Scientific Research and Essays Vol. 5 (21), pp. 3213-3221, 4 November, 2010
Available online at http://www.academicjournals.org/SRE
ISSN 1992-2248 ©2010 Academic Journals

Review

Digital mining applications in distance education environments

Tuncay Sevindik* and Necmi Demirkeser

Yildiz Technical University, Department of Computer and Instructional Tech. Educ., Istanbul, Turkey.

Accepted 12 July, 2010

It is seen that distance education environments are more frequently used due to the progress of internet today. The data acquired from these environments are commonly irregular journal records and text-based communication messages. Digital mining applications are used in order to convert these irregular data into regular, and cleaning out the information stacks. The purpose of this research is to reveal the usage of the types of digital mining application, which are being applied within the distance education environments. For this purpose, a situation-based study was carried out which was related to the subjects of data mining, web mining and text mining, being used in the distance education environments. As a result of the research, a relationship was determined between the types of digital mining.

Key words: Text mining, data mining, web mining.

INTRODUCTION

Together with the rapid progress of informatics technology today, web-based educational environments are now being used in distance education. One of the previous reasons for this situation is as a result of the developments of internet and communication technologies. Worldwide communication became stronger and educational institutions became closer to the learners in the light of these developments.

One of the most significant problems in education is that there are no adequate professional instructors and the existing instructors do train a restricted population. With the help of distance education approaches, these restrictions are removed and professional instructors are allowed to access all around the world. Although distance education is not yet widespread to the extent of competing with face-to-face education, an alternative is always necessary due to the restrictions of face-to-face education. At this point, distance education shall always be the first alternative to be preferred, since it can reach wider populations and can be used without the restrictions of time and place (Kaya et al., 2004).

While a unilateral communication was provided in the former distance education types (Letter, TV, CD etc.), the environments where students participate in and constantly interact with each other within the internet-based distance education environments, being used today, provide the opportunity of preparing the applications. It is an undeniable truth to tell that distance education would gain more significance together with the further development of the future technology platform (basement). Besides, it might be thought that visual and audial applications which are currently used with restricted opportunities would be advanced further with the acceleration of communication technologies. Distance education has a wide usage in various areas, providing an in-service education not only for students, but also for grown ups and careerists (Antalyalı, 2004).

The data, presented to students within the internet-based distance education environments which are used today, could have visual and audial features such as texts, videos and animations. It could be claimed that students' participation particularly in asynchronous education environments is text-based; e-mail, bulletin board and chat rooms, all these are the primary communication systems (Gürol and Sevindik, 2004). While the examination of the text-based feedbacks obtained from the students is not a problem to small populations, processing these data in greater communities causes some challenges. At this point, text mining, which is one of the data mining types and which constitutes

*Corresponding author. E-mail: tsevindik@gmail.com.

the core of the digital mining could step in and smooth the assessment phase. Open-ended questions which would be used as an alternative, especially for the classical assessment can only be assessed with text mining. In such cases, where the participants of the distance education are expressed with thousands and even hundreds of thousands rather than hundreds, it is assumed that application of text mining is obligatory.

DISTANCE EDUCATION

Distance education is an educational method, with roots going back to 150 years. The possibility of distance education was mentioned in a newspaper of Switzerland in 1833, for the first time (Schlosser and Anderson, 1994). Distance education which was previously applied in environments such as letter and radio transitivized into tape and book later on; and together with the spread of television is being used as television-based. During the periods in which there was an increase in computer usage, it was carried to environments such as disk and CD, and together with the development of internet, it was used within internet-based platforms. It is seen that distance education institutions have always been attentive to use the latest technology of the age.

Various distance education models are used in the world and various changes are made in order to develop them. With the development of vehicles such as radio, television after 1955, the opportunities of distance education increased as well. With the emergence of the satellite technology, education among different countries was enabled. "Open University" which is depicted as an example and which was founded in England in 1968, provided education for many countries (Daş and Varol, 2001).

