EXAM-BASED EDUCATION SYSTEM

Thesis
Submitted to
The College of Arts and Sciences of the
UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for
The Degree of
Master of Computer Science

By
Xuhang Song
Dayton, Ohio
May, 2014
EXAM BASED EDUCATION SYSTEM

NAME: Song, Xuhang

APPROVED BY:

________________________________________________
James P. Buckely, Ph.D.
Faculty Advisor

________________________________________________
Dale E. Courte, Ph.D.
Committee Member

________________________________________________
Zhongmei Yao, Ph.D.
Committee Member
ABSTRACT

EXAM BASE EDUCATION SYSTEM

Name: Song, Xuhang
University of Dayton
Advisor: Dr. James P. Buckley

Big data is growing in importance in everyday life, yet traditional models of University education do not make good use of it. This thesis proposes a system that allows students to find courses based on similarity measures and take these courses in an exam-based environment. We describe a new mining method that can efficiently search, cluster and perform related functions in the system. The basic idea of this mining is to map courses in a university to buildings in a city. This means that finishing a degree or getting a skill is analogous to finding a path in the city. A number of approaches to build the city are presented. In the process of developing an algorithm, we use machine learning, artificial intelligence, and classic mining methods (ID3, K-Mean, and Page Rank) as core ideas.
Dedicated to my parents
ACKNOWLEDGEMENTS

I am grateful for support from many people. Foremost I would like to thank my advisor, Dr. James Buckley, for giving me guidance. Special thanks to Dr. Dale Courte and Dr. Zhongmei Yao, the members of my thesis committee. They spend a lot of time reviewing my thesis draft with Dr. Buckley. Without their suggestions and encouragement, I can hardly imagine how I can complete my thesis.

I also would like to thank all faculties in the computer science department of University of Dayton. What impressed me most are their respectable personality as well as their expertise.

A sincere note of thanks to my parents for their unconditional care and help.
Table of Contents

ABSTRACT .................................................................................................................... iv
ACKNOWLEDGEMENTS ............................................................................................... vi

CHAPTER 1: INTRODUCTION ...................................................................................... 1
  1.1 Overview ............................................................................................................... 1
  1.2 Exam-base education system ............................................................................... 1
  1.3 Data mining ........................................................................................................... 2

CHAPTER 2: Background .............................................................................................. 3
  2.1 Techniques ........................................................................................................... 3
  2.2 Data Mining ........................................................................................................... 3
  2.3 Machine Learning ................................................................................................ . 4
  2.4 Genetic Algorithm ................................................................................................ . 5
  2.5 C4.5 and K-Mean .................................................................................................. 7
  2.6 System Benefits .................................................................................................... 8

CHAPTER 3: Exam-based Study System Structure ...................................................... 12
  3.1 Exam-based Study System Components ............................................................. 12
  3.2 Terminology .........................................................................................................14
  3.3 Classes of System Users ..................................................................................... 15

CHAPTER 4: System Function Using Data Mining ...................................................... 18
  4.1 Overview ..............................................................................................................18
  4.2 Classes of Users .................................................................................................. 18
  4.3 EES Database .....................................................................................................19
  4.4 What are the needs of the user? ..........................................................................22
  4.5 Data mining in the EES ......................................................................................23

REFERENCES ..............................................................................................................40
CHAPTER 1

INTRODUCTION

1.1 Overview

This thesis is divided into two parts: One is the description of a new education system that allows students, faculty, and employers a completely different perspective on learning outcomes. Then it is shown how this data can be mined for useful knowledge.

This thesis defines a new mining method that can efficiently search, cluster and perform other related functions in the system.

1.2 Exam-base education system

Big data is growing in importance in everyday life, yet traditional models of University education do not make good use of it. This thesis proposes a system that allows students to find courses based on similarity measures and take these courses in an exam-based environment. The system can give both student and teacher benefits.
1.3 Data mining

Data mining is very useful for detecting patterns in our data that are not obvious and are not trivial. This thesis presents a new mining method that, although specific for the system, it can be used in other similar areas as well.

The basic idea of this mining method is to map courses in a university to buildings in a city. This means that finishing a degree or getting skill in the system is analogous to finding a path in the city.

