences between them. The purpose of the e-commerce is to make the user to buy the products, while the e-learning is to make the learners well master the materials. Thus, this leads the many basic differences between them. The following table gives the different points.

E-commerce	E-learning
In e-commerce, the web log data is not too long time	The learning process lasts long time
The page view time represents the interests	The page view time represents the learning time, but not the interests
The domain ontology is underlying the website, but not the user interests.	The course ontology is knowledge ontology. The learner wants to master it.
The navigation is oriented by the products.	The learning sequence is based on the learning styles.

E-commerce	E-learning
The recommended items are listed by the range.	The learner wants to know why recommend this as the next.

Table 1-1 The differences between the e-commerce and the e-learning

According to these differences, we need to consider how to solve these problems in e-learning system. I give some brief view in the table 2.

E-learning	Solutions
The learning process lasts long time	Record the ID and use the function to record the whole-learning process
The page view time represents the learning time, but not the interests	Use the significance on all the pages (if possible the exam results is better)
The course ontology is knowledge ontology; the learner wants to master it.	When give the recommendation, consider the ontology.
The learning sequence is based on the learning styles.	Use the learning styles model
The learner wants to know why recommend this as the next.	Give the lien between the recommended item and the concept in the learner ontology

Table 1-2 The solutions to the e-learning

1.5 Our Approach

The goal is to well comprehend the learner behaviors and give the best suitable recommendations to make the learner more comfortable and effectively learn. The main idea is: we consider the semantic side of the learning materials and the pedagogical side of the learning pattern and use the data mining techniques to find the profiles then in the online component using the recommendation algorithms to give the recommendations. The basic ideas are in the following.

- Design a function to enable the learner record their knowledge ontology during learning process
- Data sources for mining include: the history collected by the function, learner profiles and course ontology
- The course ontology as one style of the content, then combine the content mining with the usage mining
- Find aggregate learner profiles by automatically discovering learner access patterns through Web usage mining (offline process)
- Match a learner's active session against the discovered profiles to provide dynamic page recommendation (online process)

The basic data design composes the semantic model and the trails model.

The semantic model composes two models: the learning material model and the learner personality model. We use the semantic ontology to model these two data resources: learning materials and learner information. The Figure 1-1 gives the structure design. These ontologies are written manual by the teachers.

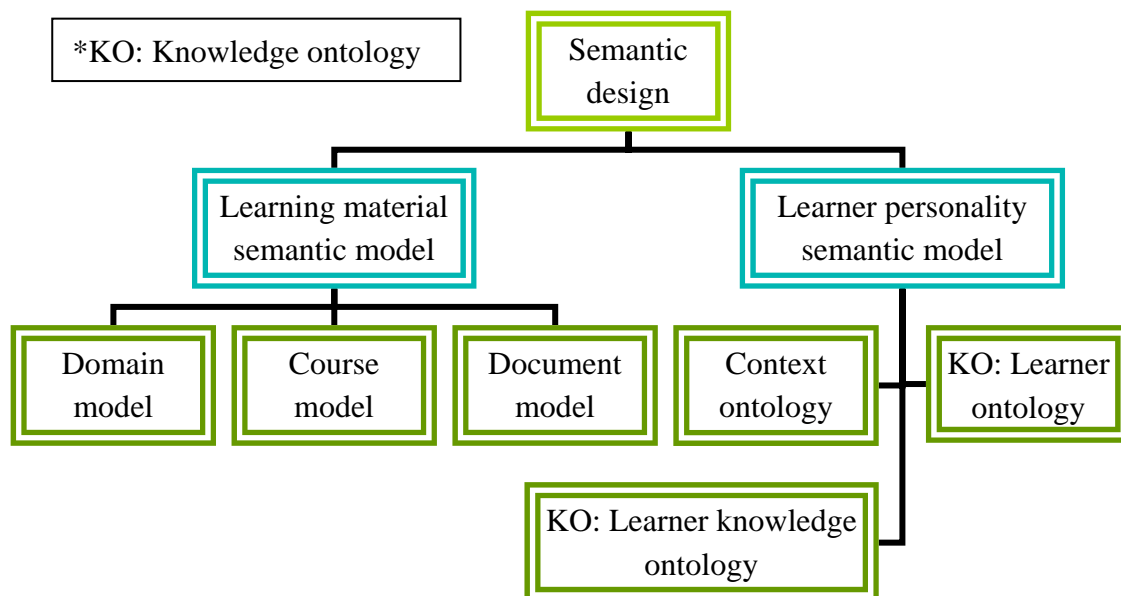


Figure 1-1 Semantic Model

The trail design uses a cognitive model to simulate the learning process and then give the detailed design. Figure 1-2 is the design of the trails. The learners input manual their learning concepts and the system capture these trails automatically.

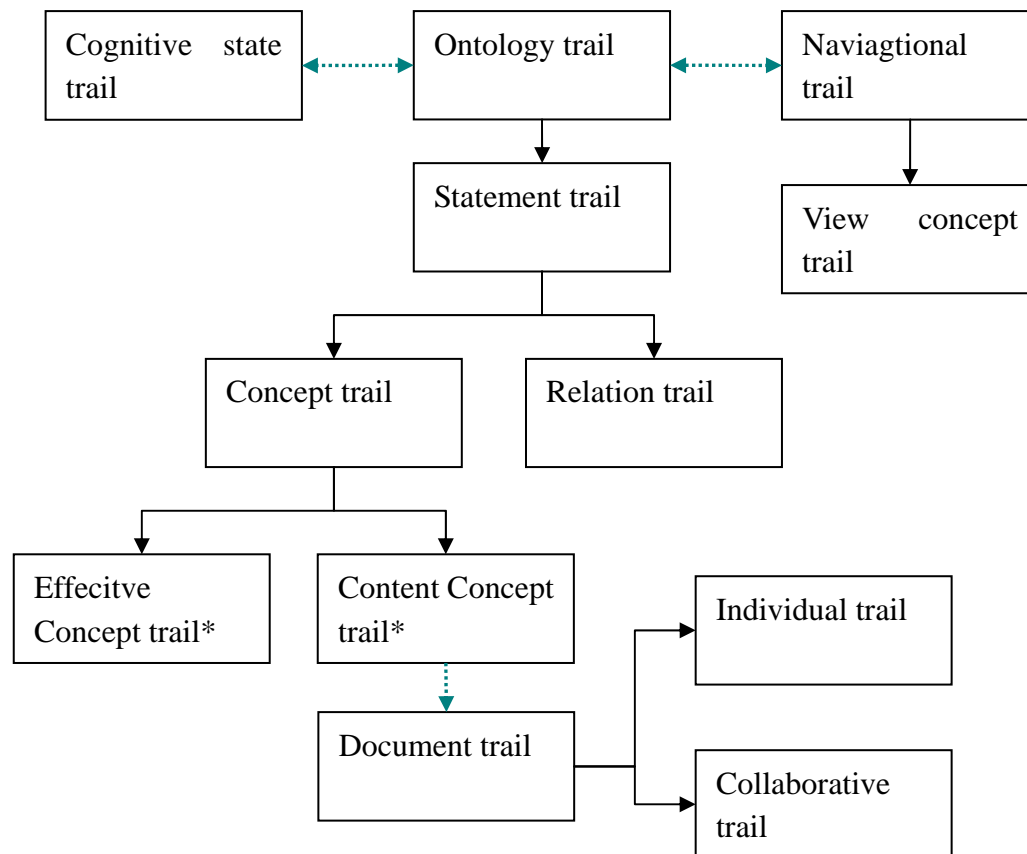


Figure 1-2 Trail model

There are three main components to generate the recommendations: the data preparation, the pattern discovery and the online recommendations. The first two phases are offline and the last is online. All these three components are automatically.

First is the data preparation component of Figure 1-3. We ignore the data-cleaning. We just pay attention to the data semantic preparation. We weight the learner ontology, mapping the learner ontology to the domain ontology and then pruning the learner ontology.

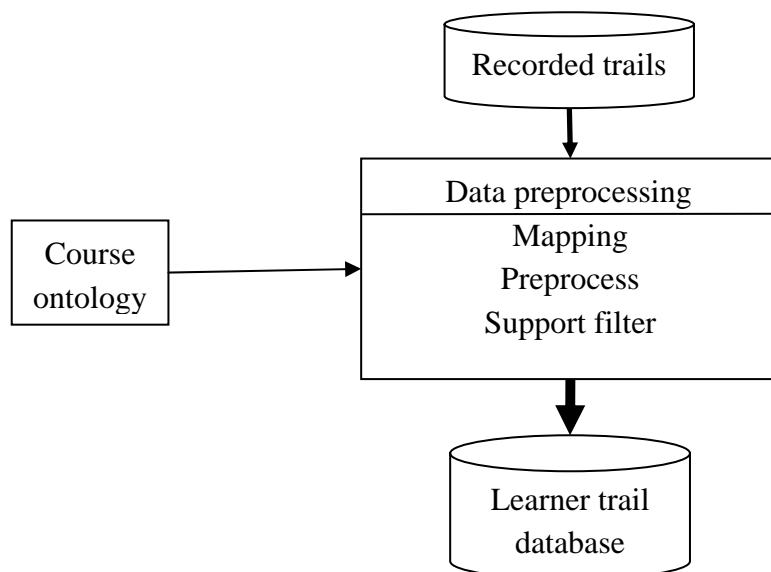


Figure 1-3 Data Preparation Component

The second component is the pattern discovery. We use the data mining techniques to the user traces database, which we obtain from the data preparation component. Then we consider the semantic information and the pedagogical information in the pattern analysis component. Like this, we can explain the mining results and analyze the aggregate profiles for the next component: online recommendations.

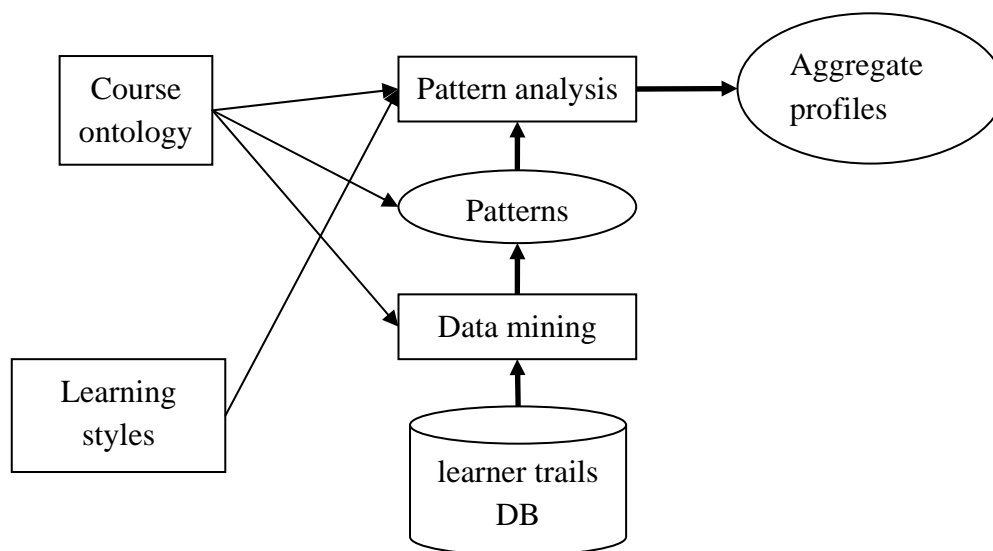


Figure 1-4 Pattern Discovery Component

The last component corresponds to the online recommendation phase Figure 1-5. In this component, we capture the learner active session and integrate the learner profile. Then we apply the recommendation algorithms.

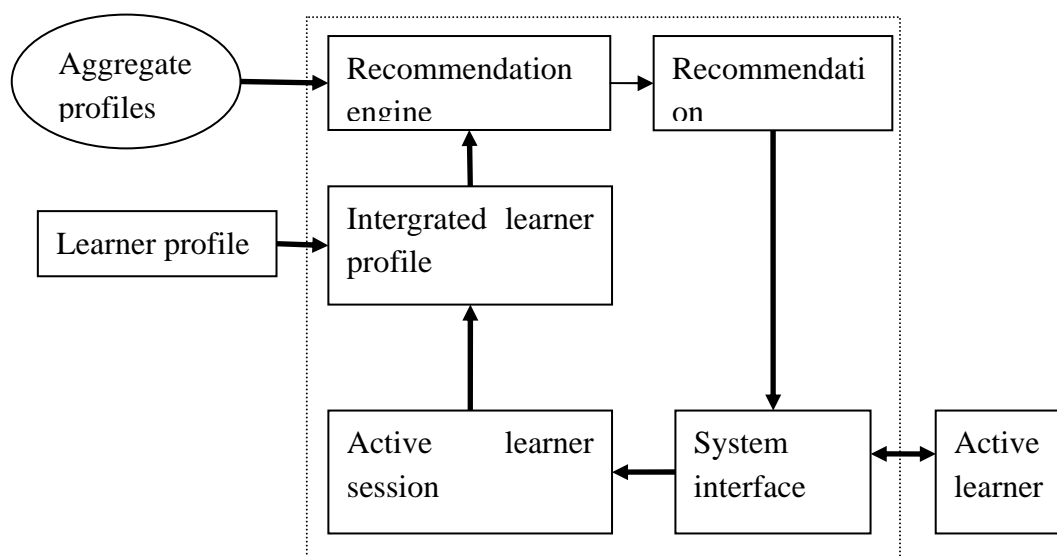


Figure 1-5 Recommendation Component

There are two parts of this paper, part one: the basic theories part, and part two: analysis and recommendations. The first part composes the chapter 2, 3, 4 and 5. The chapter 2 describes the semantic level and the chapter 4 is the personalization. These are the basic level. The chapter 3 presents the theories of the pedagogical model. The chapter 5 defines the trail of the learner. The second part composes the chapter 6, 7 and 8. In the chapter 6 we propose the data preparation and the chapter 7 is pattern discovery. At last, the chapter 8 is the recommendations.

Part I Basic Theories Statement

Problem Statement

Data capture method: The learner can record their knowledge ontology during the learning process.

This function is designed to permit the learner to record his knowledge ontology during the learning process. Like when we read a book or follow a course in class, we prefer to make the note for the reviewing and better remember. This is also the same in e-learning, this is the basic difference between the e-learning and the e-commerce.

Example 1

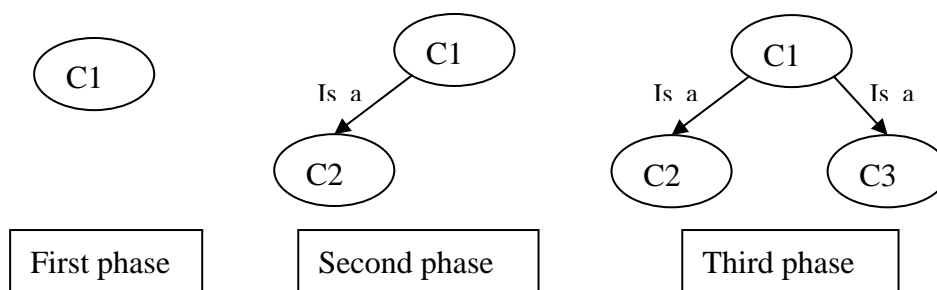


Figure I-1 the traces of the learner ontology

Data source: the traces of the learner ontology

We capture the process of the learner constructing the ontology. This is like the web page that the user marks during the visiting. This exactly affects the learner purpose and knowledge statement during the learning process. Then this is the trail of the ontology graph. We will give the detailed design of this source.

Example 2

1	Create the head concept: web data mining
2	<web data mining, is_a, web content mining>
3	< web data mining, is_a, web structure mining >
4	< web data mining, is_a, web usage mining >

Table I-1 The trail of the example 1(Figure I-1)

Recommendations

1. The next concept to learn

In this recommendation, we supply the learner the next concept to be learned. We give the list of the recommended concepts, the most range in the top. When the learner click one concept, then this concept will appear in his knowledge ontology with the color and with a relationship to the concept that has already in his ontology.

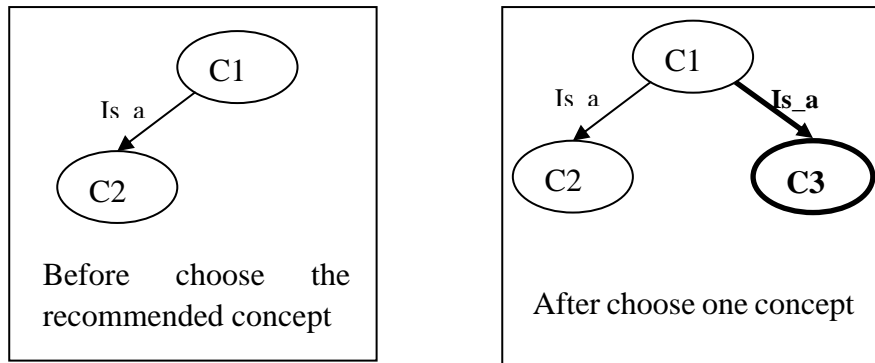


Figure I-2 The ontology before and after the next concept to be learned

2. The next document to be recommended

The next page to be recommended is the content document that the learner should be learned. In fact, it is the next document, which we recommended. When the learner chooses one concept, then this document will display in a list using the range.

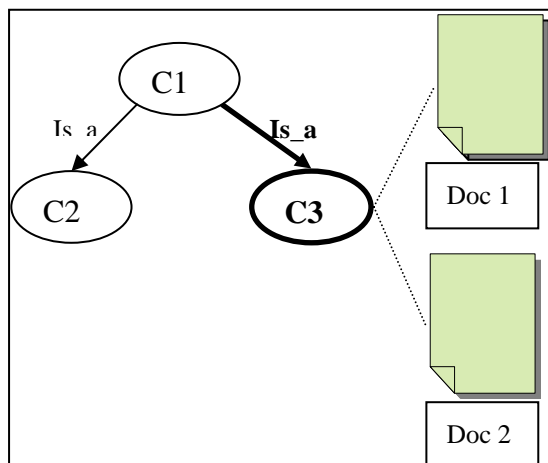


Figure I-3 The next document to be recommended.

Chapter 2 Semantic Design: Ontology Models

The semantic design is the basic level in the e-learning system. That is the key to transfer the learner navigation trail into the meaningful learning trail, as follows:

1. The learning content is organized by the semantic meaning.
2. The learning sequence is oriented by the semantic purpose.
3. It is the core of the content mining.