The advantages of distance education are as follows (İşman, 2008 -a): “It gives the opportunity to individuals for self learning, is more flexible compared to conventional learning and is applicable to individual conditions. Internet, which is the utmost of today’s technology, is an environment which is accessed by a great number of population worldwide and it is open to everyone. According to the data of the World Internet Usage Statistics and World Population Stats (2009) site, the number of internet users in the world in September 30, 2009 was determined to be approximately 1.7 billion (1.733.993.741). Considering the fact that the world population is above 6 billion (6.767.805.208), the number of internet users has a rate of 25%. Using a technology which is easily accessed by such a great population for education is inevitable and necessary.

Assessment in the distance education

Assessment phase is one of the important basic parts in any program development system. Since the success of the distance education depends on the general performance of the institution, the assessment phase should be carefully designed. In the distance education, program assessment is also one of the subjects that affects the success of the institution, as well as the assessment of the students. In the program assessment, goals to be reached should be determined clearly and during the assessment phase, deficiencies should be analysed through an objective approach. The state of reaching the goal and the pleasure state of the student and instructive staff should also be assessed (Monolescu et al., 2004).

Like in other educational branches, distance education progresses towards a student-based structure. Role of the instructors changes more in the direction of proper learning and in addition to this, many institutions made changes which is directed following the attendance status of students and their study phases. Although the facilities of technology change, the necessity of a teacher never disappears; instead, its role changes. A teacher reaches the position of a contact between students and knowledge. And it became obligatory for teachers to use technology (Beaudoin, 1990).

In public credit, the success level of the distance education is measured by the achievement of following the students. To what extent the student is successful at lessons, how well they understand the lessons, how much they improve themselves, and whether the contents are efficient or not could be measured by follow-up tools. The student should be followed very carefully and their problems regarding certain issuies will be determined. The determined deficiencies should be removed and new techniques willl be developed.

Since computer technology is open to all kinds of innovations, using every opportunity of this technology will be the most convenient approach in distance education. A sharing and communication environment should be developed in order to show that distance education is more different from open learning and to make the student be under a constant inspection. By this way, it should have been seen that distance education is not much different from the formal education. During the assessment phase of the distance education, we can watch the student’s behaviour by keeping logs. By this way, we can reach information such as how often and on what time intervals the students visit a web page and which pages they have access to, etc. In such cases, digital mining applications will be useful in overcoming the problem, during the struggle with information stacks. The concept of digital mining separates into three parts within: data mining, web mining and text mining.

DATA MINING

Data mining is a data processing method with a very broad area of usage. The constant increase of the information in the developing world increased the use of data mining as well. Particularly due to the market competition

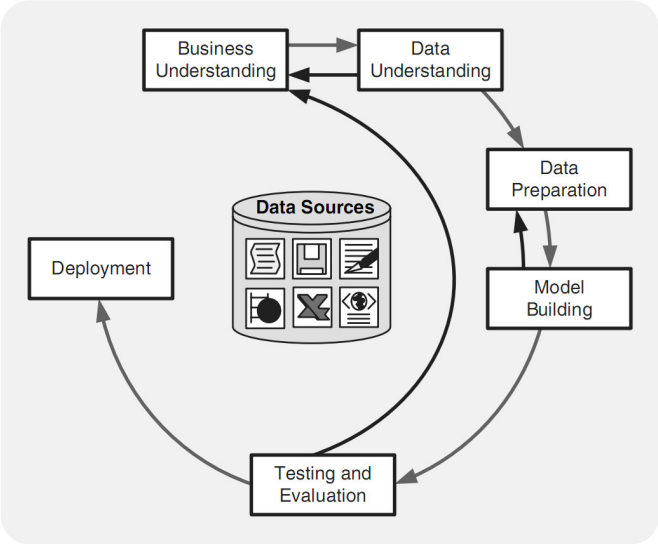

Figure 1. Phases of data mining (Olson and Delen, 2008).

in various sectors, data mining became an essential tool in following the customers and determining the product demands (Berry and Linoff, 2004).