A number of approaches to build the city are presented. In the process of developing the algorithm, we use machine learning, artificial intelligence, and classic mining methods (ID3, K-Mean, Page Rank) as core ideas.
CHAPTER 2

Background

2.1 Techniques

This thesis defines new algorithms for grouping similar courses and finding useful information for students, educators, and companies looking for future employees. These algorithms incorporate a number of existing techniques and technologies. In the following sections, I provide the necessary background information on each.

The main idea of all the technology is helping users retrieve information from the system easier.

2.2 Data Mining

Simply stated, data mining refers to extracting or “mining” knowledge from large amount of data. The process of data mining include data cleaning, data
integration, data selection, Date transformation, data mining, pattern evaluation, knowledge presentation. [1]

Data mining in general is the search for hidden patterns that may exist in large databases. Spatial data mining in particular is the discovery of interesting relationships and characteristics that may exist implicitly in spatial databases. [2]

As databases grow bigger and bigger, it is more and more important to use data mining methods. Data mining can retrieve the knowledge which is hardly to be found by human beings, especially when there are tons of information in the database.

In this thesis, data mining will be discussed in specific cases.

### 2.3 Machine Learning

Learning means to get knowledge by study, experience or being taught. Learning means to become aware by information or from observation. Machine learning means when machines change their behavior in a way that makes them perform better in the future. [3]

This thesis will make machine learns itself and then get the better result. Some of the result is hard to get without machine learning. However, the
database will become really big. The efficiency of data mining method becomes really important.

Hard problems need a bigger population and this translates directly into higher computational costs [4]

Mining information and knowledge from large databases has been recognized by many researchers as a key research topic in database systems and machine learning, and by many industrial companies as an important area with an opportunity of major revenues. Researchers in many different fields have shown great interest in data mining. Several emerging applications in information providing services, such as data warehousing and on-line services over the Internet, also call for various data mining techniques to better understand user behavior, to improve the service provided, and to increase the business opportunities.[5]

2.4 Genetic Algorithm

Genetic algorithms are a class of stochastic search algorithm based on biological evolution. GA represents an iterative process. Each iteration is called the generation. The entire generations is called a run. As the end of a run, we expect to find one or more highly fit chromosomes. [6]
Genetic Algorithms are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specific problem on a simple chromosome like data structure and apply recombination operators to these structures so as to preserve critical information. Genetic algorithms are often viewed as function optimizers, although the range of problems to which genetic algorithms have been applied is quite broad.

Usually there are only two main components of most genetic algorithms, the problem encoding and the evaluation function. [7]

Evolutionary algorithms model natural evolution by asexual reproduction with mutation and selection. [8]

Across a wide variety of fields, data are being collected and accumulated at a dramatic pace. There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. [9]

In this thesis, we use genetic algorithm to find the best position for each building in the city. In another word, we use this technology to find out the relationship between each course.
2.5 C4.5 and K-Mean

C4.5 is a software extension of the basic ID3 algorithm designed by Quinlan to address the following issues not dealt with by ID3. It is a method to make decision tree. [10]

The algorithm of this article use part of this method.

The technology for building knowledge-based systems by inductive inference from examples has been demonstrated successfully in several practical applications. An approach to synthesizing decision trees, ID3, has been used in a variety of systems. Results from recent studies show ways in which the methodology can be modified to deal with information that is incomplete. [11]

The algorithm called k-mean consists of a simple re-estimation procedure as follows. Initially, the data points are assigned at random to the $K$ sets. For step 1, the centroid is computed for each set. In step 2, every point is assigned to the cluster whose centroid is closest to that point. These two steps are alternated until a stopping criterion is met, i.e., when there is no further change in the assignment of the data points. [12]

This thesis uses the same approach to cluster the courses, but the steps that calculate difference between nodes are different from K-Mean algorithm.
In the past 30 years, cluster analysis has been widely applied to many areas such as medicine (classification of diseases), chemistry (grouping of compounds), and social study (classification of statistical findings), and so on. Its main goal is to identify structures or clusters present in the data. [13]

2.6 System Benefits

In this section we will discuss the benefit of the system. For there are three type of users, we will discuss it separately.

a) For student

1) When students attend university and take the course, they first attend the lecture, then review the material, and finally take the exam. If they are good students, they read through the material before lecture, find question, and then attend lecture, find if we can resolve the question, if not, ask the professor in class. But for most students, it is hard to do like that. Most students attend the lecture without knowing anything. In that case, it is really hard for them to focus on the lecture, because human cannot focus for a long time. The lecture is totally new for them, so they have to remember while understanding, if there is a question, thinking the question will cause some miss about the lecture. So the traditional lectures seem to waste some time and not efficiency. But almost all the people can focus them study before the exam. Cause every one want to get a good result in
the exam. If we change the order of the course, first, we let the student go through the course material, and then let them find question and ask, then take the exam, and then take the course, it will be much better. Cause less time is wasted. On the other hand, we see that after exam, the lecture is useless cause the student has known what they need. So we just make material – ask question – exam as a circle (like figure 2.1). It will save the students’ time and make things more efficiency, because students can use nervous energy to become more efficiency.