Therefore, a well semantic design facilitates the teacher to well organize his course, the learner to well understand the course, the system to better analyze the trails.

Problem

In the traditional universities, the teacher often divides the course into small lectures and the basic structure of these lectures is the knowledge ontology. Each lecture is composed by many materials like the course ppt, the support papers, the exercises and son on. In the class, the learners can well follow the teacher, have the global impact of the course, and well learn each lecture. But in the ancient e-learning environment, it uses the hierarchy structure to represent the course, and the documents are classified by the keywords. This leads many disadvantages to the learners. First of all, this leads the learner feel confused without the important points and also makes the learner just pay attention to the width but miss the depth of the knowledge. Secondly, they just take into account the content side of these materials, but miss their context information.

Solution

According to this problem, we divide the semantic model into three levels: the domain level, the course level and the document level. For each level, we design a model.

- The domain model contains the domain ontology, which is given by the experts.
- The course model describes course metadata. It is composed by two kinds of ontologies: the course structure ontology (in this, we call it structure ontology not the content ontology, the reason will be given in the session 2.2) and the course context ontology.
- The document model describes the document metadata. It contains the document content ontology and the document context ontology.

The domain ontology is in the top of this model, it is the basic of the document database. The documents are mapped to each concept in the domain ontology. The two kinds of course ontologies are designed by the teacher to give the course. The content ontology is a part of the domain ontology. Its concepts of the content ontology are supported by the document ontology. So we can say that the course ontology connects the domain

ontology and the document ontology. Then this course model makes the database with meaning and can be supplied to the e-learning. It is not like the traditional e-learning database, the LOs are just organized by the domain ontology. The Figure 2-1 gives the order of these levels.

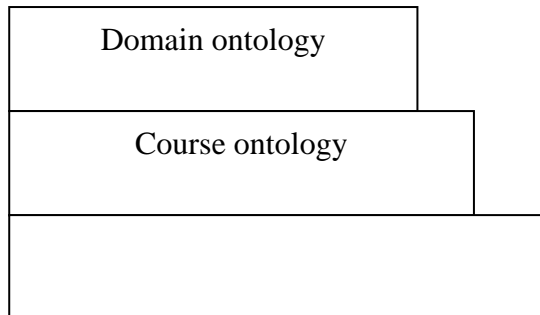


Figure 2-1 Three levels of ontology model

The course supports the concept of the domain ontology.

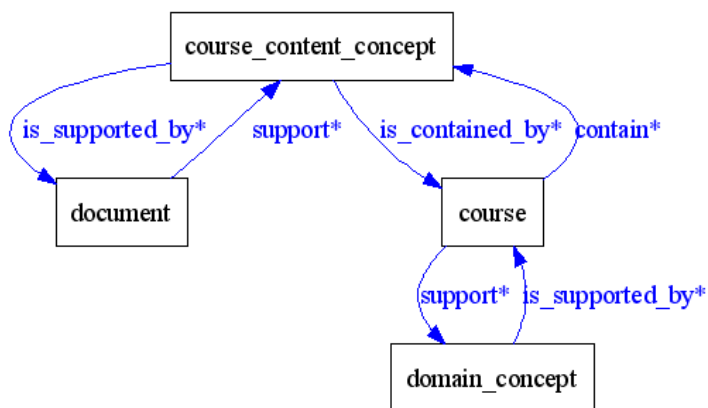


Figure 2-2 Metadata model of the relationship between the three ontology

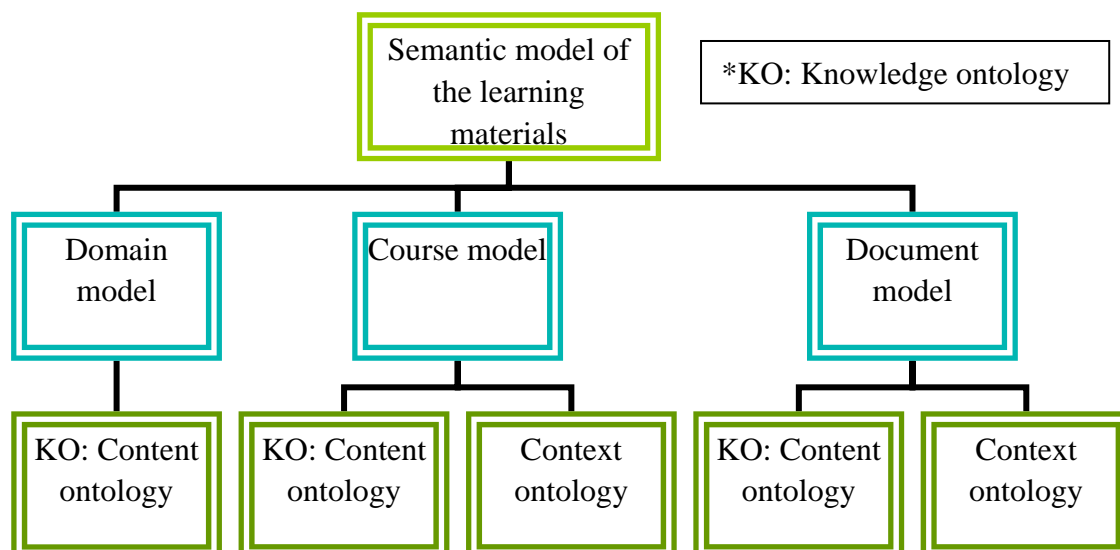


Figure 2-3 Semantic model of the learning materials

We first give the definition of the ontology. The ontology definition contains an is-a hierarchy of relevant domain concepts, relations between these concepts, further properties of concepts (attributes with value ranges), and the derivation rules to infer new knowledge. The following is the formal definition.

Definition 1: Ontology:

An ontology O is a 4-tuple $\langle C; R; I; A \rangle$, where C is a set of concepts, R is a set of relations between these concepts, I is a set of instances and A is a set of axioms.

Knowledge ontology

In these three models, they all contain a content ontology. In these ontologies, they almost have the same definition. Thus, we give the basic model of the content ontology.

Concept: in the content ontology, there is just one kind of concept: the knowledge concepts. This concept represents one knowledge node in the knowledge structure.

Relations: Relations have four main types: the relation types are mainly four types: pre-acquit, is-component, is-a (generalization/specialization), and is-application.

The following is the basic representation of the content ontology.

knowledge_concept		
is_component_of	Instance*	knowledge_concept
has_component	Instance*	knowledge_concept
has_preacquit	Instance*	knowledge_concept
is_preacquit_of	Instance*	knowledge_concept
is_application_of	Instance*	knowledge_concept
has_application	Instance*	knowledge_concept
is_a	Instance*	knowledge_concept

Figure 2-4 Content Ontology

2.1 Domain Model

Definition 2: Domain ontology

The domain model mainly gives the design description of the domain ontology. The domain ontology represents the domain knowledge content ontology. This is given by the domain experts.

The domain ontology contains all the knowledge concepts of the topic. Therefore this domain ontology is large. In order to make the domain ontology more useful and comprehensive, I define two levels of the domain ontology. The first is the main domain ontology.

Definition 3: Main domain ontology

This ontology contains the main concepts of this domain; the concept represents a large topic. This ontology is composed by the sub main topics of this domain. And some times, this ontology can present one specialty. Thus, the teachers choose the important sub topics as the courses of one specialty. In fact, its part often is the discipline ontology.

Definition 4: Sub domain ontology

This ontology supports the concepts in the main domain ontology. It is as more detailed as it can be.

2.2 Course Model

This level combines the domain ontology level and the document level. The domain ontology level is high level of the topic, so it is not detail enough. While the document ontology is often very detailed and its content just supports a small part of a concept in

domain ontology. Thus we design a course level to combine these two levels.

This model is designed to supply the meaningful information of the course from the content side to the context side.

2.2.1 Content Ontology

This ontology contains the knowledge structure of the course. This ontology is created by the teachers. It depicts the knowledge concepts that the course refers to and the course structure.

1. The content ontology units all the main concepts that the teacher wants to show to the learners.
2. The teacher uses these concepts to divide the course into small parts.
3. The teacher utilizes the relations to plan the course in order.

Relations:

The relations between these concepts are two types: the knowledge relation between the concepts and logical relation that the teacher adds and then use to order these concepts.

The relation types are four types: pre-acquit, is-component, is-a (generalization/specialization), and is-application.

When the teacher first designs the course content ontology, they can add the documents to the concepts, or if not, the system can automatically add the documents to the concepts using the document ontology.

2.2.2 Context Ontology

An ontological representation, which respects Buccholz assertion [4] “A representation of the context information should be applicable throughout the whole process of gathering, transferring, storing, and interpreting of context information”.

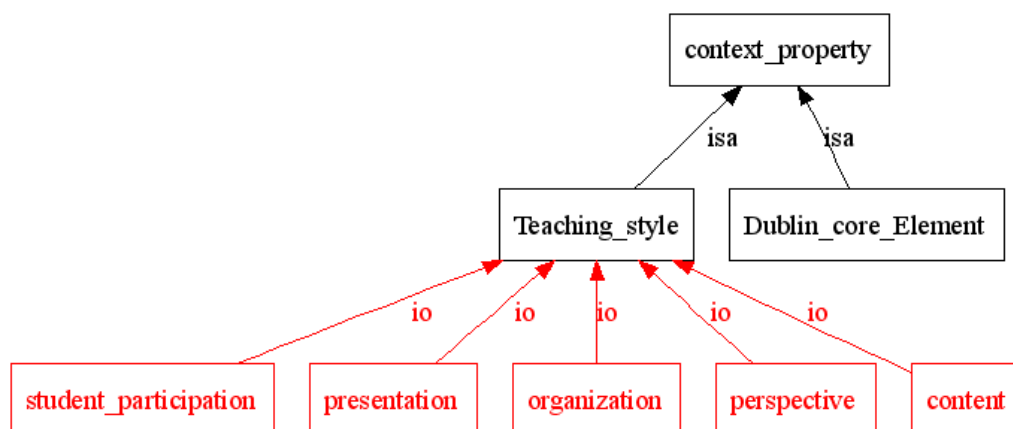


Figure 2-5 Context Ontology

The context ontology contains the information besides knowledge of the course. We use the Dublin core and pedagogical concepts to describe the metadata of the course.

- The Dublin core: we use these elements as the basic description of the course.
- The teaching style metadata: they give the teaching styles of the course. They come from the data mining results. We will give the details in the chapter 6.

1. Dublin core: we use Dublin core to define the basic metadata of the course. In the annex 1, I will give the detailed description.

Metadata	Element	Description
Coverage	Contributor	An entity responsible for making contributions to the content of the resource.
	Creator	An entity primarily responsible for making the content of the resource.
	Publisher	An entity responsible for making the resource available.
	Rights	Information about rights held in and over the resource.
Content	Coverage	The extent or scope of the content of the resource.
	Description	An account of the content of the resource.
	Type	The nature or genre of the content of the resource.
	Relation	A reference to a related resource.
	Source	A reference to a resource from which the present resource is derived.
	Subject	A topic of the content of the resource.

Metadata Coverage	Element	Description
	Title	A name given to the resource.
	Audience	A class of entity for whom the resource is intended or useful.
Instantiation	Date	A date of an event in the life cycle of the resource.
	Format	The physical or digital manifestation of the resource.
	Identifier	An unambiguous reference to the resource within a given context.
	Language	A language of the intellectual content of the resource.

Table 2-1 Dublin core element

2. Teaching style metadata

The pedagogical concepts have a main concept: learning style and it contains five sub concepts. Each concept has a value from 1 to 6. This value corresponds to the kind of the teaching style. The detailed information will give in the chapter 3.

Metadata Coverage	Element	value
Teaching style	Content_Concrete	1 to 6
	Content_Abstract	1 to 6
	Presentation_Visual	1 to 6
	Presentation_Verbal	1 to 6
	Organization_Inductive	1 to 6
	Organization_Deductive	1 to 6
	StudentParticipation_Active	1 to 6
	StudentParticipation_Passive	1 to 6
	Perspective_Sequential	1 to 6
	Perspective_Global	1 to 6

Table 2-2 The teaching style element

2.3 Documents Model

In the document model, there are two kinds of ontologies: the content ontology and the context ontology. The content ontology represents the knowledge content, which the document explains. And the context ontology presents the context information of the document.

2.3.1 Document Content Ontology

The content ontology expresses the document knowledge content. Many techniques exist to extract the content ontology from the document. We consider this ontology given by the author. So this content ontology expresses not only the knowledge concepts that the document supports, but also the structure of this document.

1. We use the RDF to represent the ontologies of the content and the titles of the documents. Using RDF statement, resource is the concept of the ontology and the property is the relationship between the concepts and titles. The value is the titles.
2. In this ontology, the relationships between the concepts are mainly four types: pre-acquit, is-component, is-a (generalization/specialization), and is-application.
3. We construct the ontology of the documents not only make the relationship between the documents to the ontology exist. In fact, we construct the ontology of the documents itself. Because we can use the ontology of the documents to manage the documents.

2.3.2 Document Context Ontology

The context ontology contains the metadata of the document. These metadata are used to give the recommendation to the learners and will be revised dynamically after mining process. This context ontology has the same structure with the course context ontology. We use the Dublin core also, but the teaching style change into the learning style.

The learning styles information comes from the learners. The system records all the learners who learn this document. From these data, we extract the main learning style characters of these learners. Then we use this as the document learning styles attributes. The more detail about how to add the learning style to the document will given in the chapter 6. the detailed information is the same as the Table 2-2.

Chapter 3 Pedagogical Model

The pedagogical model is my main contribution of this master these. Many researchers talk about the pedagogical design in e-learning system, but none of them indeed takes in account the pedagogical model. The pedagogical information is the key to comprehend the behaviors of the learners and is basis to give the recommendations.

At first, I will explain the two basic concepts: the learning styles and the teaching styles.

Definition 5: Learning style

The learning styles: learning styles are “characteristic cognitive, affective, and psychological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment”[1]. The concept of learning styles has been applied to a wide variety of student attributes and differences in classroom education. Because the learning styles directly affect the learner behaviors and the learning process. But in the e-learning domain, the research is very poor.

Definition 6: Teaching style

The teaching styles: the teaching styles are the methods how the teacher transfers the knowledge to the learners. This corresponds to the learning styles. In the classroom education, the teaching styles work for the professors to ameliorate their teaching results and to meet the learner’s needs. But in the e-learning system, we notice that the recommendations come from the ontology in some researches and some other come from the real teachers. The teaching results are not as good as in the classroom. Thus, we need to bring this model to the e-learning recommendation mechanisms to best suitable the learners’ needs and automatically generate.

In this chapter, I will first give a model of the learning styles and the teaching styles. Secondly, I will explain in the e-learning system, how it works. At last, I will discuss how to apply it in our system.

3.1 The Felder-Silverman Model

The learning styles models have been developed in several dozen models. I choose the model of Felder and Silverman[3]. This model calls Felder-Silverman model and is populate used in the engineering education. (Considering of the future work, I want to realize this system, so I use the model of the engineering education). To this model, Richard Felder and Barbara Soloman develop the Index of Learning Styles (ILS), which is a forty-four-item forced-choice instrument to access preferences on the four scales of the Felder-Silverman model. The ILS is the important reason that I choose this model. I will introduce this model and ILS in the following.

3.1.1 Learning Styles

In the Felder-Silverman model, a student learning style defines by the answers to four questions. We record the five questions, while the third question has been deleted by the author because of considering that many learners drop the examples during the learning process. I have no experiment data now, so I hold this for the future work to examine its function in the e-learning. The question 3 is the deleted one.

The five questions are the following:

1. What type of information does the student preferentially perceive: sensory—sights, sounds, physical sensations, or intuitive— memories, ideas, insights?
2. Through which modality is sensory information most effectively perceived: visual— pictures, diagrams, graphs, demonstrations, or verbal—written and spoken words and formulas?
3. With which organization of information is the student most comfortable: inductive— facts and observations are given, underlying principles are inferred or deductive—principles are given, consequences and applications are deduced?
4. How does the student prefer to process information: actively—through engagement in physical activity or discussion, or reflectively— through introspection?
5. How does the student progress toward understanding: sequentially—in a logical progression of small incremental steps, or globally—in large jumps, holistically?

According to the answers to these five questions, we can give the recommendation to the learner if there is no set of the trails data of the learner as the training resource and if we have the set of the trails data resulting from the mining phrase, we can use this to explain the mining results and anticipate the mining process.

3.1.2 Teaching Styles

Teaching style are defined in terms of the answers to five questions:

1. Which type of information is emphasized by the instructor: concrete— factual, or abstract—conceptual, theoretical?
2. What mode of presentation is stressed: visual—pictures, diagrams, films, demonstrations, or verbal— lectures, readings, and discussions?
3. How is the presentation organized: inductively—phenomena leading to principles, or deductively— principles leading to phenomena?
4. What mode of student participation is facilitated by the presentation: active—students talk, move, reflect, or passive—students watch and listen?
5. What type of perspective is provided on the information presented: sequential—step-by-step progression (the trees), or global—context and relevance (the forest)?

The learning styles are the view from the learner side and the teaching styles are the view from the teacher side. The following table describes the correspondance the comparison of the learning and the teaching styles.

Preferred Learning Styles			Corresponding Teaching Style		
Sensory	Perception	Question 1	Concrete	Content	Question 1
Intuitive			Abstract		
Visual	Input	Question 2	Visual	Presentation	Question 2
Verbal			Verbal		
Inductive	Organization	Question 3	Inductive	Organization	Question 3
Deductive			Deductive		
Active	Processing	Question 4	Active	Student participation	Question 4
Reflective			Passive		
Sequential	Understanding	Question 5	Sequential	Perspective	Question 5
Global			Global		

Table 3-1 Dimension of Learning and Teaching styles

3.1.3 ILS: Index of Learning Styles

The Index of Learning Styles is a forty-four-items forced-choice instrument developed in 1991[2]. In 1994 several hundreds sets of responses to the initial twenty-eight-item version of the instrument were collected and subjected to factors analysis. Items that did not load significantly on single factors were discarded and replaced by new items to create the current version, which was put on the WWW in 1997[2]. The ILS is available at no cost to the individuals who wish to assess their own preferences. Thus, this is very suitable for us to make all the students to assess their learning styles using the ILS when they register the e-learning system. Then we use this as the learning styles personal information of the learners to analyze the whole trails and generate the recommendations.