Data mining aids the analysis of a very large scale data, stored in the computer. For instance, the sale list of products, being sold in the market, could be considered as such kind of data. Since barcode system is used in sales, a large number of products are sold in a very short time and they are priced rapidly. Since follow-ups are made with computer system, new products could be supplied by reaching to the manufacturer before any product sells out. It is only the barcode system that enables us to obtain much information easily and to store them. Data mining is not only used as a marketing instrument, but is heavily used during election periods, at the preview phase of the electors. In the medicine sector, data mining has long been used for following the patients, processing high amounts of patient information and determination of the effect of the applications. The effect of various drugs on the same kind of illness could be compared with data mining. In the banking sector, data mining is used for following the credit card systems and prevention of knavery. And lastly, in the producing companies, data mining is used for quality control (Olson and Delen, 2008).

Phases of data mining

According to the Figure 1, which is determined by Olson and Delen (2008), the phases of data mining are determined as thus explained.

Examination of the application area

This phase consists of the examination of the area on which data mining is applied and determination of the goals, making the situation assessment and charting the project plan.

Examination of the data

After the determination of the goals and plan, the phase of determining the situation of the available data and determination of data types comes.

Preparation of the data

After the data resources are determined, the information is chosen and utilized by getting rid of the parasites. The operations of data cleaning and transformation are carried out in this phase. Besides, the information is deeply examined and additional models are determined in this phase as well.

Developing model

Software tools and grouping tools used for processing data (programs of determining relationship and printing tools) are convenient for inner analysis. The rules that show the relationships of the data types and connections inbetween the information are supposed to be determined. Once the characteristics of the data are efficiently comprehended, thus more developed designment types could be produced. It is in this phase that data separates into two groups as test and control groups.

Test and assessment

After the determination of the necessities of application area during the first phase, the modelling results are assessed due to that. This phase could enable the determination of other necessities. The phase of test and assessment is in a close relationship with the examination of the application area. According to the assessment results received, the first phase could be rechecked.

Application

Data mining could be used in order to determine the accuracy of a predetermined hyphothesis or to explore a non-explicit information. A non-expected relationship could be revealed as a result of the data analysis.

These six phases do not have to be applied, for certain. According to the situation, these could be reduced or some of them could be separated into sections. New sections could be added depending on the demands.

Data types used in data mining

No matter the difference it makes, the data source is definitely supposed to be included in one of the three type of data mining: Demographic, behavioural and psychographic. Every data type has its own weak and strong sides (Rud, 2001).

Demographic data

Demographic data are generally the types of data that are hard to change, such as personal characteristics and present situation. Characteristic features could be ranked in this group as well; Age, gender, marital status, monthly income, being landlord, educational level, ethnic origin, number of children. Stronger aspects of demographic data are considerable. It becomes a good reference point for future estimations since it does not change easily. Especially educational level, being landlord, marital status is among features that do not change frequently. However, the data such as bank deposit, political view could change very quickly. Compared to other data types, it is easy to obtain and does not have much cost. On the other hand, it has such a feeble aspect that many people may not want to share such kind of data and it is seen that they give wrong information in some cases.

Behavioral data

These data are acquired through measuring a particular process and behaviour. Among the data types, it is by far the most preferred one. According to the characteristics of the area to be used, data such as amounts of the products which are sold, dates of sale, dates of payment, process of the customer services could be entered within these information. As well as purchase following, clicking rates and link following status could also be taken into consideration on the web page. Processing the behavioral information gives better results in estimating prudential behaviours compared to other data types. It is a data type which is hard to get from external sources.

Psychographic data

Data of this type are information acquired from the views of the users. Personal assessments and criticisms could be counted within this type. Sampling according to the market data as the previous ones, customer attitude surveys, requests and complaints could be counted among them. Attitudinal data bring a different dimension to the analysis in the estimation models. Product development and purchase directed at special situations such as marriage and retirement could be estimated much better with results to be obtained from such data. The greatest deficiency of this data is that the specified behaviour does not reflect the real behaviour.