2) Students have fewer limits when using this system. They can arrange their own time to read through material, ask question and choose the best time to attend exam. The can use their own speed to finish all the tasks. I always think reading book is a better entertain than watching TV or movie, because the speed we read is under control. We do not need to follow the time the TV or movie gives us. It is the same as the system. Student can follow their own steps and do not need to follow the time others give them.

3) When I choose my major, I know nothing about my major; I do not know what my major can give me. I believe most student meet that problem. They just guess what the major looks like, and then choose them. May be it is totally different from the previous thought. When they find the major is not what they want, there has been nearly ten course past, it hard to come back at that time. But a system is helpful. The system can let the student choose what ability they actually want and recommend exam to the
student. The student do not need to care about the major or something useless, just take the exam then they can get what they want.

4) This system let the student discuss the knowledge easier. When we have class in university, we can only discuss within this class. We cannot discuss with the one who take the class in previous year. But use this system we can discuss with the one who take the class all over the world.

5) In a lot of situations, we just need a few knowledge to resolve problem. For example, when we need to figure out how to build a button in IOS, most of persons won’t try to find a book related to the technology, they just search online, and find specific method to resolve the problem. Students do not need to get whole courses in most cases. They just need to learn one element so that they can get all they want.

b) For teacher

1) Everyone can be a teacher. Anyone who wants to share their knowledge can upload element. They do not need to be a real professor.

2) In general courses, the professors must prepare for a whole course. At this system, they can upload whatever they want and build the related exam. It is much easier for a professor because they can only upload the most interesting part they want to the system. Element is a really small course, so teachers must be really good at them.

c) For organization
1) Companies always have training part for new employee, so that they can make sure their stuffs have the knowledge they want. When they use this system, they can set up the ability they want. So they can get the skilled employee in the beginning.

2) They can put some training course directly in the system, it will make the company closer to education.

d) For developer

We will put all education courses together, so it is much easier to do data mining or source arrangement.
CHAPTER 3

Exam-based Study System Structure

3.1 Exam-based Study System Components

The Exam-base study system helps university students gain skill and knowledge by learning material at their own pace. There are three kinds of users, student, teacher and organization.

The system consists of many small courses called *elements*. A *course* consists of one or more elements. Some specific courses are required to get *ability*, which is like a skill. Figure 3.1 shows the interrelationship between these concepts. A specific example of these concepts is shown in Figure 3.2.
Figure 3.1: Relationship among ability, course and element
3.2 Terminology

In this section I more formally define the components and associated terms involved in the exam-based study system.

This system has many small course segments called course elements. Element courses are fairly short in duration, generally less than 1 hour long. Student can use 1 hour or less to finish the lecture. Element courses are uploaded by the professor. Some element courses can only read by the student. A normal element course is one where the professor set up tests and assignments. Cost to students is greater in this case.
3.3 Classes of System Users

There are three basic classes of users of the exam-based study system:

- Students
- Teachers
- Organizations.

**Students**

The student takes exams, gains knowledge and obtains grades. Grades demonstrate their ability so that they know what they learned.

**Teacher**

The teachers are the people who want to share their knowledge and build the related exam.

**Organization**

The one who set up really exam for the student and set up ability (the student finish some related course can get ability, it is used for company to find good employee)

Students, gain knowledge through these steps:

1. determine desired ability
2. find which courses are required
3 find which elements make up these courses
4 learn the material related to the element
5 take the online test for the element
6 take the exam for entire the course
7 gain ability.

Students learn by using cycles, like “read material – take exam – get grade and knowledge – read other material”, to help students remember and understand the material. This is shown in Figure 2.3 below.

Figure 3.3: Step 4 -6 for student users to use the system

For teachers, they try to teach like this:
1 Upload the materials they want as elements
2 Build related test for the elements.
3 Find several elements and make them as a course.
4 contact organization and hold exam with them.