In the annex 2, I give the ILS questionnaire and the annex 3 is a result example.

3.2 Application of the Felder-Silverman Model for E-learning

I will give the synthesis of how this model affects the learner behaviors in the e-learning environment. We explain from the five questions of the FS model. In the following, it must be attention that the course is not only just the structure of the course, but also contains the documents, which the teacher assigns to the course concepts. So it is an entire course pattern.

1. **The type of information in the document and the course:** Which type of

information is used in this document or this course: concrete—factual, or abstract—conceptual, theoretical? This mainly concerns with the representation method of the course or document. For the course, if a teacher is prone to the learning style of concrete, he may assign many documents, which represent the information in the same type. Thus this can affect the course choice.

2. **The type of the presentations of the documents:** What mode of presentation is stressed: visual—pictures, diagrams, films, demonstrations, or verbal—lectures, readings, and discussions? For example, a ppt is more visual than a word document. And also, if some document contains more pictures than others in the same topic.
3. **The organization of the presentation of the documents and also of the course:** How is the presentation of the course or the document organized: inductively—phenomena leading to principles, or deductively—principles leading to phenomena?
4. **The type of learning activities:** What mode of student participation is facilitated by the presentation: active—students talk, move, reflect, or passive—students watch and listen? This will be very important to give the recommendations in the large e-learning system, which contains several components like the chat, email.
5. **The perspective way of the learner to learn the course:** What type of perspective is provided on the information presented: sequential—step-by-step progression (the trees), or global—context and relevance (the forest)? This is mainly about how the learner views the learning materials.

Preferred Learning styles	Affection in the e-learning
Perception	The document information type The main course documents infor. type
Input	The documents types The main course documents types
Organization	The document organization method The main course documents organization method
Processing	The course activities design
Understanding	The learning sequences

Table 3-2 The main parts which the learning styles affect

Preferred styles	Learning	Corresponding teaching styles	Corresponding Recommendations
Perception		Content	Document Course
Input		Presentation	Document Course
Organization		Organization	Document Course
Processing		Student participation	Course
Understanding		Perspective	Course

Table 3-3 The recommendation item that corresponding to the styles.

Chapter 4 Personality Model of the Learner

In this chapter, we describe the design of personality model of the learner. The personalization contains the personal information of the learners. We define three kinds of information: the basic information, the learning styles and the knowledge state.

Problem

Our purpose of the first part is to comprehend and study the learner behaviors. The main two factors of the behaviors are the knowledge structure and the personality. In the chapter 2, we have already used the semantic design to discover the semantic meaning of the behaviors. Thus the second main problem is to find the determinant factors of the behaviors. Each learner behaviors are different, because they are determined by many factors like the specialty, the degree, the learning styles, and the knowledge state. These factors refer the personality of the learners. So the problem is how to design the learner personality

Solution

According to the characters of the learner, we use three parts to define the learner personality: the basic information, the learning styles and the knowledge state.

1. The basic information of the learner is used in all the e-learning systems, and this is the basic rule to classify the learners. Thus, we use this also.
2. The next basic characters of the learner come from the pedagogical model. In the pedagogical domain, the educators find that every student has his own learning styles and these-learning styles strongly affect their learning behaviors. Thus, we add the learning styles in the personality model of the learner.
3. The last is the knowledge state of the learners. This knowledge state is represented by the learner knowledge ontology and contains all the concepts that the learner has already learned or mastered.

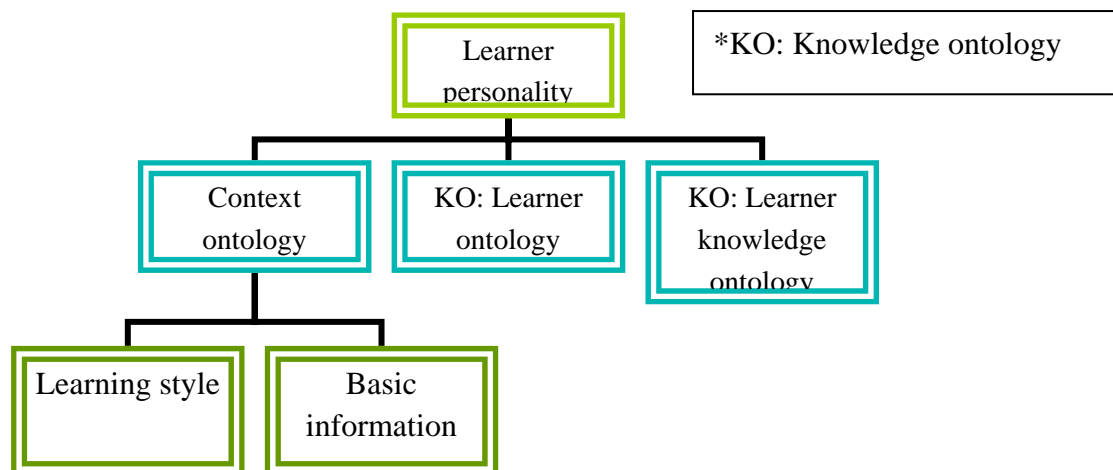


Figure 4-1 Structure of the Learner Personalization

We give definitions and description of these three parts in follow.

4.1 Learning Styles of the Learners

The learning styles information of the learner is the preferences learning styles of the learner and come from the ILS done the register phase(3.1.3 ILS: Index of Learning Styles). This personal data is used to comprehend the learner behaviors during the learning process. I have already discussed in the chapter 2.

The Table 4-1 is the detailed design. For the results of the ILS, they give a result, in which each learning style has a rate, which represents the degree of the learning style. In our system, we define an integer to represent the degree of the learning styles.

The restriction: for each learning style the two types are exclusives: just one type can be not zero. For example, if one learner sensory rate is 5 then his intuitive rate must be zero.

Preferred Learning Styles	L	Type	Value (v)
Perception	l_1	Sensory	0 to 6
	l_2	Intuitive	0 to 6
Input	l_3	Visual	0 to 6
	l_4	Verbal	0 to 6
Organization	l_5	Inductive	0 to 6
	l_6	Deductive	0 to 6
Processing	l_7	Active	0 to 6

	l_8	Reflective	0 to 6
Understanding	l_9	Sequential	0 to 6
	l_{10}	Global	0 to 6

Table 4-1 Concept Details

$$ls_i = \{(l_1, v_1), \dots, (l_{10}, v_{10})\}$$

$$v_j \in \{0, 1, 2, \dots, 6\}$$

$$l_j \in L$$

Where

— ls_i is the learning style vector.

— v_j is the value of the l_j , the type value is an integer from 0 to 6

4.2 Knowledge State of the Learners

The knowledge state concerns about the learner specialty and the knowledge structure. The knowledge structure is the knowledge that the students have learned before. This is organized by the content ontology (Chapter 1). Then we use the learner course ontology to cluster the learners in different groups. Because we know that the learners in the same knowledge background has the high probability to choose a similar content to learn in the same course. In the first version, we just use the trail (vector) to represent the learner knowledge background with the specialty. But in the future, we need to consider the ontology aspect.

4.3 Basic Information of the Learners

The basic information of the learner comes form the inscription information of the learners. They are used to classify the learners in the first time, when the learner first uses the e-learning system.

Element
Name
Sex
Birth date
Specialty
Nationality
Language

Table 4-2 The elements of the basic information

Chapter 5 Trail Design

5.1 Introduction

Problem

The conventional learning trail is the navigational trail. This trail represents the learner navigation activities. The determinant factor to the navigation is the cognitive trail. So how to give the best recommendation is determined by how to best learn about this cognitive trail. To extract the cognitive trail from the navigational trail using the semantic knowledge is very difficult and is easy to miss the original purpose of the learners.

Solution

I design a function to capture the more accurate trail of the learners: the learner ontology construction. This method is to let learner construct his learning ontology himself during the learning process and we capture the data trail of this process.

Advantage

1. For the learner, the trail is for the retrieval of the course.
2. For the system, by analyzing the trails, it can apply three important data.
 - a) For the learner, it can give the personalization.
 - b) For the teacher, it can supply the syntheses data about the learners and the courses.
 - c) For itself, we can use the results of the recommendations to ameliorate the machine-learning train set and the learning rules to make the system smarter.

In this chapter, we will first discuss basic theories of this method and then give the overview design of this method. At last, we will give the definitions of the trails that we capture and use to be analyzed in the Chapter 6. In the future work, we try to automatically extract this trail from the navigational trail and analyze the cognitive trail; this can help not only the course-learning, but also can be applied on the e-commerce and other areas.

5.2 The Basic Theory of the Learning Process: A Cognitive Model

The learning process is the process to connect the new knowledge to the old knowledge process in our cognition. Derry supposes this model in 1990 (Figure 5-1).

1. The learner begins to learn the new information.
2. He finds the prior knowledge in his mind which has the relation with the new information.
3. He tries to connect the new information with the prior information. In fact, this phase is to comprehend the new information.
4. Then he gives the relation definition of the new information and the prior knowledge. This process indicates that the learner has already taken a place for the new information.
5. The second part is the function of the prior knowledge: retrieval and inferred new information from the prior knowledge.

Learning and Remembering Meaningful Information A Cognitive Model

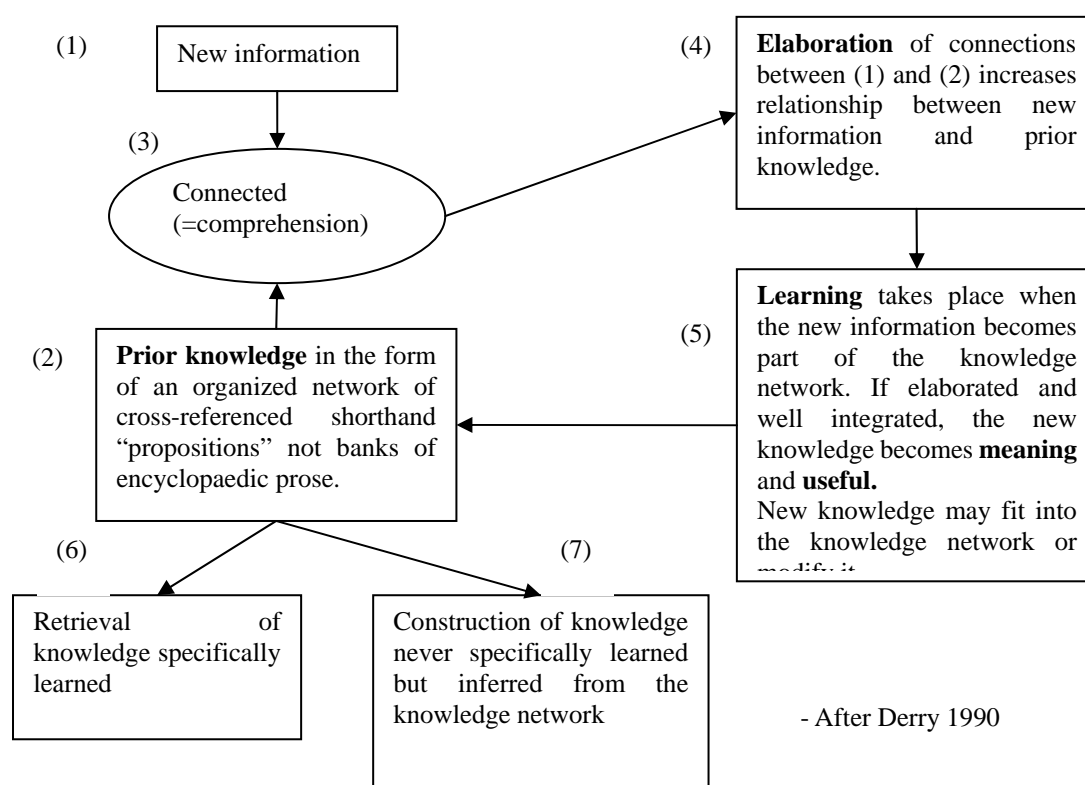


Figure 5-1 Learning and Remembering Meaningful Information: A Cognitive Model

This is the cognitive model for the new information. And for the learner to learn on course, how this model works? In fact, the learning new course is to make a new small network in the mind, and this new small network is the ontology of the course. But this ontology is not the course ontology or domain ontology, because each learner has his own comprehension to the course. We give the analysis in the following part.

5.3 The Activities of Constructing the Ontology by the Students

We simulate the learning process of the learners and design a function, which will make the learner record their learning ontology during the learning process. We use the cognitive model of Derry to simulate the learner learning process.

Compare our method and the theory learning cognitive model in Figure 5-2.

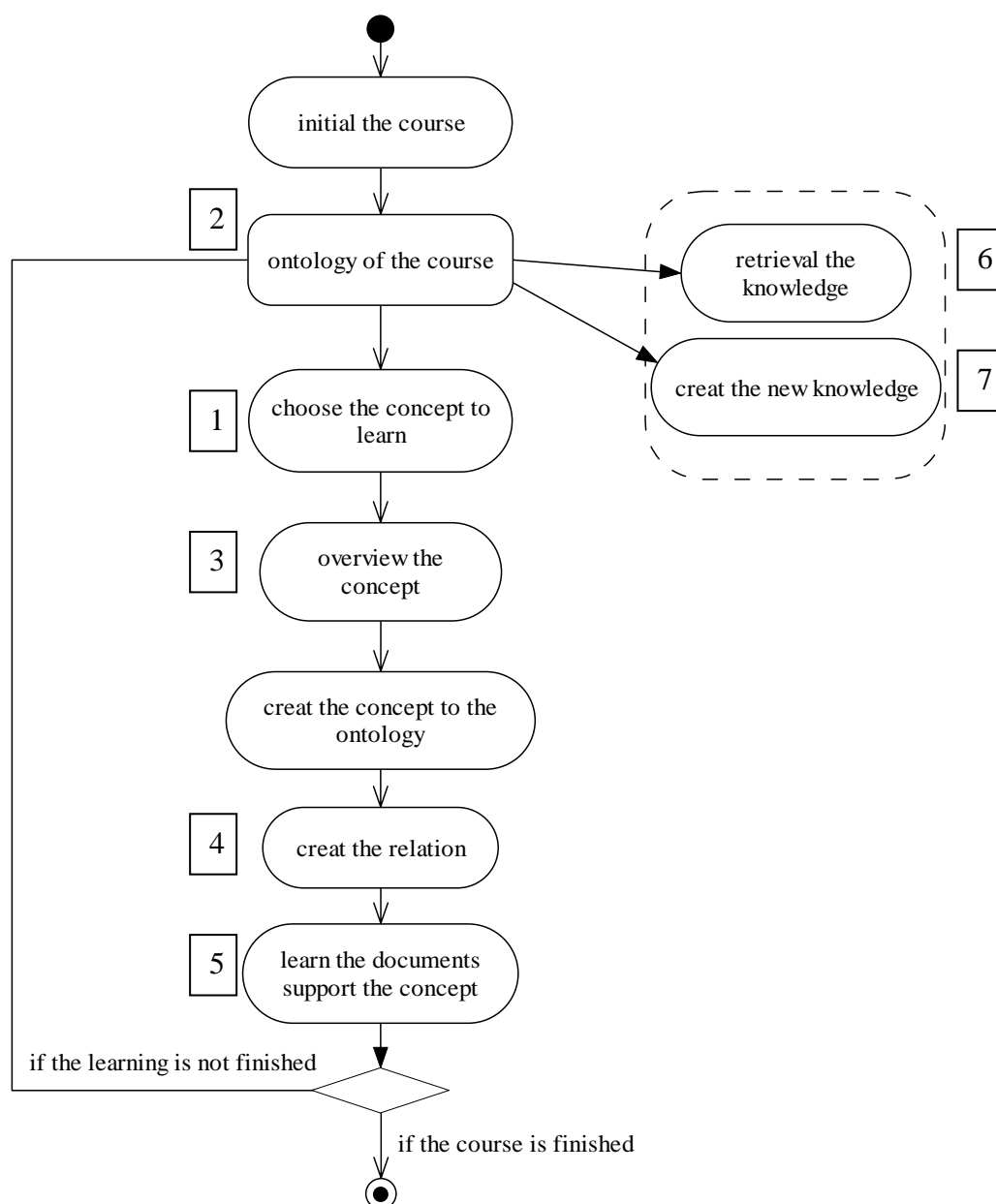


Figure 5-2 The activities of the learners to construct the ontology

- The ontology of the course is the **2** in Derry model.
- The new concept that the learner is now learning is **1**.
- The learner learns the overview of the new concept is the **3**.
- Then the learner adds the relation of the new concept with an old concept is **4**.
- Learning documents of the concept is the **5**.
- The learner can retrieval his course ontology as the note and make up in his mind. Also he can infer some new knowledge or information from this ontology that he has made. This maps the **6** and **7** in Derry's model.