Algorithms used in data mining

Methods of data mining are considerable in number and the following methods are frequently used: C45, K-Means, SVM: Support Vector Machines, Apriori, EM, PageRank, AdaBoost, kNN: k-Nearest Neighbors, Naive Bayes, CART: Classification and Regression Trees

Types of data mining

Data mining takes various names according to the areas it is used in. Considering as distance education, web mining and one of its components, text mining, comes into prominence. While data mining searches for large scale information, text mining is similarly used in searching for patterns within a great quantity of texts. Significant information is revealed through examining the text-like data. Text-based data are found to be unconfigured according to the general data type and they are hard to process. Besides, writing is the most frequently used form in information interchange and transmission in the modern culture. It does not matter how successful the text mining is, there is a chance of using only a certain part of the data and the result is acquired as classified. Although the data acquired in the data mining are configured, the results that are sought are not always clearly determined and it is possible to get unlooked results. In the text mining, the sought information is clearly seen within the text; it has no hidden parts. However, many writers have a trouble with expressing themselves precisely while preparing an article and are scared of being misunderstood. As well as classifying the information, text mining also enables a person to make sense without reading the whole text. Text mining could be used within various areas with different application methods (Dolgun, Özdemir and Doruk, 2009).

Another type of text mining is data mining which is based on the inference of Meta information. Meta is a type of information that summarizes the information. Writer, title, subject classification and keywords could be counted within this group. Meta data consist of data which are configured at high extent. Meta is generally identified with the term “existence” in the world. Many documents are full of such data, for example, telephone numbers, fax numbers, street addresses, e-mail addresses, e-mail signatures, abstracts, content list, reference list, pictures, titles, announcements, internet addresses etc.

Many short passages give information about a particular subject and object. These sections come together and constitute higher-level documents. Identifying the structure in the composition and determining the

relationship between the singular parts is a method called “information inference”. After the existence is determined, the text separates into parts and the relations between them are determined.

The information, obtained from internet pages is a kind of text data. However, due to special definitions used during the scripting of the internet pages, it differentiates from prose at a certain rate. While some signs show the sections within the document, a connection is made between the link structure and different documents. Such characteristics enable obtaining more useful data compared to text mining. However, there is no automatic method to extract data from these sources. Using various softwares, the requested sections on the page could be extracted and separated into parts. HTML tags within the text could be used for classification, for instance: sections, emphasized with <p> (passage limits) or <li> (list elements) or bold fonts within the page could be taken separately (Witten and Frank, 2005).

Usage areas of data mining

Data mining has become an analysis method, being applied in many sectors and various disciplines due to the large scale of data, encountered as a result of the developing technology (Çaliş, 2007). Arrangements of some of its usage areas are thus explained.

Applications in the business area

A company could apply to data mining in order to prevent its customers from heading towards its opponents. Here, the purpose of data mining is to obtain the characteristics of the customers and determine the customers who would head towards the opponent companies. And then a strategy might be developed in order not to lose them, through benefiting from the results of data mining. This determines the reasons why products and services are preferred and which characteristics influence the customers and to what extent they influence them. It also examines the credit situations and payments of the bank customers, determines the precarious customers among them and estimates other customers who would come up to the same risk group.

Applications in the medicine area

This entails being able to make predictions of the effect a drug would have on patients, and the age group of the patients. It aslo determines the most suitable candidate for the recently-found cancer treatment method.

Applications in the sports area

This involves making predictions of the future performance of a player who would be received within a basketball team and the services he would provide for the team.

Applications in the librarianship area

This involves making predictions of when a customer would turn the book and which book he would choose the next time.

Applications in the tourism area

This is arranging of advertising campaign by determining why the tourists come to a region and increasing the number of the tourists in the season.

Applications in the education area

This is separating the students into branches according to their performances and increasing the harmony of person-branch.

Applications in the web area

This determines the profiles of the users and publishes the advertising campaigns of products, which are convenient for them on the page.

Especially in recent years, data mining in the health area has been used in various dimensions. They generally enable acquiring important results as a consequence of keeping the patient records and effects of the drugs as data. The following results could be obtained through examining the data in the health area:

1. Estimation of common traits of people having a particular disease
2. Estimation of the conditions of patients after the medical treatment
3. Estimation of the hospital costs
4. Estimation of death rates and epidemic illnesses

The studies of data mining, concerning the management of diseases, include researches such as identifying the diseases and their conditions and modelling the expenditures. The purpose in these studies is to attain positive results. It is possible to previse a possible disease in the future through classifying the similar characteristics, seen in patients. For example, using age, gender and some symptoms, the patients who carry the possibility of cancer in the future could be determined. Besides, various data could be accessed through using the departments visited by patients and demographic information. Another examination area is applied on examining the hospitalization periods which constitute the greatest cost in terms of the hospitilized patients.