Of course they can only choose one or two steps to teach, like figure 2.4, for example. They do not need to upload their own elements, they can just choose step 3 and 4 then they can have a course.

This is the same thing as, teachers do not need to write text book themselves, but they can teach the related material in the text book.

Figure 3.4 how teachers use the system

And for organization, they can build up the ability based on which course they want the student to have. So they can get what they want and student can get what they should learn, too.
CHAPTER 4

System Function Using Data Mining

4.1 Overview

This section describes the operational aspects of the system as well as the associated algorithms. Since there is a machine learning component to the system, data mining is also described in terms of its relational to the system.

4.2 Classes of Users

The Exam-based Educational System (EES) has the following classes of users:

- Students
- Teachers
- Companies/Organizations

For students, the system offers the following uses:

- Search ability or course
- Find courses directly
• Find courses through ability
• Register for course
• Take exams
• Get grade
• Get related ability

For teachers the system offers the following uses:

• Upload course material
• Put elements together to form a regular course.

For the organizations the system offers the following uses:

• Set up an exam for regular course
• Set up ability.

4.3 EES Database

The data in the EES is represented by a relational database.

Figure 5.1 shows the database conceptual design and Figure 5.2 shows the database schema.
Figure 4.1 Conceptual Model of the EES database
User

| User_id | Username | Password | profile | Student | teacher | organization |

Element

| Element_id | Element_name | profile | detail | X | y | z | c | value |

Course

| Course_id | Course_name | profile |

Ability

| Ability_id | Name | describe | builder_id |

Learn

| User_id | Element_id | Grade | detail | evaluate |

Teach

| User_id | Element_id | Good_num | Bad_num |

Hold_exam

| User_id | Course_id | time | location | detail |

Take

| User_id | Course_id | Grade |

Required

| Course_id | Ability_id | grade |

Make_up

| Course_id | Element_id |

Path

| Previous_element | Element_id | Next_element |

Figure 4.2 EES Database Schema
4.4 What are the needs of the user?

It is desirable that students are able to find the best courses in the EES easily. Teachers need a way to increase enrollment in the courses they teach. For organizations, the EES should allow them to determine how reliable an ability is as a predictor if future success. Therefore, we need to add a machine learning aspect to the EES.

The EES must match the requirements of each class of users, but it is often hard to discern. We will use data mining to come up with information that best matches the requirements of the users. For students, we can group courses so that the student can easily find the course they want. A good search engine is important as well. Using data mining, we can collect student opinions of each course element and put each course element in its best position.

The EES should gather information from all the users, so that students will know which course is better. The EES should also recommend the best course for students. This entails gathering enough information so as to find the relationship between courses. For instance, a database course might be recommended to a student who just completed a JAVA course.

The EES should also allow teachers to attract more students for courses they teach. This also makes the EES more viable as the amount of data increases.
We should determine why students want to take a course. For instance, they may take a course to gain more knowledge and increase their skill set. Organizations can evaluate a course and/or provide feedback to the teacher. The system will gather all the information and give a rank for each teacher and course. So if students want to work at a specific company, they can use the ranking provided by this company. In some cases, teachers may also change their course to tailor it for a specific company.

To some degree, all students want to gain knowledge. The EES should allow students to accomplish this easier; Companies can provide an exam to determine if the student has actually gained the ability. The EES can also provide feedback in order to make exams more accurate.

We now discuss how to achieve these goals by using data mining,

### 4.5 Data mining in the EES

The basic idea using data mining is to give every element a location such that it will be accessible in a manner that provides each class of user some useful information.

This process is similar to constructing buildings in a city. Each building is an element. Learning becomes the process to find the path. When we build the city, we put two building close to one another if they are similar. In this way, the user can start from any location and try to find buildings that are similar to where they are.
Here is the sequence of steps to perform this mining:

1) **Pick all the elements from database. In this level there are no coordinates.**

   That is, all the x, y, z values are 0.

2) **Evaluate the course. We can use multiple ways to evaluate course. For example, if there is a lot of people choose the course, the course has higher value. This step is used to decrease the number of course that are being modify.**

   We can skip this step if there are not too many elements in the database.

   In the system, every user can upload elements. Some of the elements may not have any users. Few student will learn it, few organization will concern about it. So we do not need to locate them because it is not necessary to locate this kind of elements.