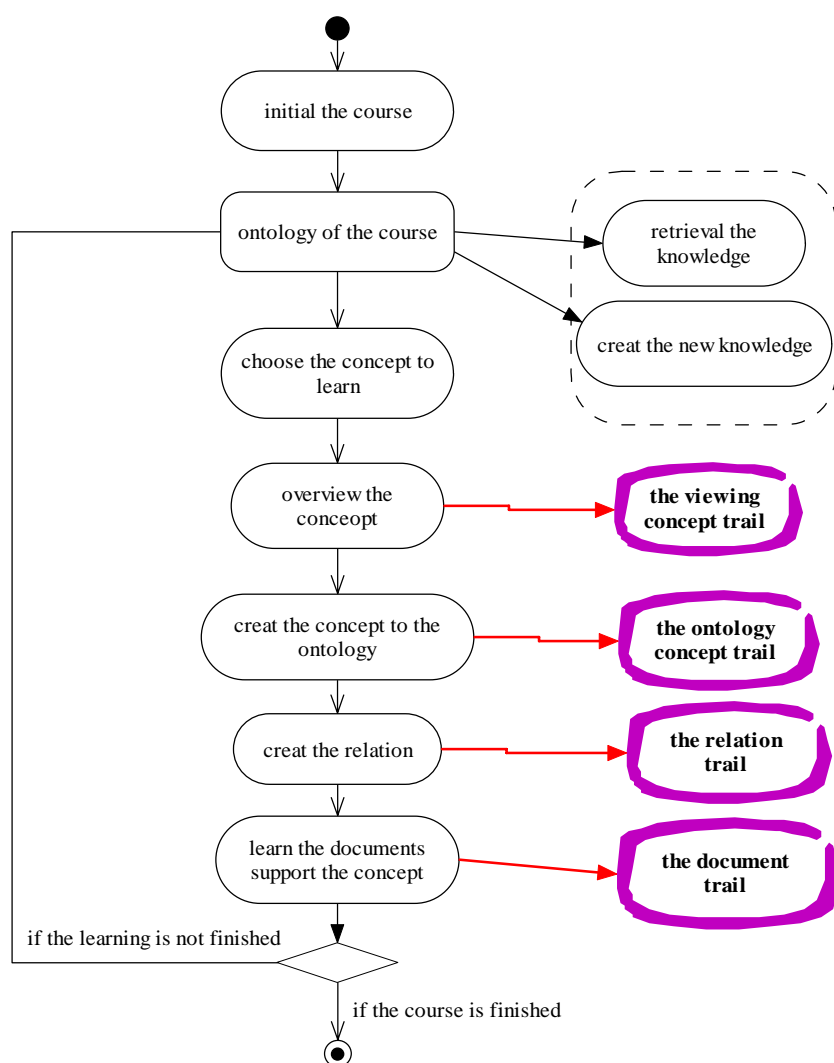


Figure 5-3 The trails versus the activity

5.4 The Trail Design

Before we give the trail definitions, we analyze the learning process. We can capture the navigational trail of the learner, which is composed by the learner activities. And the learner's purpose is to construct the ontology and we capture the trails of constructing the ontology. The ontology trail is extracted from the navigational trail. And then which leads the learner to construct the ontology? The answer is the knowledge state (learner knowledge ontology). The learner wants to construct a whole ontology, and the knowledge state in fact is "pseudo-ontology" of the learner at each learning state. Then at each knowledge state, the goal of constructing an integrated ontology leads the learner to continue to choose the next concept and learn. So the knowledge state of the learner determines the ontology activities of the learners. Then the ontology activity leads the learner navigational trail to accomplish the activity of the ontology.

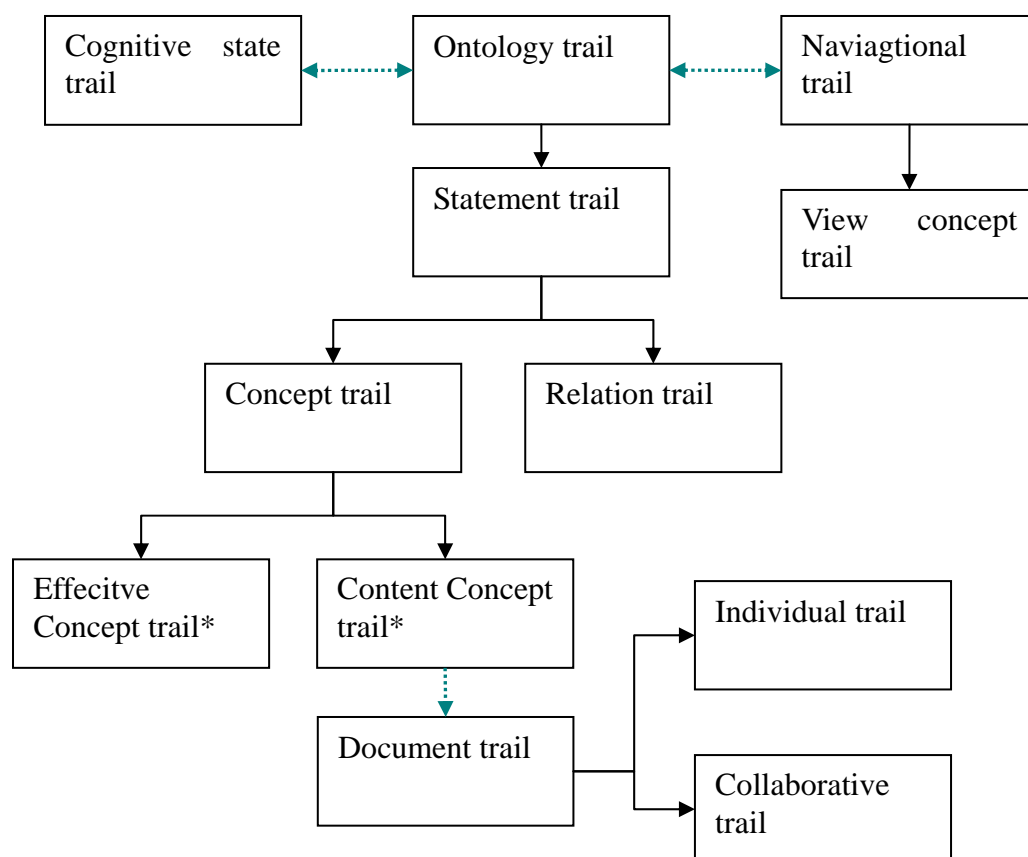


Figure 5-4 Trails design

Legend: *: The trail cannot be captured directly; it is obtained by the data processing process.6.2 Preprocessing)

There are three main branches of the trails: the cognitive state trail, the ontology trail and the navigational trail.

Definitions to these three kinds of trails

Definition 7: Cognitive trail

Cognitive trail presents the learner knowledge state during the learning process. This trail is extracted from the ontology trail (Chapter 6Data Preparation).

Definition 8: Learner ontology trail

The learner ontology is constructed by the learner to represent his knowledge structure. Thus, the learner ontology trail records all the status during constructing the ontology.

Definition 9: Navigational trail

The navigational trail records all the items that the learner visits during the learning process. Its purpose is to construct the ontology trail.

In the learner ontology, there may be some concept that the learner does not learn, but he has already added to their ontology. So we cannot say the learner ontology is just the learner knowledge state. For this problem, in the Chapter 6 we will give the solution to find the real knowledge state of the learner by the learner ontology.

5.4.1 Learner ontology trail

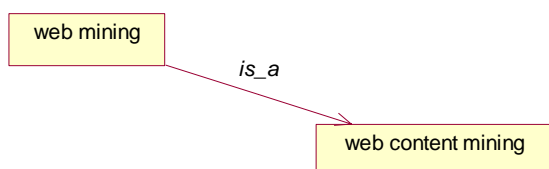
For this trail, we will give an example to better comprehend. The following 4 ontologies construct the ontology trail.

Example 1: the learner constructs the web data mining ontology process.

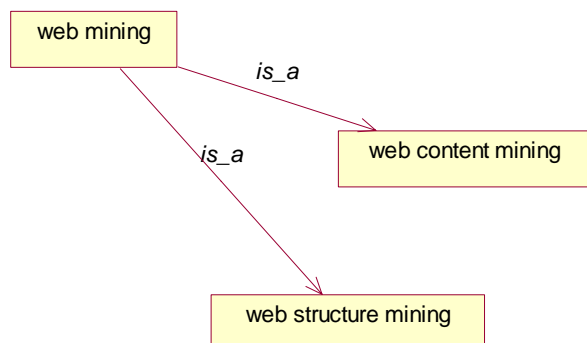
The following ontologies are the learner constructing during the learning process. And from these ontologies, we can extract the useful trails to apply in the mining process.

web mining

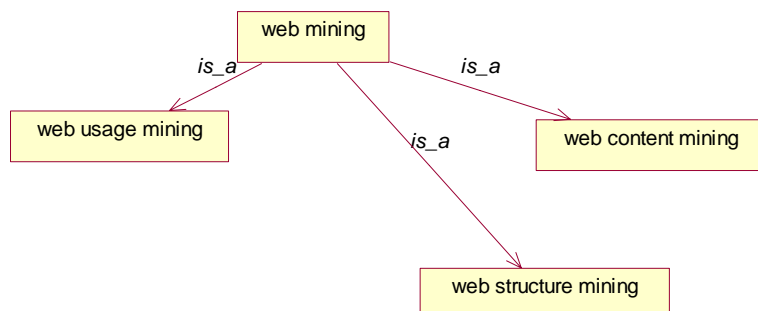
Ontology 5-1



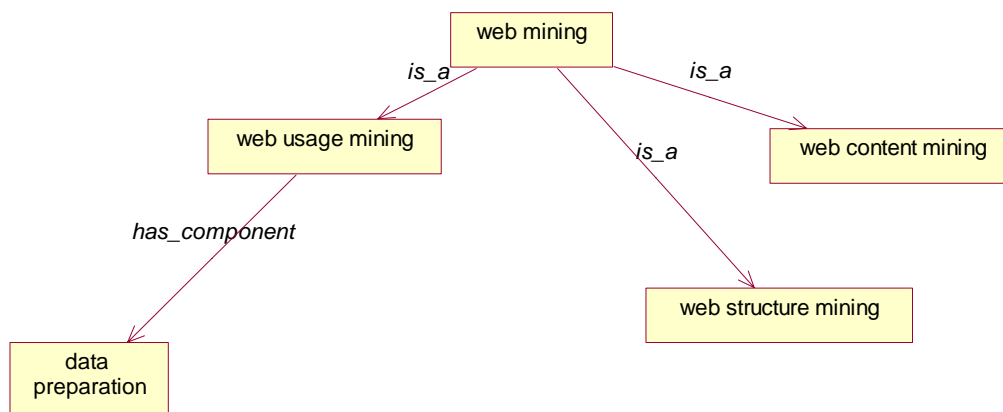
Ontology 5-2



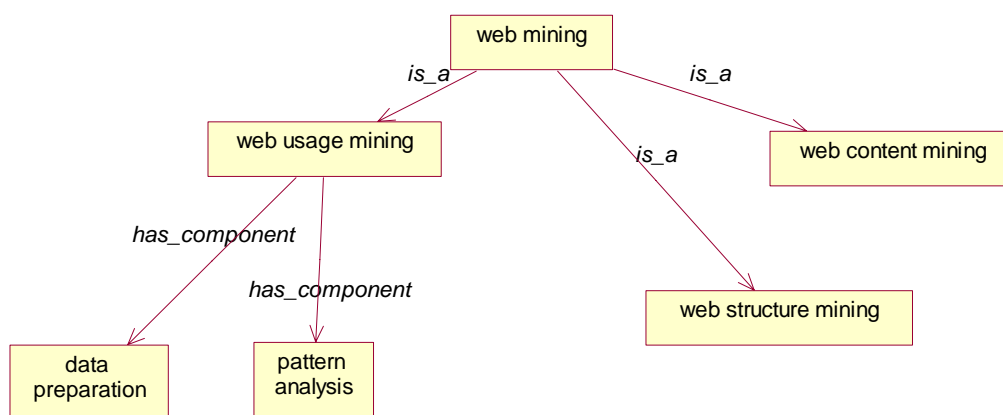
Ontology 5-3



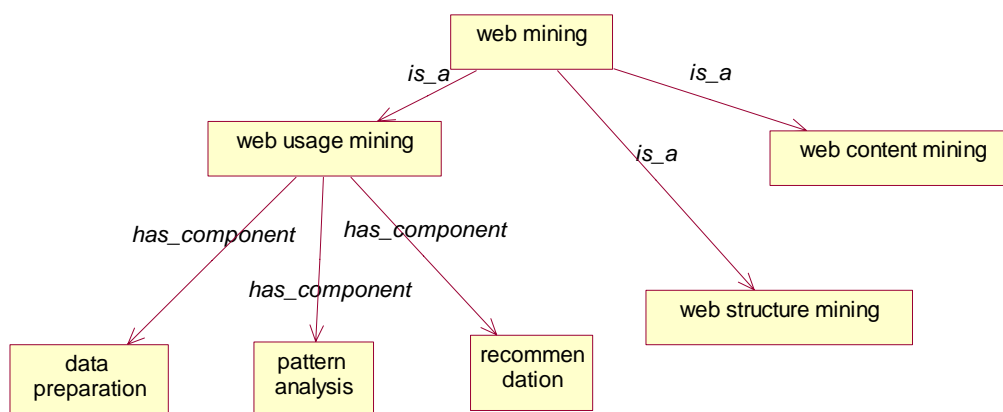
Ontology 5-4



Ontology 5-5



Ontology 5-6



Ontology 5-7

5.4.2 Statement trails

Because the learner constructs an ontology, we can use the RDF statement to present the ontology constructing process. The beginning concept is the subject in the statement, the relation between them is the predicate and the new concept is the object.

Definition 10: Statement

Let S represent the learner add a new concept and a relation to his ontology. We use a triple $S = \langle c_b, r, c_n \rangle$, where:

- c_b is the beginning concept from which the learner adds the new relation with the new concept
- c_n is the new concept which learner adds to his ontology or the concept that the learner is learning or just after learning
- r is the relation between c_b and c_n .

Definition 11: Statement Trail

The statement trail is the sequential set of the statement that records the learner constructing process. $t_s = \{s_1, s_2, \dots, s_n\}$. Where t_s is the statement trail and s_1, s_2, \dots, s_n are the statements that the learner constructs.

This trail affects the learner navigation trail of this topic and also represents the hidden learner motivation during the learning the topic. In Chapter 6 Data Preparation, we will analyze this trail and the learner ontology to give the recommendations.

Example 3: Statement Trail

We use the example 1 to continue. If the learner constructs the ontology of the web data mining like this sequence: web data mining, web content mining, web structure mining and at last the web usage mining. Then we have the statement trail like this.

Index	Statement
1	Create the head concept: web data mining
2	$\langle (\text{web mining}), \text{is_a}, (\text{web content mining}) \rangle$
3	$\langle (\text{web mining}), \text{is_a}, (\text{web structure mining}) \rangle$
4	$\langle (\text{web mining}), \text{is_a}, (\text{web usage mining}) \rangle$
5	$\langle (\text{web usage mining}), \text{has_component}, (\text{data preparation}) \rangle$
6	$\langle (\text{web usage mining}), \text{has_component}, (\text{pattern analysis}) \rangle$
7	$\langle (\text{web usage mining}), \text{has_component}, (\text{recommendation}) \rangle$

From this trail, we can extract some sub-trails. We use these trails to apply the different analysis.

5.4.2.1 The concept trails

The trail of all the concepts that the learners learn and add to the ontology is ordered by the time.

Definition 12: Concept Trail

$$t_c = \{c_1, \dots, c_n\} \text{ where}$$

- t_c is the concept trail that the learner creates by time as learning the course.
- c is the concept that the learner constructs during the learning the course.

With each concept, we give two types of time to each concept: itself time and total time.

- Itself time: comes from all the time spending by the student on its own documents.
- Total time: is the sum of all the sub-concept time and itself time.

Index	Concept	Itself time	Total time
1	Web mining	0	18
2	Web content mining	3h	3h
3	Web structure mining	1h	1h
4	Web usage mining	0	15h
5	Data preparation	3h30m	3h30m
6	Pattern analysis	6h30m	6h30m
7	Recommendation	5h	5h

A. The efficient concept trails

Definition 13: Efficient concept trail

The efficient concept trail contains the concepts, which the learners learn, or the sub-concepts of this concept that the learner effectively learn. Thus it excludes the concepts which are just created by the learners but not learn indeed. The algorithms about how to extract this trail will be presented in the chapter 6.

$$t_{ec} = \{c_1, \dots, c_m\}$$

Where

- c_i is the concept which has been efficiently learned by the learner.

B. The content concept trails

Definition 14: Content concept trail

The content concept trail is the trail extracted from the concept trail, which just contains the content concepts. The algorithms about how to extract this trail will be presented in the chapter 6.

$$t_{cc} = \{c_1, \dots, c_m\} \text{ where } c_i \in C_c$$

5.4.2.2 The relation trails

Definition 15: Relation trail

This trail is composed by all the relations, which the learner creates during constructing the ontology.

$$t_r = \{r_1, \dots, r_n\}$$

Where

- r_i is the relation between two concepts

The relation trail is extracted from the statement trail and we try to use this to find the rules that the learners choose the concepts to learn.

5.4.3 The navigation trail

The navigational trail (Definition 9: Navigational trail) records all the activities of the learner. This trail can help us find the relationship between the ontology trail and the navigational trail. With this comparison, we can find some rules to help us realize extract the ontology trial directly from the navigational trail.

5.4.4 The document trail

Definition 16 : Document trail

The document trail represents the documents, which the learner learns during the learning process and is ordered by time. There are two kinds of the document trail: the collaborative trial and the individual trial.

The next important trail is the document trail. There is a problem in the trail domain whether we should combine the transaction trail and the document together. First, the document can be added a meta ontology, which is projected to the domain ontology or the course ontology. So like this, we can analyze the document trail to find the patterns and give the recommendations to the learners. But, the target of learner reads the documents is to learn the concept, which is supported by these documents. So if we separate the document from the learner ontology, then the analysis will miss some interesting information. So we combine the documents to the concepts trails (statement trails).

I design two kinds of document trails: the individual trail and the collaborative trail.

5.4.3.1 The Collaborative Document Trail

Definition 17 : Collaborative document trail

The collaborative document trail is the trails of all the documents which learner learns during the entire course learning process.

$$t_d = \{d_1, \dots, d_n\}$$

Where

- d_i is the document that the learner views.

We record two main metadata for this trail: the subject concept and the time.

- The subject concept: is the document topic corresponding to which concept in the domain ontology.
- The time: the learner learns the document time.

Example

$$t_d = \{d_1, d_2, d_3, d_4, d_5\}$$

Index	document	Subject concept	Time
d ₁	Doc 1	Web content mining	3h
d ₂	Doc2	Web structure mining	1h
d ₃	Doc 3	Data preparation	3h30m
d ₄	Doc 4	Pattern analysis	2h50m
d ₅	Doc 5	Pattern analysis	3h20m

Table 5-1 Example of the document trail

5.4.3.2 The Individual Document Trail

Definition 18 : Individual document trail

$t_{dci} = \{d_1, \dots, d_l\}$ is the document trail for the concept c_i that the learner follows.

Where

- d_i is the document, which the learner learns.

So I call it concept doctrail. To each concept, we have the document set of each concept $D_{ci} = \{d_1, \dots, d_m\}$. That is to say that each concept contains the documents set D_{ci} . For the whole ontology, we have the $D = \{d_1, \dots, d_n\}$, which is the set of all the documents in the ontology (for this course). But in the practice, the documents belong to one concept is very rare. Sometimes the teacher, they assign just one document to one concept, so our trail t_{dci} is not very practical. Then we capture another doc trail $T_d = \{t_{dc1}, \dots, t_{dck}\}$ called whole doctrail. This trail captures all the concepts that the learner learns and ordered by the time. Then we have:

$t_{dci} = \{d_1, \dots, d_l\}$: Individual document trail

$T_d = \{t_{dc1}, \dots, t_{dck}\}$: whole doctrail

$D_{ci} = \{d_1, \dots, d_m\}$: set of documents supporting the concept c_i

$D = \{d_1, \dots, d_n\}$: set of documents supporting in the content course ontology

where

$$t_{dci} \subset D_{ci} \quad T_d \subset D$$

$$t_{dci} \subset T_d \quad D_{ci} \subset D$$

In Chapter 6, we will analyze this trail using association rules and the sequential pattern, and we will use the pedagogical theories to analyze this trail to find the learner learning preference styles.

Example

$$t_{dci} = \{d_4, d_5\} \text{ where } c_i = (\text{pattern discovery})$$

Part II Analysis of the trails and the recommendation

In this part, we now begin to analyze the learner trails using web mining techniques. The main idea is to use the data mining techniques to find the aggregate profile of the course, document and the learner, then according to the aggregate profile to generate the recommendations. There are three main components in the analysis phase: the data preparation, the pattern discovery and the recommendation component.

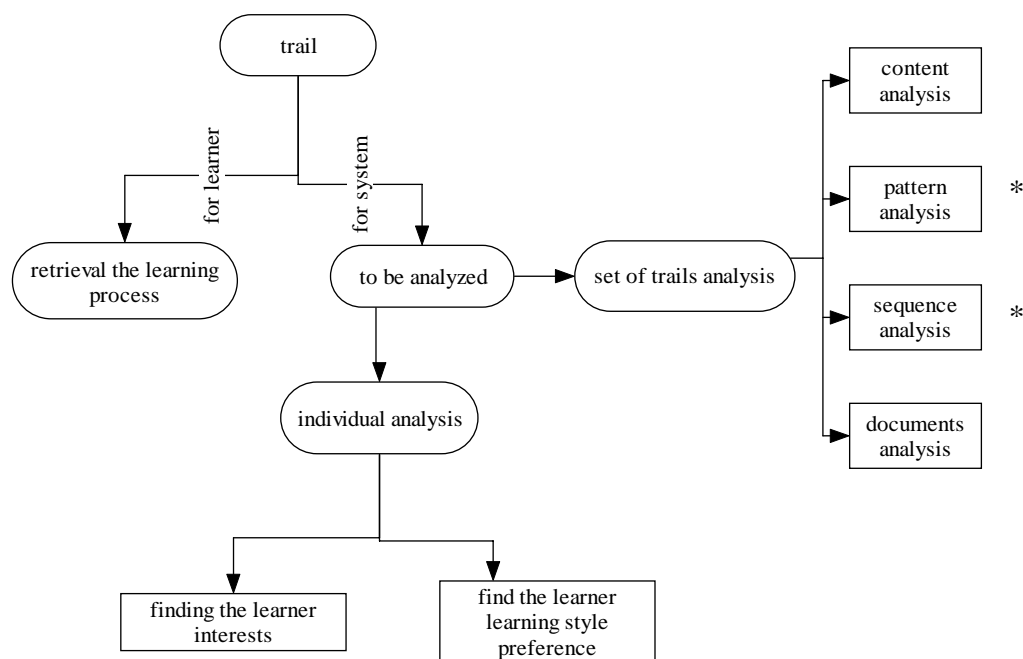


Figure 5-1 The Application of the Trails

Legend: *: future work

For each of these three steps, we propose algorithms..

1. Semantic data mining. This is important to us, because we know that the hidden trail for the learner is the ontology. So we analyze the trails of the learners, in fact mainly analyze the ontologies trails. So we need to consider the association rules and clustering with the semantic, because of the semantic trails. We cannot think all the concepts without hierarchy and relations. For us, we have three tasks:
 - 1.1 Consider the concept hierarchy in association rules and clustering mining.
 - 1.2 Hierarchy mining: in this part, we need to mine the hierarchy pattern that the learner visits. For example, one learner visits the first level of the ontology and then the second and then the third with breadth. But another one perhaps visits the ontology with depth, the first level and the second level and the third, and then returns to the second and then return to the first.

2. Domain level aggregation and representation: for the learner ontology, we will find there is not the algorithm to compare them and give the similarities of the ontologies. But we cluster the learners and give the recommendations.
3. Ontology based recommendation: many systems give the recommendation basing the collaborative filters. But they miss two important things: the first the content filter misses the ontology structure, and the user filter misses the analysis of the learning styles of the learners.

Chapter 6 Data Preparation

Problem

The resource trails are the learner ontology trails and the navigational trails. The learner ontology trail will be mainly discussed in this document. But there are some problems. We have already stated in the trail design that the learner ontology is not exactly the knowledge state ontology trail. It needs to clean the disturbed concepts to make it more accurate to consider as the knowledge state ontology. The following three reasons are why they are not be equal.

1. We can consider the learner learning process as a navigation of the learner ontology, but the problem is that during the learning process, the learner may meet some concepts that he cannot give the definition or he cannot add this to the learner ontology.
2. There are some concepts that the learner adds to the ontology, but he doesn't interest in this after, so he doesn't learn about this concept.
3. There are some concepts in the ontology, which the learner wants to learn but just opens some documents without reading indeed.

Solution

1. For the first problem, although the learner doesn't know the relation, but we know the domain ontology. So we can use the domain ontology to find the relation. The mapping process is designed to avoid the learner make the mistakes or cannot give the relation right like this.
2. For the second the problem, we define four types of concepts, like this we can find the content concept, which the learner effectively learns.
3. For the third problem, we consider add a metadata weight to each concept. This weight represents the degree of the learner pay in this concept. Thereby, this can be considered as a scale to filter the noisy concept.

In this part, we clean the ontology data of the learner trails and prepare data for the analyzing in data mining. The Figure 6-1 represents main design of the data preparation phase. The input resource is the data trails that we get from the learners. Then after the mapping, preprocessing and the support filter phases, these trails are stored in the learner trail database to prepare for the next component-pattern analysis.

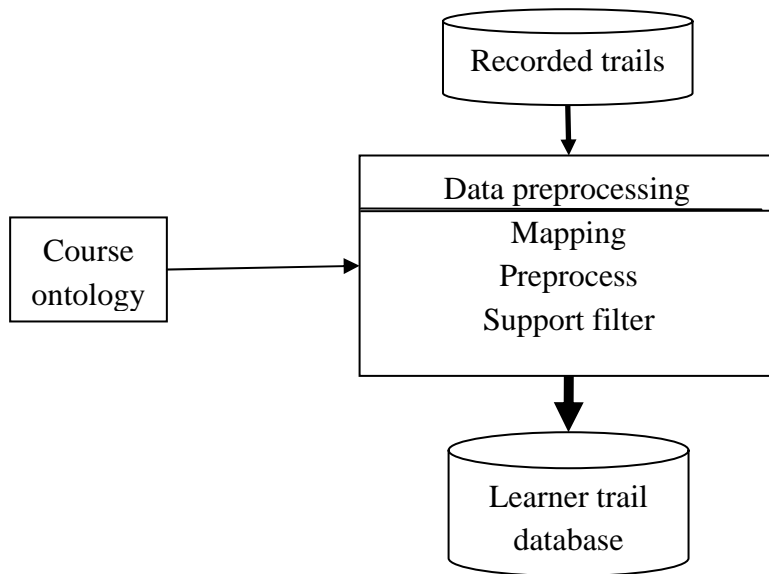


Figure 6-1 Data Preparation Component

Before explain the data preparation component, we use an example to explain the phase and to explain our hypothesis.

Example: Hypothesis

Resource 1. domain ontology

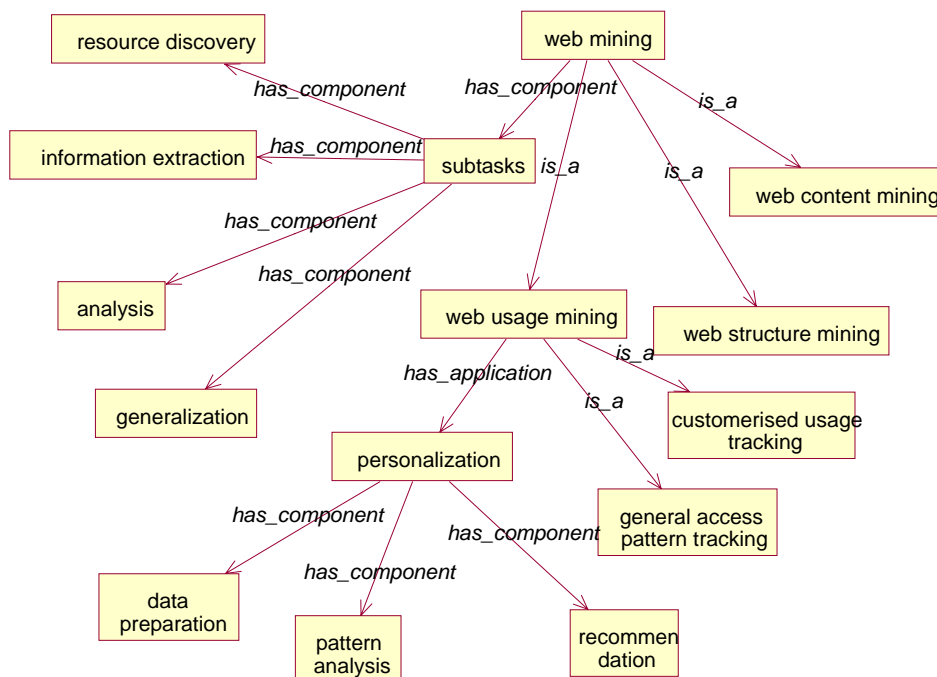


Figure 6-2 Domain ontology

Resource 2. Learner ontology

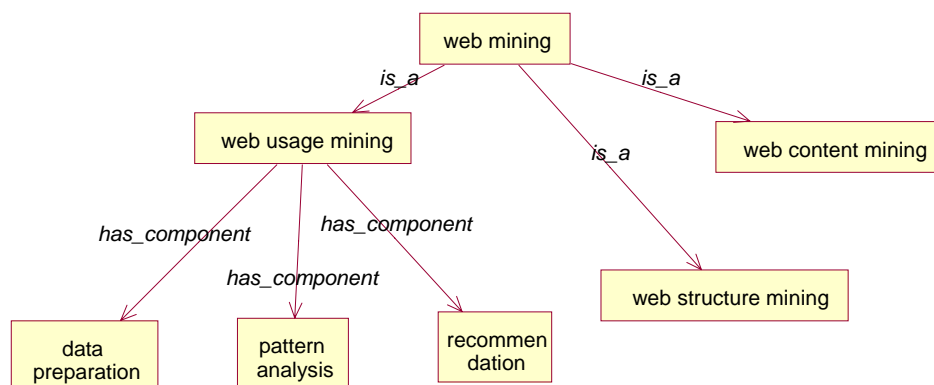


Figure 6-3 Learner ontology

Resource 3. statement trail

Index	Statement
1	Create the head concept: web data mining
2	<(web mining), is_a, (web content mining)>
3	<(web mining), is_a, (web structure mining)>
4	<(web mining), is_a, (web usage mining)>

Index	Statement
5	<(web usage mining), has_component, (data preparation)>
6	<(web usage mining), has_component, (pattern analysis)>
7	<(web usage mining), has_component, (recommendation)>

Resource 4. concept trail

Index	Concept	Itself time	Total time
c ₁	Web mining	0	18
c ₂	Web content mining	3h	3h
c ₃	Web structure mining	1h	10m
c ₄	Web usage mining	0	15h
c ₅	Data preparation	3h30m	3h30m
c ₆	Pattern analysis	6h30m	6h30m

Resource 5. document trail

Index	document	Subject concept	Time
d ₁	Doc 1	Web content mining	3h
d ₂	Doc2	Web structure mining	10m
d ₃	Doc 3	Data preparation	3h30m
d ₄	Doc 4	Pattern analysis	2h50m
d ₅	Doc 5	Pattern analysis	3h20m

6.1 Mapping the Individual Ontology

The mapping process is to map the learner ontology to the domain ontology. By the mapping process, we can correct the learner mistakes in the learning process. For the mapping, we have two main steps.

1. We map the concepts of the individual ontology to the domain ontology. That is to find the concepts of the learner ontology in the domain ontology.
2. We complete the path of the concepts in the mapping ontology.

Example: mapping the individual ontology

Data preparation phase 1: Mapping

Step 1: Map the concepts in the learner ontology to the domain ontology.

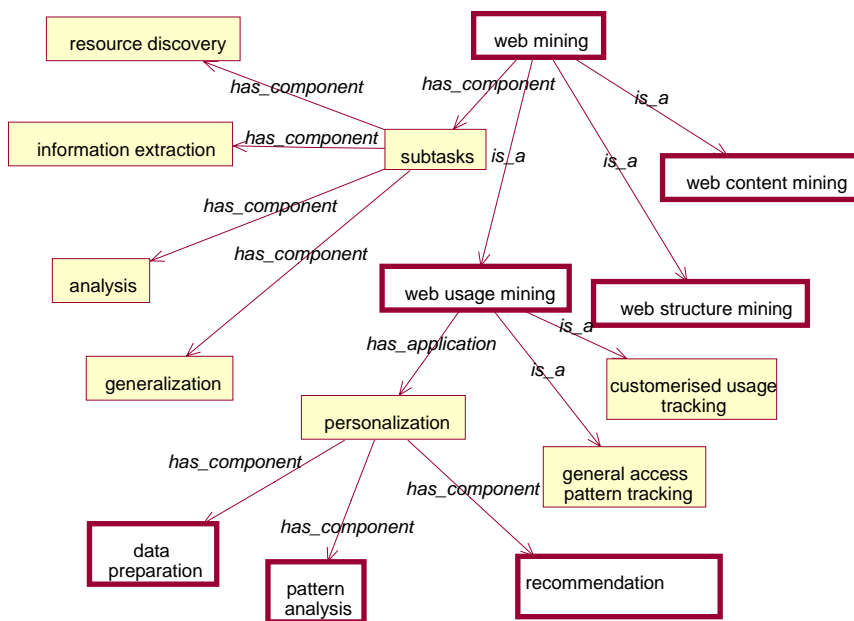


Figure 6-4 Mapping the learner ontology to the domain ontology

Legend: The gross concept represents a concept that happen in the learner ontology.

Step 2: complete this ontology that we add by completing the path of the red concepts in the domain ontology.

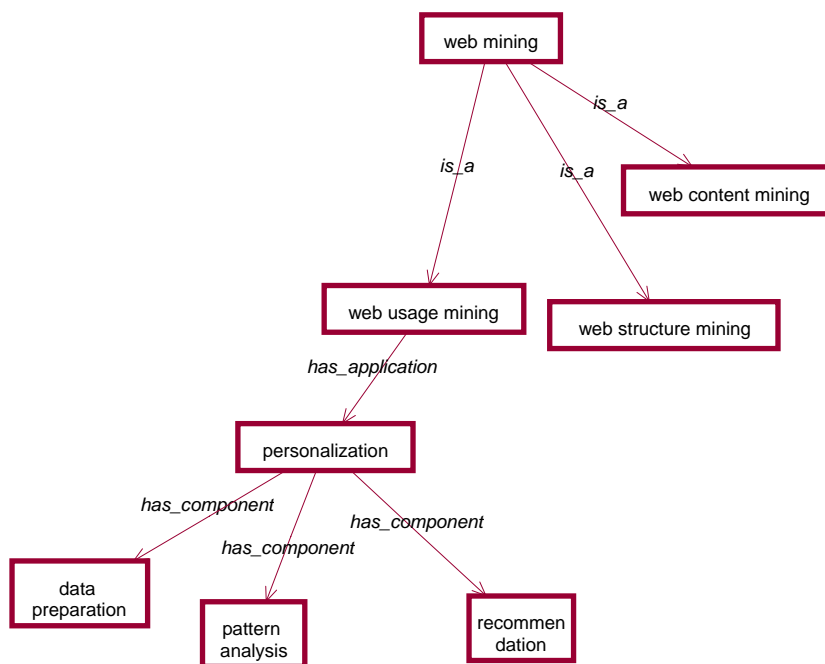


Figure 6-5 The mapped learner ontology

6.2 Preprocessing

6.2.1 Assign the type to the learner ontology concepts

In order to analyze the learner ontology, we separate the concepts into four types:

Definition 19: Type of the concepts

1. Head concept (Ch): a concept is the root of the learner ontology and the topic of the course.
2. Mi-concept (Cm): a concept contains the sub-concepts and some documents are connected to it.
3. Auxiliary concept (Ca): a concept contains the sub-concepts but no document is connected to it.
4. Content concept (Cc): a concept just contains the documents but no sub-concepts.

In fact, we use the nodes, the links and the documents connected to classify the concepts. But we need to pay attention that the mi concepts contains the sub-concepts, so their documents sometimes contain just a little information like the definition, history or introduction. And for the content concepts, they are always in the leaves of the ontology. In contrary, the auxiliary concepts are always in the branches of the ontology. The following Table 6-1 explains the usage characteristics of these concepts.

Concept type	Content, link characteristics	Usage characteristics
Head concept	Root of the ontology	First concept in the trail
Mi concept	Documents, sub-concepts	Low doc_weight but high sub_weight, short time spend on doc and itself also, high frequent
Auxiliary concept	Sub-concepts	High sub_weight, short time spend, high frequent
Content concept	Documents	High doc weight, long time spend, low frequent

Table 6-1 Characteristics of the four type of concepts

There are two main steps to assign the type to the concepts.

1. We use the definition of the four types to assign the concepts.
2. There is one kind of concept not belonging to the four type. This concept contains neither sub-concept nor the document. This kind of concept is meaningless to us. Thereby we delete it in this step.
3. Generate the content concept trail.