   We can evaluate elements by multiple ways. For example, we can calculate how many students have learnt this element, and put the number of student as the element’s value. The course that has the higher value, in another word, the value number is a larger number, has more possibility to be consider. We can also use other way to evaluate the elements, but that is not the main part of this thesis.

3) **Find the similarity between every two elements. We can use multiple ways to get the similarity of each elements. We choose the method below.**
In this case elements and student id are already sorted.

Each element is followed by student id(s).

\[ \text{Element 1} \ \ 2 \ 4 \ 5 \ 6 \ 7 \ 8 \]

The element id is 1, and the students with id 2, 4, 5, 6, 7, 8 have learned it.

We use merge to sort two different elements which have same students.

**Algorithm for get similarity:**

Let M students choose course 1, N student choose course 2

Let C1[M] and C2[N] to be the two array that kept the student who chosen element 1 and element 2.

Let s be the numbers of student that has choose both course 1 and course 2.

\[
i = 1, \ j = 1 \\
\text{while } (i < m \&\& j < n) \\
\{ \\
\quad \text{if } (c1[i] == c2[j]) \\
\quad \quad s++; \\
\quad \text{else if } (c1[i] > c2[j]) \\
\quad \quad j++; \\
\quad \text{else} \\
\quad \quad i++; \\
\} \ \text{(reference?)}
\]
Ex:

Element 1: 2, 3, 4, 7, 9, 10, 12

Element 2: 1, 2, 3, 4, 5, 6, 7, 8.

First comparing first item,

If they are matched, we pick it up. If not we move the small student id in the one list.

1 is small than 2, and we move the pointer to second item in the element 2, and compare those items.
2 and 2 are equal; we pick it up, and move both pointers to the next items.

And continue to do those steps; finally we will get the same students in both elements.
For example: Element 1 and 2: 2, 3, 4, 7

Since we want to find the elements that are similar, we should first develop a method to determine that two elements are similar. We can assume that if two elements have the same students, then they are similar. Ex: (format is: element id [student ids])

1 \[1, 3, 4, 5, 7, 9\]
2 \[2, 4, 6, 8, 10, 11\]
3 \[1, 3, 5, 6, 7, 9\]

In this example, we determine that element 1 and element 3 are more similar, with similarity rate of about 88% (same students/total students). For 1 and 2, the similarity rates are lower (17%).

Figure 4.1 shows a representation where the nodes are elements and the edges represent the similarity values between the two nodes.
Figure 4.1 similarity measures between nodes

So we can see that all the similarity is in the graphic. Next we find the node that is a centroid for a group of elements. We choose the element that has the highest value as the centroid. In this case, the highest value means this element is learnt by the most students, as we mentioned before. We call it node C.

We have calculated the similarity for each nodes in previous steps. If we extract all the nodes related to node C and put them in order of the similarity with node C, we can find the first four nodes in this sequence are the four who are closest to C. We get a tree structure like Figure 4.2. C means node C, and 1, 2, 3, 4 is the nodes that have the highest similarity to node C.

![Figure 4.2 Tree to locate elements](image)

4) For each leaf node, we can find the two closest nodes. This will extend our tree and is shown in figure 4.3. We can continue to extend the tree in a similar manner. Each time we add one node to the tree, we will not add
it again. So every node can only be in the tree once. At last we cannot find any element that can attend to the tree, because all the nodes that has the possibility to add into this tree is in the tree. Then we can find another centroid, which has the greatest value in the remaining nodes (the nodes haven’t been put in the tree), and build another tree. In the end, we will have many trees.

5) Now we determine the x, y values for each element. The x and y means the location of the nodes in two dimensions. First, we give node C the values (0, 0) as x, y, so that it is located in the center of the map. Then we put nodes 1,2,3,4 around it With the following values: Node 1 (1,0), node 2 (-1,0), node 3 (0,1), node 4( 0,-1). At this point the map looks like Figure 4.4.
6) We traverse the tree and continue putting nodes on the map. This is represented in Figure 4.5.
7) We continue in this manner until all nodes in the tree are put on the map. Notice that each location can locate multiple elements. It means that we can put the node closest to node 21 in the same location as node 2. So we do not have to find if a location is occupied. Note that this cannot happen in the real world. In the real world we cannot put two building in exactly the same location. However, the nodes are not the same as the building in real world, so we can do this in two dimension just for convenience.
8) Next we put all trees in their associated maps. The final map can be represented as shown in Figure 4.6.