Example: preprocessing

Step 1: assign the types

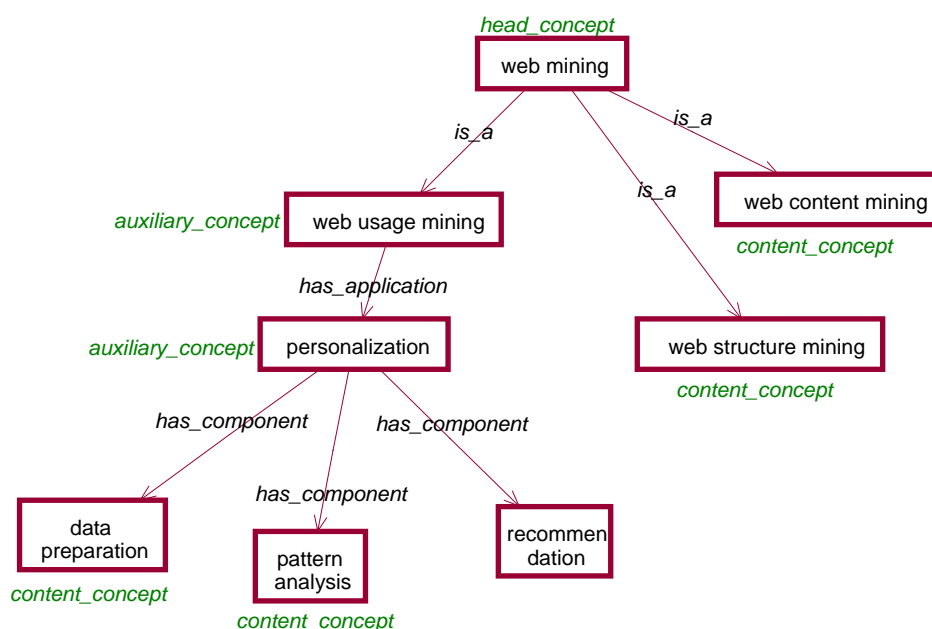


Figure 6-6 The mapping ontology after add the type of the concepts

Step 2: Delete no meaningful concept

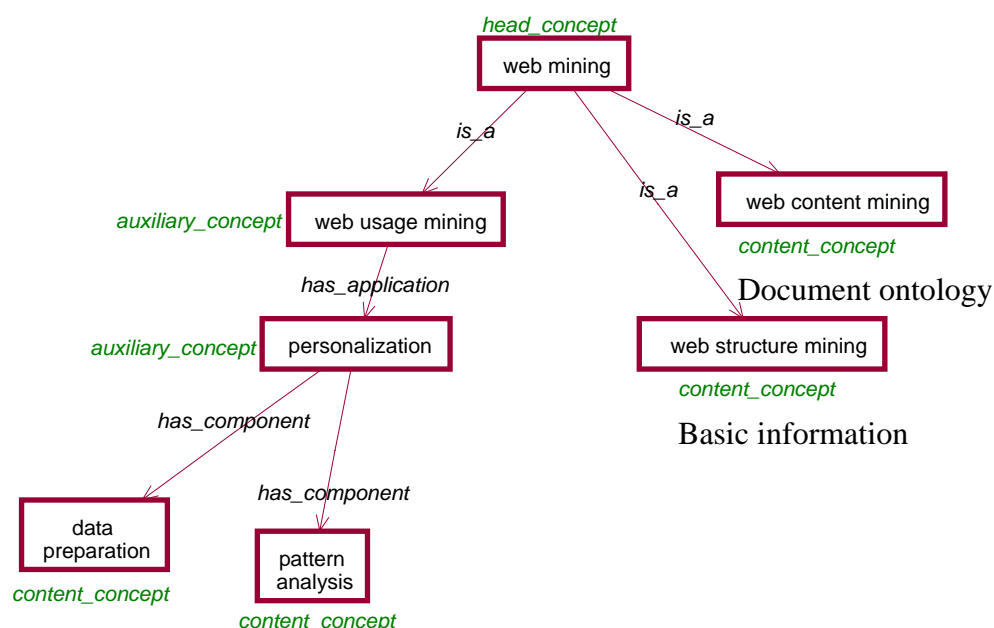


Figure 6-7 After pruning

Step 3 : generate the content trail

$$t_{cc} = \{c_2, c_3, c_5, c_6\}$$

Index	Concept	Itself time	Total time
c ₂	Web content mining	3h	3h
c ₃	Web structure mining	1h	10m
c ₅	Data preparation	3h30m	3h30m
c ₆	Pattern analysis	6h30m	6h30m

Table 6-2 Content trail concepts

6.2.2 Assign the weight

The individual concept weight indicates how the learner takes attention to this concepts or whether the learner effectively learn about this concept. We have the learner ontology and the trail of the learner. We will notice that some algorithm like the association rules just use the frequent times to generate the candidature, but in the individual trail, the content concepts just are visited by the learner one time. So we use the association rules directly on the trails, we will find that some content concepts will be pruned, because the top auxiliary concepts and the

mi-concepts appears so many times in the trails. So we lose much useful information. So we must add a meta weight to the concept to ensure that the important concepts like the content concept are not to be pruned. The weight can represent the time and the depth of the concept that the learner learns.

1. We assign the weight to the document to represent the efforts of the learner to learn this document.
2. The weight of the concepts represents the degree of the learner to learn this concept.
3. For the weight of the ontology, it is to stand for which degree of the learner learns this topic.

Segment the learner navigational trail

There are many methods to segment the log trails of the users. But all these methods miss the ontology structure of these concepts and are not easy to control. So in our system, we make the learner to construct his ontology and then we use this to segment the trails. When the learner finish one concept in the ontology, we have a record of the pages and the documents that he visits for constructing this concept. Thus when the learner constructs the ontology, he has already divided the trails into small segments himself. Each concept can represent one segment, so this is easy to analyze the structure of the learner trails.

6.2.2.1 The document trail

We have the document d_i and our purpose is to use the time as basic scale to assign the weight.

We propose three kinds of weight which are assigned at two different moment: before and after.

A. Pre-processing

Binary weight: If the learner views this document, we assign the weight of this document to 1; if not, we assign it to 0. But, the learner perhaps just reads the document a little time and doesn't interest at all. So we cannot just use this binary weight to measure the learner's efforts.

$$w_{di} = \begin{cases} 1 \\ 0 \end{cases}$$

So we need to use the learner learning time as the scale to assess the learner efforts.

Time weight: Then now the weight is the time duration weight.

$$w_{di} = \text{time} \quad \text{where } d_i \text{ is the document } i$$

But this weight also has a problem. Due to the document length, difficulties, and other factors to

affect the learning time, how we know that whether the learner effectively learn this document? For this reason, we realize the post processing.

B. Post-processing

We use the significance testing firstly to filter the concepts and then use the normalization give the formal weight to the document.

b.1 Significance testing

We use the statistical measure such as significance testing and normalization to systematically analyze the document weight. If there is not the mistake, the weight distribution should be a normal distribution (Gaussian distribution).

Statistical significance testing can help us to find the document in the trail, which is not learned by the learner. The weight of one document in a document trail should be 0, if the amount of time spent on this document is significantly below the mean time duration of this document across all learner trails.

$$score(w_{d_i}) = \frac{w_{d_i} - mean(d_i)}{\sigma(d_i)}$$

Where

- The threshold θ is the significance threshold.
- $mean(d_i)$ is the mean time duration of the document d_i .
- The $\sigma(d_i)$ is the standard deviation for the concept d_i across all the learner trails document trail Td_{c_i} that contain the d_i .

Therefore, if the $score(w_{d_i}) \leq \theta$, it indicates that the duration time for this document is too short to learn. This expresses the results that the learner does not learn it in fact. Thus, we need to filter these document from the learner document trails.

b.2 Normalization of the Document weight

We use the normalization of the document weight across all the learners' trails to give the new and formal definition of the weight. This normalization is across the learner document trail. Thus, this method can make us to compare the different documents using this weight.

1. During the significance test step, we find the low weight of the doc and filter them. But in fact, there are also some document with the abnormal high weight. For example, if one learner spends 2 hours on one document, while the other learners just 30 minutes. This is abnormal and we can infer that this document is difficult to comprehend for this learner. If $score(w_{di}) \geq (-\theta)$, that is to say that the weight is abnormal, so we reduce the abnormal weight to the max normal weight. Then we assign the $score(w_{di}) = (-\theta)$ if $score(w_{di}) \geq (-\theta)$.

If the $score(w_{di}) = (-\theta)$ then we assign all the w_{di} meets $score(w_{di}) \geq (-\theta)$ to $w_{di} = x$.

2. Now we apply the normalization to the w_{di}

$$norm_d(T_d) = \{v_{di}^t \mid 1 \leq i \leq n\}$$

where

$$v_{di}^t = \frac{w_{di} - \min_{i \in T} \{w_{di}\}}{\max_{i \in T} \{w_{di}\} - \min_{i \in T} \{w_{di}\}}.$$

3. We give the new weight definition. $w_{d_i}^{formal} = v_{di}^t$ We change the weight is to make the value from 0 to 1 and then this weight can be easy to compare across the docs.

The formal weight definition

$$w_{d_i}^{formal} = v_{di}^t$$

Now we have three kinds of weights: the binary weight, the time weight and the normal weight.

So now the new weight is from 0 to 1. Then we can say that the learner attention is from 0 to 1. But when we give the recommendation, we should notice that when the $w=0.5$ around, the learner actually learn the document but not very long time. We cannot say that they just pay half efforts. So how we deal with this problem? We will discuss in the follow.

1. We use this to the significant filter to analyze the common weight of the document in order to find the most popular document in the document list. Inversely, from the weights that are filtered, we can find which documents waste the learner time. Because the high click rate,

but the low effectively learn time.

2. We can use this to calculate the relative weight of the document and to find the learners' interests and calculate the concept weight.

So the trail of the document becomes $t_{dci} = \{(d_1, w_{d1}), (d_2, w_{d2}), \dots, (d_l, w_{dl})\}$.

b.3 Normalization the weight across the ontology

Now we have already filtered the document with the low weight. But the weight of the document is the time and this cannot represent the learner efforts. Why? If the learner learns this doc using 30 minutes, but this doc is so long that other ones use 45 minutes. But we cannot filter this doc, because the time is long enough to prove that the learner has already learned. Then there are three possibilities. Firstly, the learner doesn't learner careful, not interesting enough. Or the learner learns is very quickly. This case is too difficult to analyze and predict. At last, the learner reads all the docs with careless. We use the normalization of the weight of the document to find the two first possibilities. For the third possibility, we have no method to find. But this not important, this is too rare to happen. Thus, we do not study the third case.

Then we use the normalization to deal with the above problems. This normalization is across the weight of the documents in the single document trail of one concept. We call this wholedoctrail normalization. For the whole doctrail, we simplify the trail using the weight to represent the doc, then $T_d = \{w_{d1}, \dots, w_{dn}\}$. Now we use the normalization to the doctrail.

$$norm_t(T_d) = \{v_{di}^t | 1 \leq i \leq n\}$$

Where

$$v_{di}^t = \frac{w_{di} - \min_{1 \leq j \leq n} \{w_{dj}\}}{\max_{1 \leq j \leq n} \{w_{dj}\} - \min_{1 \leq j \leq n} \{w_{dj}\}}.$$

Then the trail becomes $T_d = \{v_{d1}, \dots, v_{dn}\}$. Use this trail to find the learner interests and preferences doc in the learning process.

6.2.2.2 The concept trail

A. The content concept trail

We use the method like the document trail analysis to analyze the concepts trails. Because we have already said that the concept in the course ontology doesn't contain so many documents

indeed. Normally it contains just one document. So for the content concept, we use the time as the raw weight of the concept and then use the methods above significance testing and two kinds of normalization to analyze the concept weight. The w_{c_i} is the weight of the concept c_i .

The definition of the time weight

$$w_{c_i}^{time} = \text{time}.$$

a.2 Significance testing

The statistical significance testing can help us find the concept which the learner not indeed learn. The duration time of these concepts often very little. The method is in the following.

$$\text{score}(w_{c_i}^{time}) = \frac{w_{c_i} - \text{mean}(c_i)}{\sigma(c_i)}$$

Where

- The threshold θ is the significance threshold.
- $\text{mean}(c_i)$ is the mean time duration of the concept c_i .
- The $\sigma(c_i)$ is the standard deviation for the concept c_i across all the learner trails that contain the c_i .

If $\text{score}(w_{c_i}^{time}) \leq \theta$, we filter the concept c_i in the learner ontology.

a.2 Formal weight: Normalization the weight across all the trails

1. During the significance testing, we can find the abnormal concepts from the other side. If $\text{score}(w_{c_i}^{time}) \geq (-\theta)$, that is to say that the weight is abnormal, because it is too heavy. So we reduce the abnormal weight to the max normal weight.
2. Then we assign the $\text{score}(w_{c_i}^{time}) = (-\theta)$, if $\text{score}(w_{c_i}^{time}) \geq (-\theta)$.

If $\begin{cases} \text{when } \text{score}(w_{c_i}^{time}) = (-\theta) \\ \text{then } w_{c_i}^{time} = x \end{cases}$ then we assign all the $w_{c_i}^{time}$ meets $\text{score}(w_{c_i}^{time}) \geq (-\theta)$ to

$$w_{c_i}^{time} = x.$$

3. We apply the normalization to w_{c_i}

$$norm_c(Tcc) = \{v_{ci}^t | 1 \leq i \leq n\}$$

Where

$$v_{ci}^t = \frac{w_{c_i}^{time} - \min_{t \in T} \{w_{c_i}^{time}\}}{\max_{t \in T} \{w_{c_i}^{time}\} - \min_{t \in T} \{w_{c_i}^{time}\}}.$$

Now we give the new weight definition. We change the weight to make the value from 0 to 1 and then this weight can be easy to compare across the different concepts.

$$w_{c_i}^{formal} = v_{ci}^t$$

Thus the content trail changes into a vector. $t_{cc} = \{(c_1, w_{c_1}^{formal}), \dots, (c_n, w_{c_n}^{formal})\}$

B. The mi-concept

For the mi-concepts, the situation is a little more complex, because it contains the docs and the sub-concepts at the same time. But the docs often contain just the introduction, history, or something like this. So we can think this as one of the sub-concept of the mi-concept. So firstly, we treat the mi-concept as the content concept to calculate the weight of the docs. And secondly, we think the doc weight as one sub-concept and use the method of the auxiliary concept to calculate the weight. Then

$$w_{c_j}^{formal} = \frac{1}{n+1} \left(\sum_{i=1}^m w_i^{formal} + w_{dc} \right)$$

Where

- c_i is the sub concept of the c_j
- m is the number of the sub concepts that the learner learn of the concept j
- n is the number of the sub concepts which belong to the concept j in the domain ontology.
- w_{dc} is the weight of the document which subject is the concept j

C. The auxiliary concept

For the auxiliary concepts, the weight represents the importance that the learners consider. So we use the sub-concepts to calculate the weight of these concepts.

If the concept contains n sub-concepts and m concepts are learned by the learners. Then

$$w_{c_j}^{formal} = \frac{1}{n} \sum_{i=1}^m w_{c_i}^{formal}$$

Where

- c_i is the sub concept of the c_j
- m is the number of the sub concepts that the learner learn of the concept j
- n is the number of the sub concepts which belong to the concept j in the domain ontology.

D. The head concept

The head concept can be treated to the two types of concepts above: the mi-concept or the auxiliary concept. And then we can use the method above to calculate. But this concept is so important because it can present the whole ontology weight. So we consider it as separate.

Conclusion

The concept trace of the learner becomes as follow.

$$t_{oi} = \{(c_1, w_{c_1}), (c_2, w_{c_2}) \dots (c_n, w_{c_n})\}$$

Using the significant testing and the normalization, we prune the abnormal data and then we solve the problem of the tree. If we don't apply the normalization, we will find that for the head node in the ontology, the weight is too small and we emphasize the leaves and miss the branches. So we use the normalization method to compare the different weights.

Using this method to calculate the weight, implies two problems consider that we miss first problem is important differences between the concepts. For example, the sub-concept A is more important than the sub-concept B, but we think them equally and just divide. The second problem is that, this method for the content concepts, it needs enough set to use the significance testing and normalization. So how to deal with these two problems?

For the first problem, we can use the probabilistic to solve this problem. If the more learners

visit the more probabilistic of these concepts and then we normalize the probabilistic and we can get the importance differences between the concepts.

6.3 Support Filter

After adding the weights to the ontology, we use these weights to prune the individual ontology in order to provide the efficient information. We prune the concepts that the learners do not effectively learned. It is to say that we prune the concepts with a low weight in the learner ontology. We use the significance testing to prune the individual ontology.

The algorithm 1 consists to search the leaf concepts of the ontology, and we give a threshold of the weight. If the weight of the concept is smaller than this threshold, we delete this concept from the ontology. We repeat this function until there are not the leaf concepts that are smaller than the threshold.

```

prune(o)
  generate the leaf concept set  $C_L$  from  $o$ 
  forall concepts  $c_i \in C_L$  do
    if( $score(w_{c_i}) \leq \theta_b$ ) then
      delete  $c$  from  $o$ ;
      prune ( $o$ );
    end
    esle if (all  $score(w_{c_i}) \geq \theta_b$ ) then
      end
  end
end
end

```

In this algorithm, $score(w_{c_i}) = \frac{w_{c_i} - mean(c_i)}{\sigma(c_i)}$

Where

- θ_b is the significance threshold
- $mean(c_i)$ is the mean weight of the c_i
- $\sigma(c_i)$ is the deviation for the concept c_i .

Example following the above

The content trail changes into: $t_{cc} = \{(c_2, w_2), (c_5, w_5), (c_6, w_6)\}$

6.4 The Data Integration

In order to provide the most effective framework for pattern discovery and analysis, data from many sources are integrated. Integrate the three kinds of data sources: ontologies (content and structure), pedagogical information (learner data) and usage data.

Chapter 7 Pattern discovery from the trails of the learners

In view to analyse the trails and the learner ontologies, we mainly use two different techniques: the statistical theories (frequent itemset) and clustering.

1. The statistical theories can help us find out the frequent ontology of the learner and the frequent documents. This is the basic for the following mining techniques.
2. The clustering techniques can discover the learner groups and we can use this to group the new learner and give the recommendations.