Figure 4.7 Final Mapping of All Elements

Here is the algorithm for the previous step. If we have built the tree, we can use the tree to get the map.

Function Location (node T)

{
    n = T->numOfChildren;
    if(n>=1)
    {
        T->child1->x = T->x + 1;
        T->child1->y = T->y + 1;
        Loction (T->child1);
    }
}
If (n>=2)
{
    T->child2->x = T->x + 1;
    T->child2->y = T->y - 1;
    Location (T->child2);
}

If (n>=3)
{
    T->child3->x = T->x - 1;
    T->child3->y = T->y + 1;
    Location (T->child3);
}

If (n>=4)
{
    T->child4->x = T->x - 1;
    T->child4->y = T->y - 1;
    Location (T->child4);
}

9) We now use a method called Page Rank to further determine if an element is in its right place in the graph. We rank each node based on user feedback. When he map is completely built, users can use it to find elements. We can determine if an element is in its correct position by seeing how many students choose this element using the map. If few (the threshold will be set depend on specific situation, if there are fewer nods, the threshold can be set lower, and if there are more nodes, the threshold will be set higher. The purpose for this calculation is relocate the elements ) students choose this element through the map, we call it a poorly matched element. Otherwise it is well matched element.

10) If a course is poorly matched, there are two possible reasons for this. Either the element is in wrong place or it is not good enough. We can use a genetic algorithm to properly place these elements.
11) First, we choose the 20% most poorly matched elements. Then we copy them 10 times, putting them in 10 random locations. Then we check to see if it is better matched. This is shown in Figure 4.7.

Figure 4.7 An Element That is Poorly Matched

Let node 1 to be one of the poorly matched elements. We then copy this node in ten different places in the graph. This is shown in Figure 4.8.
We then gather users input again. We delete 80% of node1 occurrences because they are not in the good place. 80% is just a number to get better location. We can set it to other percentage if we want. If a node performs better, then it will have less possibility to be deleted. At this point we have about 2 labeled nodes. We copy them to different locations again. We do this until we finally find the right place for node 1 (this process will accomplish when users use the system and may be cost a long time). If a node did not improve, then it means the element itself is not so good. So we lower the value of this node. Next time they won’t have the chance to go to a better location through application of our genetic algorithm.
12) We now have the value of \((x, y, c, \text{value})\) for each element. Now we want to get the value of \(z\). In another word, we want to extend the map to a three-dimensional space. We know that when students take elements, they take them one by one. For example, we can take elements in order of 1, 2, 3, 4 or in order of 2, 1, 3, 4. As elements are taken, they increase in difficulty.

13) When a student takes elements 1, 2, 3, 4, we make the difficulty of element 1 +1, element 2 +2, element 3 +4, elements 4 +8. After we calculate this for all students, we now have the difficulty of each element. For each element, the value of \(z\) becomes this difficulty measurement.

14) Teachers can build courses from these elements. They will give each course an index that can be used to put all courses in a catalogue.

15) We can adjust the map again when we know all the elements' content. For example, element 1: 90% computer science, 10% art. We can cluster again by exchange the positions of elements. We have built a map by the previous steps. First we should put the nodes as the same major (ex: computer science > 70%), so we get a cluster of all the computer science major. Then we can find the center of the cluster, and there must be one element that is furthest from the center, we called it node F. We then find the center without node F, and find the node T who is nearest to the center with less computer science content. (Example: computer science < 30), we exchange the location of T and F, until there is no this kind of T in a specific distance from the central. Now we find that all the nodes near
the central have sufficient computer science content. In another word, we have finished cluster the nodes.

16) Building a course is accomplished by finding a path on the map. When a particular path is taken by many/most students, we can make all the node in the path as a course sequence and move the sequence to the course level. After that, students need only find a particular course sequence. See Figure 4.9 and figure 4.10 as examples of paths and course sequences.

Figure 4.9 element Sequence
Figure 4.10 Course Sequence
REFERENCES

[1] Data Mining Concepts and Techniques, by Jiawei Han and Micheline Kamber, p5-7.

[2] Efficient and Effective Clustering Methods for Spatial Data Mining (1994) by Raymond T. Ng, Jiawei Han


[5] Data Mining: An Overview from Database Perspective (1996) by Ming-syan Chen, Jiawei Han, Philip S. Yu, Ibm T. J, Watson Res