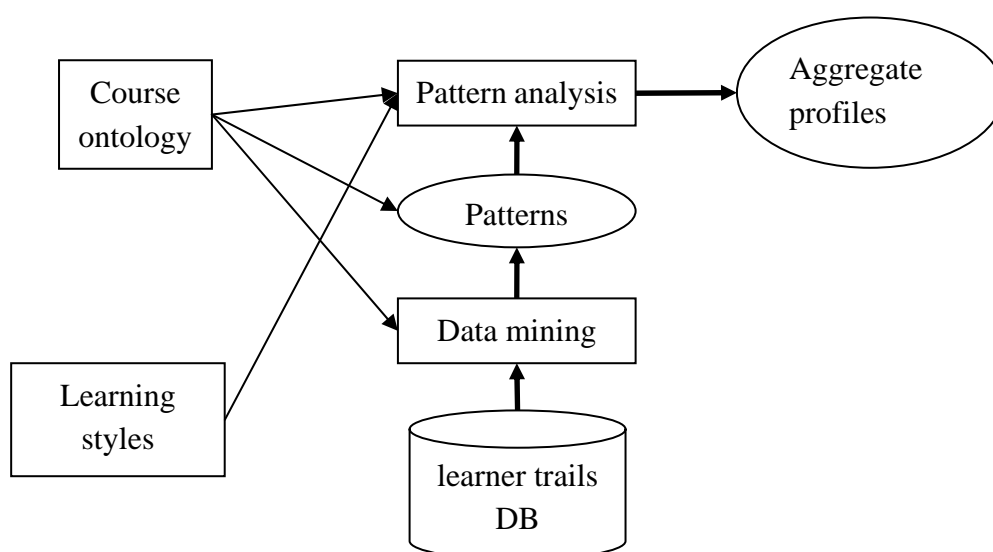


Figure 1-4 Pattern Discovery Component

7.1 Statistical Frequent Analysis

7.1.1 The frequent ontology Discovery

7.1.1.1 Discovery Phase

We try to find the common ontology of a topic. This ontology contains the concepts that the learners most frequently visit and with the high weight.

First, we have already pruned the individual ontology according to the weight in the preparation process in the chapter 6. So this is important to ensure our frequent ontology is valid.

Second, we map the learner ontology to the domain ontology. So like this, the ontology is easy to be analyzed and more formal.

At last, we calculate the support of each concept. The support of the concept c_i $\text{support}(c_i)$

presents the percentage of the ontologies that contains the concept c_i in the entire learner mapping ontology set O_m .

$$\text{support}(c_i) = \frac{|\{o \in O_m : c_i \in o\}|}{|O_m|}$$

From the support of the concepts, we can get ontology with the most frequently appearing concepts. If $\text{support}(c_i) \geq \text{min sup}$, we mark this concept as the frequent concept. And then we preserve these concepts in the domain ontology and obtain a frequent ontology. If the leaf concepts are frequent concepts, because we use the mapping ontology, then the concepts in the path from the root concept to the frequent leaf concepts must frequent also. Thus the frequent concepts compose a frequent ontology.

7.1.1.2 Metadata of the frequent ontology (pruning)

We have the frequent ontology, now we need to know the degree of the learner interest in the frequent concepts. So we need to use the statistical theories to give weight of these concepts, and find out the average time of the learner to spend in these concepts and then we will use the average times to give the recommendations.

A. The frequent ontology weight

We have given three weight types in the chapter 5. But in this part, the weight of the concept is not the one of the three ones, but we use the probabilistic of the concept to represent the weight of the concept in the frequent ontology. This is called the probabilistic weight. This weight can tell us the learner interests. If the more learners learn this concept, this concept is more important in the frequent ontology.

$$w_{c_i}^p = \text{support}(c_i)$$

B. The average time

The time is important factor. During the learning process, when we want to learn something, we must need to plan how much time we will spend in this topic. The learners need to know how much time will be necessary to learn a topic; this is in view to plan their time and efforts.

Therefore, we need to calculate the average time. We have defined in the data preparation phase (chapter 6), the raw weight likes the itself time of the learner spend in the concept and document.

Therefore, the average time is the average raw weight of the concept and document.

$$\bar{w}_{ci} = \frac{\sum_{j=1}^n w_{ci}^{o_j}}{|o|}, \quad o_j \in O_m$$

where

- o_j is the mapping ontology of the learner j and contains concept c_i .
- O_m is the set of the learner mapping ontology.
- $|o|$ is the number of the o_j .
- $w_{ci}^{o_j}$ is the weight of the concept c_i which belongs to the mapping ontology of the learner j .

7.1.2 The course frequent metadata

7.1.2.1 The course weight

The course weight is like the ontology weight. We use the support of the course to represent the weight of the course. The support of the course is the probability of how many learners choose this course to learn. This is called the frequent weight. The higher weight, the more popular of this course.

$$w_{course_i}^f = \text{support}(course_i)$$

Formula 7-1 the frequent weight of the course i

7.1.2.2 The course teaching style

The course teaching style is calculated by the learning styles of the learners who choose this course. In the chapter 3, we indicate the correspondences between the teaching styles and the learning styles. Thus, we calculate all the learners' average-learning style data as the course teaching style.

In the learner learning style, we know that in the same-learning style, the two type value are exclusive and must be one zero and one different of zero.

Preferred Learning Styles	Type	L	Value	Metadata Coverage: Teaching style	
				Element	T
Perception	Sensory	l_1	0 to 6	Content_Concrete	t_1
	Intuitive	l_2	0 to 6	Content_Abstract	t_2
Input	Visual	l_3	0 to 6	Presentation_Visual	t_3
	Verbal	l_4	0 to 6	Presentation_Verbal	t_4
Organization	Inductive	l_5	0 to 6	Organization_Inductive	t_5
	Deductive	l_6	0 to 6	Organization_Deductive	t_6
Processing	Active	l_7	0 to 6	StudentParticipation_Active	t_7
	Reflective	l_8	0 to 6	StudentParticipation_Passive	t_8
Understanding	Sequential	l_9	0 to 6	Perspective_Sequential	t_9
	Global	l_{10}	0 to 6	Perspective_Global	t_{10}

Table 7-1 The corresponding the learning style and the teaching style

Input

We have the set of learner learning style vector.

$$L_{learner}^{course_j} = \{ls_1, \dots, ls_n\}$$

$$ls_i = \{v(l_1), \dots, v(l_{10})\} ; v_q \in \{0, 1, 2, \dots, 6\}; l_q \in L$$

Where

— ls_i is the learning style vector of the learner i

—Learner i chooses the course j .

— $L_{learner}^{course_j}$ is the set.

— ls_i is the learning style vector.

— v_q is the value of the l_q , the type is integer form 0 to 6.

Calculate the teaching style

$$w_q = \frac{1}{n} \sum_{i=1}^n v_q^{ler-i}, i \in \{1, 2, \dots, 10\}, v_q^{ler-i} \in ls_i, ls_i \in L_{learner}^{course_j}$$

where

— t_q is the teaching style.

— v_q^{ler-i} is the value of the l_q of the learner i .

Thus the teaching style vector of the course j is in the following.

$$ts_j = \{w(t_1), \dots, w(t_{10})\} ; w_q \in [0, 6]; t_q \in T$$

7.1.2.3 Aggregate course list

The aggregate course list is the three-tuple vector.

$$List_{course}^{C_i^d} = \{(course_1, w_{course_1}^f, ts_{course_1}), \dots, (course_n, w_{course_n}^f, ts_{course_n})\}$$

$$subject(course_j) = C_i^d$$

Where

— C_i^d is the concept of the domain ontology.

7.1.3 The concept frequent metadata

7.1.3.1 The concept frequent weight

The concept weight is like the ontology weight. We use the support of the concept to represent the weight of the concept. The support of the course is the probability of how many learners choose this concept to learn. This is the frequent weight of the concept.

$$w_{c_i}^f = support(c_i)$$

7.1.3.2 The average formal weight

The formal weight presents the learner efforts to this concept. After the significance testing and the normalization, this weight can be considered as a rule to scale the different kind concepts. Thus we calculate the average formal weight as the concept usage metadata combine into the aggregate profiles.

$$w_{c_i}^{AF} = \frac{1}{n} \sum_{i=1}^n w_{learner_j}^{formal}$$

$$w_{learner_j}^{formal} = w_{c_i}^{formal} \text{ where } w_{c_i}^{formal} \in T_c^{learner_j}$$

Where

— $w_{c_i}^{formal} \in T_c^{learner_j}$ explains the $w_{c_i}^{formal}$ belongs to the concept trail of the learner j

$$(T_c^{learner_j}).$$

- Learner j is the learner who choose the concept c_i to learn, and the $w_{c_i}^{formal}$ is not zero.

7.1.3.3 Aggregate concept list

The aggregate concept list is the two-tuple vector.

$$List_{concept}^{course_j} = \left\{ (c_1, w_{c_1}^f, w_{c_1}^{AF}), \dots, (c_n, w_{c_n}^f, w_{c_n}^{AF}) \right\}$$

$$c_i \in C_{course_j}$$

Where

- C_{course_j} is the set of the concepts contain in the course j .

7.1.4 The document frequent metadata

7.1.4.1 The document weight

The document weight is like the course weight. We use the support of the document to represent the weight of the document. This support is after the filtering phase using the formal weight of the document in the chapter 6.

$$w_{d_i}^f = \text{support}(d_i)$$

7.1.4.2 The document teaching style

The document teaching style is the same as the course teaching style. It comes from the learning styles of the learners who effectively learn this course.

Input

We have the set of learner learning style vector.

$$L_{learner}^{d_j} = \{ls_1, \dots, ls_n\}$$

$$ls_i = \{v(l_1), \dots, v(l_{10})\} ; v_q \in \{0, 1, 2, \dots, 6\}; l_q \in L$$

Where

— ls_i is the learning style vector of the learner i

— Learner i chooses the course j .

— $L_{learner}^{d_j}$ is the set.

— ls_i is the learning style vector.

— v_q is the value of the l_q , the type is integer form 0 to 6

Calculate the teaching style

$$w_q = \frac{1}{n} \sum_{i=1}^n v_q^{ler-i}, i \in \{1, 2, \dots, 10\}, v_q^{ler-i} \in ls_i, ls_i \in L_{learner}^{course-j}$$

where

— t_q is the teaching style.

— v_q^{ler-i} is the value of the l_q of the learner i.

— w_q is the value of the t_q of the document teaching style T.

Thus the teaching style vector of the document j is in the following.

$$ts_j^d = \{w(t_1), \dots, w(t_{10})\}; w_q \in [0, 6]; t_q \in T$$

7.1.4.3 Aggregate document list

The aggregate course list is the three-tuple vector.

$$List_d^{C_i^d} = \{(d_1, w_{d_1}^f, ts_{d_1}), \dots, (d_n, w_{d_n}^f, ts_{d_n})\}$$

$$subject(dj) = C_i^c$$

Where

— C_i^c is the concept of the course ontology.

7.2 Clustering Approaches

Now we cluster the learners into groups in the offline phase, so like this in the online phase, we can easily target the new learner to the group and then according the analysis of this group to give the recommendations. There are two main steps in the clustering.

1. Content trail clustering: we cluster the learners according their content trials.
2. Learning styles: we use the content trail clustering as the base. We cluster the learners using their learning styles in each first clustering group.

7.2.1 K mean algorithm

7.2.1.1 Main procedures

The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. The algorithms steps of the k-means are in the following.

1. *Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.*
2. *Assign each object to the group that has the closest centroid.*
3. *When all objects have been assigned, recalculate the positions of the K centroids.*
4. *Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.*

7.2.1.2 Centroid

We use the centroid of the cluster to represent the cluster, so like this in the recommendation phase, we can calculate the similarity of the target learner and the centroid of the cluster, then find the most similar cluster to as the recommendation database.

7.2.2 Clustering by content trails

In this data analysis, we want to cluster the learners based on their content trails.

A. Input

1.1 Set of the content concepts

The set (This content concept trail is the last version in the data preparation component.)

$$C_c^{course-i} = \{c_1, \dots, c_n\}$$

1.2 Set of the learner content concept trails

$$T_{cc} = \{t_1^{cc}, \dots, t_m^{cc}\}$$

1.3 content concept trail vector

Each t_i^{cc} (content concept trail) is content concept vector.

$$t = \langle w(c_1, t), w(c_2, t), \dots, w(c_n, t) \rangle$$

where

- t is the learner content concept trail after preprocessing.
- $w(c_i, t)$ is the weight of the concept c_i in learner content concept trail t .

B. Content concept trail clusters

We use the cosine coefficient (the cosine of the angle between two vectors) to measure the similarity of the vectors.

If there are two content concept trails t_i^{cc} and t_j^{cc} , then the similarity is measured by the cosine angle of the vectors.

$$\text{sim}(t_i^{cc}, t_j^{cc}) = \frac{t_i^{cc} \bullet t_j^{cc}}{|t_i^{cc}| \times |t_j^{cc}|}$$

7.2.3 Clustering by learning styles

In the chapter 4, we have discussed the personalization information of the learners. And in this phase, we cluster the learners using the personalization information. Because we have already known that the personalization directly affects the learning processes: the course choosing, the sequence and the documents choose. So we cluster the learners using the personalization information in order to find the groups of the learners. Like this, in the online recommendation, we can target the learner to the group according his personalization information and then give the recommendations. In the chapter 4, we have already given the basic theories to prove this method.

In fact, the learning styles just affect the learning pattern of the learning process and the document choosing. It does not affect the content of the learning process. Thus if we just base on this to cluster the learners, the results will not sufficient to give the complete recommendations. Therefore, we consider clustering the learners base on the first cluster of the content concept trail. In each cluster of the content concept trail cluster, we cluster the learners second time use the learning style as the scale.

The learning style vector ls_i of the learner is in the following definition. (we give the definition in the chapter 4.

$$ls_i = \{(l_1, v_1), \dots, (l_{10}, v_{10})\}$$

$$v_j \in \{0, 1, 2, \dots, 6\}$$

$$l_j \in L$$

A. Input

1. Set of learner learning style

$$L_{learner} = \{ls_1, \dots, ls_n\}$$

2. Learner learning styles vector

$$ls_i = \{v(l_1), \dots, v(l_{10})\}$$

$$v_j \in \{0, 1, 2, \dots, 6\}$$

$$l_j \in L$$

Where

— ls_i is the learning style vector.

— v_j is the value of the l_j , the type is integer form 0 to 6.

B. Learning style clusters

$$sim(ls_i, ls_j) = \frac{ls_i \bullet ls_j}{|ls_i| \times |ls_j|}$$

If the sample set is large, we can use two level clustering: the content trail clustering and then the learning style clustering. This can save the time during the recommendation component, because the cluster is small and easily to find the nearest neighbor. However, if the sample is not large enough, this will be adaptive. In case of this problem, we will not apply the second cluster. While in the recommendation, we will directly use the learning style to find the nearest neighbor in the first cluster group, but not the content .

Chapter 8 Recommendations for the Learner

The recommendation is the final phase of the system. This component is about the algorithm of generating the recommendation.

1. The course recommendation: we generate the rank of the course by combining the course usage information and the learning style information.
2. The concept recommendation:
 - a. After the course, at the beginning learning phase, we use the KNN of the learning style to generate the range of the concept.
 - b. When the learner have a certain history, we consider the learning style and the content trail to find the nearest neighbor and to generate the concept range.
3. The documents design: we generate the range of the document by combining the document usage information and the learning style information.

In this component, the main design is in the Figure 8-1. The aggregate profiles are the input. We compare the integrated learner profile to compare with the aggregate profile to generate the recommendation. The main algorithm used is the KNN algorithm.

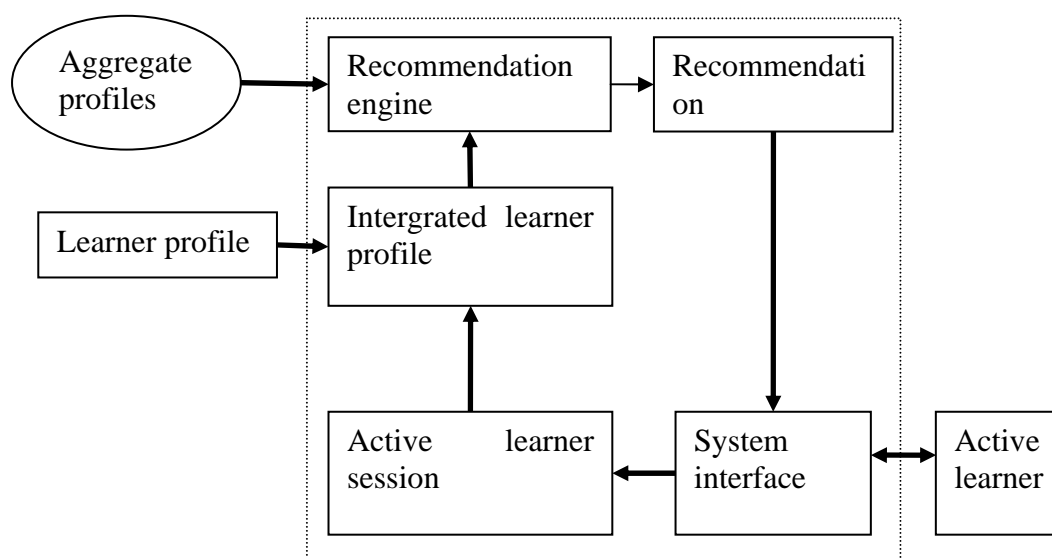


Figure 1-5 Recommendation Component

8.1 The KNN Based Approach

Collaborative filtering based on K-Nearest-Neighbor (KNN) approach involves the active learner record as a target learner and compare with the other learners' records in order to find the most similar k top learners. Thus, there are two main phases in the collaborative filtering: the neighborhood formation phase and the recommendation phase.

Purpose of this algorithm is to classify a new object based on attributes and training samples. The classifiers do not use any model to fit and only based on memory. Given a query point, we find K number of objects or (training points) closest to the query point. The classification is using majority vote among the classification of the K objects. Any ties can be broken at random. K Nearest neighbor algorithm used neighborhood classification as the prediction

value of the new query instance.

Similarity measure

We use the cosine coefficient to measure the similarity. The similarity is defined as the following formula.

$$\text{sim}(\bar{t}_i, \bar{t}_j) = \frac{\bar{t}_i \bullet \bar{t}_j}{|\bar{t}_i| \times |\bar{t}_j|}$$

$$\bar{T}_i = (d_{i1}, \dots, d_{in})$$

The cosine of the two vectors are defined in the follow.

$$\cos(\bar{T}_i, \bar{T}_j) = \frac{\sum_{i=1}^n t_i \cdot t_j}{\sqrt{\sum_{i=1}^n t_i^2 \cdot \sum_{i=1}^n t_j^2}}$$

Thus we can get the k neighborhood of the query instance.

$$NB = \{\bar{T}_1, \dots, \bar{T}_n\}$$

8.2 The Course Recommendation

When the learner first chooses the topic to learn, he chooses one course in a courses list. (These courses all belong to this topic). Thus we need to rank these courses. In the following, I give the algorithms to generate this rank from the course usage information and the learning style information.

A. Input: the course aggregate profile

1. In this profile, we have the course list vector. Each list supports the same concept in the domain ontology.

$$List_{course}^{C_i^d} = \left\{ (course_1, w_{course_1}^f, ts_{course_1}), \dots, (course_n, w_{course_n}^f, ts_{course_n}) \right\}$$

$$subject(course_j) = C_i^d$$

Simplification

$$List_{course}^{C_i^d} = \left\{ (w_{course_1}^f, ts_{course_1}), \dots, (w_{course_n}^f, ts_{course_n}) \right\}$$

Where

- Each t_s is a vector. The detail definition in Chapter 7.

$$ts_j = \{w(t_1), \dots, w(t_{10})\} ; w_q \in [0, 6]; t_q \in T$$

2. The query instance: the learner learning style vector

$$ls_i = \{(l_1, v_1), \dots, (l_{10}, v_{10})\}$$

$$v_j \in \{0, 1, 2, \dots, 6\}$$

$$l_j \in L$$

Where

— ls_i is the learning style vector.

— v_j is the value of the l_j , the type is integer form 0 to 6

2. The neighborhood formation: Algorithms of the KNN

We use the learner ls_i as the query instance to find the k nearest neighbors from the courses list.

$$sim(ls_i, ts_j) = \frac{ls \bullet ts_j}{|ls| \times |ts_j|}$$

$$ts_j = \{w(t_1), \dots, w(t_{10})\} , w_q \in [0, 6], t_q \in T$$

$$ls = \{(l_1, v_1), \dots, (l_{10}, v_{10})\}, v_j \in \{0, 1, 2, \dots, 6\}, l_j \in L$$

Then we get the NB_s of the learner learning style. These neighbors are the courses which are suitable for the learner learning styles.

$$NB_{course} = \{ts_1, \dots, ts_k\}$$

3. The recommendation

In order to determine which courses are to be recommended, a recommendation score is computed for each $course_i \in List_{course}^{C_i^d}$ based on the neighborhood for the active learner. Two factors are used in determining this recommendation score: the similarity of the learner ls to the neighborhood and the frequent weight of the each course.

$$rec(ls, course_j) = \sqrt{w_{course_j}^f \times sim(ls, ts_j)}$$

8.3 The Concept Recommendation

After the course is chosen, the next step is to give the concept recommendation.

A. Input: the concept aggregate profile

In this profile, we have the concept list vector. All the concept in the list belongs to the same course j .

$$List_{concept}^{C_{course_j}} = \left\{ (c_1, w_{c_1}^f, w_{c_1}^{AF}), \dots, (c_n, w_{c_n}^f, w_{c_n}^{AF}) \right\}$$

$$c_i \in C_{course_j}$$

Simplification

$$List_{concept}^{C_{course_j}} = \left\{ (w_{c_1}^f, w_{c_1}^{AF}), \dots, (w_{c_n}^f, w_{c_n}^{AF}) \right\}$$

B. The recommendation

In order to determine which concepts are to be recommended, a recommendation score is computed for each $c_i \in C_{course_j}$. Two factors are used in determining this recommendation score: the course frequent weight and the average formal weight.

$$rec(c_i) = \sqrt{w_{c_i}^f \times w_{c_i}^{AF}}$$

8.4 The Document Recommendation

After the learner chooses one concept prepares to learn, we need to give the recommendation: which documents belong to this concept should learn. Thus for each document individual list, we have the aggregate profile explained in the chapter 7.

A. Input: the document aggregate profile

1. In this profile, we have the document list vector. Each list supports the same concept in the domain ontology.

$$List_d^{C_i^c} = \left\{ (d_1, w_{d_1}^f, ts_{d_1}), \dots, (d_n, w_{d_n}^f, ts_{d_n}) \right\}$$

$$subject(dj) = C_i^c$$

Simplification

$$List_d^{C_i^c} = \left\{ (w_{d_1}^f, ts_{d_1}), \dots, (w_{d_n}^f, ts_{d_n}) \right\}$$

$$subject(dj) = C_i^c$$

Where

— Each t_s is a vector. The detail definition in the chapter 7.

$$ts_j = \{w(t_1), \dots, w(t_{10})\} ; w_q \in [0, 6]; t_q \in T$$

2. The query instance: the learner learning style vector

$$ls_i = \{(l_1, v_1), \dots, (l_{10}, v_{10})\}$$

$$v_j \in \{0, 1, 2, \dots, 6\}$$

$$l_j \in L$$

Where

— ls_i is the learning style vector.

— v_j is the value of the l_j , the type is integer form 0 to 6

2. The neighborhood formation: Algorithms of the KNN

We use the learner ls_i as the query instance to find the k nearest neighbors from the document individual list.

$$sim(ls_i, ts_j) = \frac{ls \bullet ts_j}{|ls| \times |ts_j|}$$

$$ts_j = \{w(t_1), \dots, w(t_{10})\}, w_q \in [0, 6], t_q \in T$$

$$ls = \{(l_1, v_1), \dots, (l_{10}, v_{10})\}, v_j \in \{0, 1, 2, \dots, 6\}, l_j \in L$$

Then we get the NB_{course} of the learner learning style. These neighbors are the courses which are suitable for the learner learning styles.

$$NBdocument = \{ts_1, \dots, ts_k\}$$

3. The recommendation

In order to determine which documents are to be recommended, a recommendation score is computed for each $d_j \in List_d^{C_i}$ based on the neighborhood for the active learner. Two factors are used in determining this recommendation score: the similarity of the learner ls to the neighborhood and the frequent weight of the each document.

$$rec(ls, d_j) = \sqrt{w_{d_j}^f \times sim(ls, ts_j)}$$

Chapter 9 Conclusion and Future Work

9.1 Conclusion

In this paper, we introduced the framework of a personalized recommendation system based on the students' trails. We mainly use the semantic ontology and the pedagogical models as the basic to capture the learner trails. The following, we use the data mining to analyze and generate the recommendation.

For our system, in fact, we consider the preferred learning styles of the learners, but we not just use these learning styles to the learners. After the learners master the knowledge, we consider using other learning styles, which the learners do not like. We use these to complete the learners' competences to make them master all the learning styles.

9.2 Future Work

In the future, there are many works from two sides.

1. We need to create software to realize this system. And if it is possible, we need to test in the computer course. Like this, we can get the first data. But we know that the learning process is long, thus it must need one semester at least.
2. For the system design, in the pattern analysis design, we just use the clustering, but there are many other data mining techniques which I have already consider to add, the association rules, the sequence mining.
3. In the recommendation phase, the algorithms need to ameliorate when get the data results.
4. Relation mining: we need to study the relations in the mining phase, like which kind of relations are first be learned by the learners.

9.3 Experiment

For our approaches, there are three main phase to execute in order to make the experiments. For this experiment, we consider to apply in the specialty of the mater 2 informatics researches in the university of the Montpellier II.

1. We need to design the course and the documents. This phase is done by the teachers.
2. For the learners, we need to design a function to permit the learners to record their learning concepts and construct their learning ontology.
3. For the third phase, it is the main analysis component of the approaches. We need to develop the prototype of the three components: the data preparation, the pattern analysis and the online recommendation.

The difficulty is the time of capture the learner trails. The 50 students' trails are enough to analyze but the time to capture the learners' trails, perhaps one semester at least.

Reference

- [1] Keefe, J.W., "Learning Styles: An Overview," in Keefe, J.W., ed., *Student Learning Styles: Diagnosing and Prescribing Programs*, Reston, Va.: National Association of Secondary School Principals, 1979.
- [2] Felder, R.M., and Soloman, B.A., *Index of Learning Styles*, [Http://www.ncsu.edu/felder-public/ILSpage.html](http://www.ncsu.edu/felder-public/ILSpage.html).
- [3] Felder, R.M., and Silverman, L.K., "Learning and Teaching Styles in Engineering Education," *Engineering Education*, Vol. 78, No. 7, 1988, pp. 674-681.
- [4] H. Dai and B. Mobasher, Integrating Semantic Knowledge with Web Usage Mining for Personalization, *In Web Mining: Applications and Techniques*, Anthony Scime (ed.), Idea Group Publishing, 2005
- [5] B. Mobasher, Web Usage Mining and Personalization, *In Practical Handbook of Internet Computing*, Munindar P. Singh (ed.), CRC Press, 2005
- [6] B. Mobasher, H. Dai, T. Luo, Y. Sung, M. Nakagawa and J. Wiltshire, Discovery of Aggregate Usage Profiles for Web Personalization, *Proceedings of the Web Mining for E-Commerce Workshop (WebKDD'2000)*, Held at KDD'2000, August 2000, Boston
- [7] R. Cooley, B. Mobasher, and J. Srivastava, Data Preparation for Mining World Wide Web Browsing Patterns, *Knowledge and Information Systems*, Vol. 1, No. 1, 1999.
- [8] O.R. Zaïane, Web usage mining for a better web-based learning environment, in: *Proceedings of Conference on Advanced Technology for Education*, 2001, pp. 450-455.

Vocabulary

Definition

Definition 1: Ontology:.....	12
Definition 2: Domain ontology	13
Definition 3: Main domain ontology	13
Definition 4: Sub domain ontology.....	13
Definition 5: Learning style	18
Definition 6: Teaching style	18
Definition 7: Cognitive trail.....	30
Definition 8: Learner ontology trail.....	30
Definition 9: Navigational trail.....	30
Definition 10: Statement.....	32
Definition 11: Statement Trail.....	33
Definition 12: Concept Trail	33
Definition 13: Efficient concept trail	34
Definition 14: Content concept trail	34
Definition 15: Relation trail	34
Definition 16 : Document trail.....	35
Definition 17 : Collaborative document trail	35
Definition 18 : Individual document trail	36
Definition 19: Type of the concepts.....	44

Figure

Figure 1-1 Semantic Model	4
Figure 1-2 Trail model	5
Figure 1-3 Data Preparation Component	6
Figure 1-4 Pattern Discovery Component	6
Figure 1-5 Recommendation Component	7
Figure 2-1 Three levels of ontology model.....	11
Figure 2-2 Metadata model of the relationship between the three ontology	11
Figure 2-3 Semantic model of the learning materials.....	12
Figure 2-4 Content Ontology	13
Figure 2-5 Context Ontology	15
Figure 4-1 Structure of the Learner Personalization.....	24
Figure 5-1 Learning and Remembering Meaningful Information: A Cognitive Model	27
Figure 5-2 The activities of the learners to construct the ontology	28
Figure 5-3 The trails versus the activity	29
Figure 5-4 Trails design	30
Figure 5-1 The Application of the Trails.....	37
Figure 6-1 Data Preparation Component	40
Figure 6-2 Domain ontology.....	41
Figure 6-3 Learner ontology	41
Figure 6-4 Mapping the learner ontology to the domain ontology	43
Figure 6-5 The mapped learner ontology.....	44
Figure 6-6 The mapping ontology after add the type of the concepts	45
Figure 6-7 After pruning	46

Table

Table 1-1 The differences between the e-commerce and the e-learning.....	3
Table 1-2 The solutions to the e-learning.....	3
Table I-1 The trail of the example 1(Figure I-1).....	8
Table 2-1 Dublin core element.....	16
Table 2-2 The teaching style element	16
Table 3-1 Dimension of Learning and Teaching styles.....	20
Table 3-2 The main parts which the learning styles affect.....	21
Table 3-3 The recommendation item that corresponding to the styles.	22
Table 4-1 Concept Details.....	25
Table 4-2 The elements of the basic information.....	25
Table 5-1 Example of the document trail.....	36
Table 6-1 Characteristics of the four type of concepts.....	45
Table 6-2 Content trail concepts	46
Table 7-1 The corresponding the learning style and the teaching style	59

Annex

Annex 1 Index of Learning Styles Questionnaire

NC STATE UNIVERSITY

Index of Learning Styles Questionnaire

Barbara A. Soloman
First-Year College
North Carolina State University
Raleigh, North Carolina 27695
Richard M. Felder
Department of Chemical Engineering
North Carolina State University
Raleigh, NC 27695-7905

Directions

Please provide us with your full name. Your name will be printed on the information that is returned to you.

Full Name

For each of the 44 questions below select either "a" or "b" to indicate your answer. Please choose only one answer for each question. If both "a" and "b" seem to apply to you, choose the one that applies more frequently. When you are finished selecting answers to each question please select the submit button at the end of the form.

I understand something better after I

(a) try it out.

(b) think it through.

I would rather be considered

(a) realistic.

(b) innovative.

When I think about what I did yesterday, I am most likely to get

(a) a picture.

(b) words.

I tend to

(a) understand details of a subject but may be fuzzy about its overall structure.

(b) understand the overall structure but may be fuzzy about details.

When I am learning something new, it helps me to

(a) talk about it.

(b) think about it.

If I were a teacher, I would rather teach a course

(a) that deals with facts and real life situations.

(b) that deals with ideas and theories.

I prefer to get new information in

(a) pictures, diagrams, graphs, or maps.

(b) written directions or verbal information.

Once I understand

(a) all the parts, I understand the whole thing.

(b) the whole thing, I see how the parts fit.

In a study group working on difficult material, I am more likely to

(a) jump in and contribute ideas.

(b) sit back and listen.

I find it easier

(a) to learn facts.

(b) to learn concepts.

In a book with lots of pictures and charts, I am likely to

(a) look over the pictures and charts carefully.

- (b) focus on the written text.

When I solve math problems

- (a) I usually work my way to the solutions one step at a time.

- (b) I often just see the solutions but then have to struggle to figure out the steps to get to them.

In classes I have taken

- (a) I have usually gotten to know many of the students.

- (b) I have rarely gotten to know many of the students.

In reading nonfiction, I prefer

- (a) something that teaches me new facts or tells me how to do something.

- (b) something that gives me new ideas to think about.

I like teachers

- (a) who put a lot of diagrams on the board.

- (b) who spend a lot of time explaining.

When I'm analyzing a story or a novel

- (a) I think of the incidents and try to put them together to figure out the themes.

- (b) I just know what the themes are when I finish reading and then I have to go back and find the incidents that demonstrate them.

When I start a homework problem, I am more likely to

- (a) start working on the solution immediately.

- (b) try to fully understand the problem first.

I prefer the idea of

- (a) certainty.

- (b) theory.

I remember best

- (a) what I see.

- (b) what I hear.

It is more important to me that an instructor

- (a) lay out the material in clear sequential steps.

- (b) give me an overall picture and relate the material to other subjects.

I prefer to study

- (a) in a study group.

- (b) alone.

I am more likely to be considered

- (a) careful about the details of my work.

- (b) creative about how to do my work.

When I get directions to a new place, I prefer

- (a) a map.

- (b) written instructions.

I learn

- (a) at a fairly regular pace. If I study hard, I'll "get it."

- (b) in fits and starts. I'll be totally confused and then suddenly it all "clicks."

I would rather first

- (a) try things out.

- (b) think about how I'm going to do it.

When I am reading for enjoyment, I like writers to

- (a) clearly say what they mean.

- (b) say things in creative, interesting ways.

When I see a diagram or sketch in class, I am most likely to remember

- (a) the picture.

- (b) what the instructor said about it.

When considering a body of information, I am more likely to

- (a) focus on details and miss the big picture.
- (b) try to understand the big picture before getting into the details.
- I more easily remember
- (a) something I have done.
- (b) something I have thought a lot about.
- When I have to perform a task, I prefer to
- (a) master one way of doing it.
- (b) come up with new ways of doing it.
- When someone is showing me data, I prefer
- (a) charts or graphs.
- (b) text summarizing the results.
- When writing a paper, I am more likely to
- (a) work on (think about or write) the beginning of the paper and progress forward.
- (b) work on (think about or write) different parts of the paper and then order them.
- When I have to work on a group project, I first want to
- (a) have "group brainstorming" where everyone contributes ideas.
- (b) brainstorm individually and then come together as a group to compare ideas.
- I consider it higher praise to call someone
- (a) sensible.
- (b) imaginative.
- When I meet people at a party, I am more likely to remember
- (a) what they looked like.
- (b) what they said about themselves.
- When I am learning a new subject, I prefer to

(a) stay focused on that subject, learning as much about it as I can.

(b) try to make connections between that subject and related subjects.

I am more likely to be considered

(a) outgoing.

(b) reserved.

I prefer courses that emphasize

(a) concrete material (facts, data).

(b) abstract material (concepts, theories).

For entertainment, I would rather

(a) watch television.

(b) read a book.

Some teachers start their lectures with an outline of what they will cover.
Such outlines are

(a) somewhat helpful to me.

(b) very helpful to me.

The idea of doing homework in groups, with one grade for the entire group,

(a) appeals to me.

(b) does not appeal to me.

When I am doing long calculations,

(a) I tend to repeat all my steps and check my work carefully.

(b) I find checking my work tiresome and have to force myself to do it.

I tend to picture places I have been

(a) easily and fairly accurately.

(b) with difficulty and without much detail.

When solving problems in a group, I would be more likely to

(a) think of the steps in the solution process.

(b) think of possible consequences or applications of the solution in a wide range of areas.

When you have completed filling out the above form please click on the Submit button below. Your results will be returned to you. If you are not satisfied with your answers above please click on Reset to clear the form.

Dr. Richard Felder, felder@ncsu.edu

Annex 2 The result of the ILS example

NC STATE UNIVERSITY

Learning Styles Results

Results for: yulin

ACT							X					REF
	11	9	7	5	3	1	1	3	5	7	9	11
							<--	-->				

SEN										X		INT
	11	9	7	5	3	1	1	3	5	7	9	11
							<--	-->				

VIS				X								VRB
	11	9	7	5	3	1	1	3	5	7	9	11
							<--	-->				

SEQ										X		GLO
	11	9	7	5	3	1	1	3	5	7	9	11
							<--	-->				

- If your score on a scale is 1-3, you are fairly well balanced on the two dimensions of that scale.
- If your score on a scale is 5-7, you have a moderate preference for one dimension of the scale and will learn more easily in a teaching environment which favors that dimension.
- If your score on a scale is 9-11, you have a very strong preference for one dimension of the scale. You may have real difficulty learning in an environment which does not support that preference.

We suggest you print this page, so that when you look at the explanations of the different scales you will have a record of your individual preferences.

For explanations of the scales and the implications of your preferences,
click on

[Learning Style Descriptions.](#)

For more information about learning styles or to take the test again, click on

[Learning Style Page.](#)