Accordingly, it was determined that while the patients in the cities and patients with any addiction are hospitilized more, compared to the ones in the rural areas, the hospitalization period in public hospitals is longer than in private hospitals. Since there is the possibility of irregularity within the health sector, many insurance companies choose the way to determine the people who are constantly hospitalized at various hospitals concerning this subject by the method of data mining. Extraordinary data attract attention in such analysis and by this way, the irregular people could be determined easily. Another important area in which data analysis is used in the medicine area is related to drugs. In case of side effects and negative results got at different rates according to the laboratory testings as a result of the follow-up of a recently-released drug through data analysis, the drug could be pulled from the market (Yıldırım et al., 2008).

It is seen that a great majority of the data in medicine and health areas were kept within irregular texts. For example, medical conditions of the patients, diagnosis, treatment information and clinical documents are stored as texts. Besides, the bills of the applied proceedings and reports documenting the workflow are in the form of text. Scientific articles within the medicine area are also valuable information sources for researches and innovations made in the health area and they are kept within textual structures. Methods of text mining are used in order to make knowledge detection on these data.

WEB MINING

Web mining is a type of data mining which is applied on the internet pages. Using the web mining, the behaviours of the visitors could be watched on the internet pages and solution offers could be developed together with the determination of the periods when the pages are busy. Interpreting the extraordinary behaviours which are determined as a result of the analysis, possible attacks could be detected and then precautions might be taken according to the methods of attack.

Making the convenience test for the rules of W3C (World Wide Web Consortium) which determines the standards especially for new technologies to be used on the web pages, precautions could be taken for displaying the web pages on various browsers smoothly. By this way, the sites should have been kept updated (Daş et al., 2007).

Through web usage mining, which is another dimension of web mining, the behaviours of the visitors who visit the web pages could be analized. The architecture of the web page might be reshaped according to the pages visited by visitors and frequencies of transition between these pages. The links of pages which would frequently be visited by the people who visit a particular page could be made easily-seen. Besides, advice pages could be displayed by presuming the pages which would possibly be visited by a visitor who comes to the web page in order to increase the lexical dimension right after visiting the main page (Daş and Türkoğlu, 2009).

Data mining in education

Together with technological developments and emergence of new requirement areas, education was freed from being restricted within school and it acquired brand new dimensions through methods, called “Continuous Education” and “Lifelong Education”. Rapid progress of internet technologies and spreading into every section of the society worldwide led to the production of internet-based softwares and together with the fall of technology prices, the computers that would use these softwares were able to assess every section of the society. Today, the opportunity of transmitting audio, video and other data at large scales increased to acceptable levels together with the extension on the internet networks (Oral, 2004).

In the distance education, it became possible to assess students by new methods and analyse the study activities of students through registering to a database. Two different methods are used in determining the student ‘s behaviours. The first one is the application of the survey. The reliability of this method depends on the sincerity of the answers of the survey questions to be given by students. The second method is the examination of the entrance registrations. The entrance is kept within the tables on the database of daily registrations. As a result of following the registrations, large amounts of data stacks - to be used by researchers – are constituted. The proceeding to be applied to this information increases in the proportion of its capacity. While registrations may not look quite significant as a result of tabulating and record keeping techniques, being designed, some restricted meanings could be made when a person wants to interpret. What matters for the analysis is not the abundance of the registration number, but registration types and varying behaviours of the users. Different approaches will be used in order to make convenient analysis.

At the end of the training, a real assessment should have been made through revealing either the achievement of the system, of the student, or the assessment of student’s achievement status. This assessment also helps to determine a student’s right for a degree, certification or certificate of achievement at the same time. Assessment of achievement status will become an indicator of whether the prior condition is performed in the training program, during the subsequent phases. Data mining could be used in analysing the information on the database in which the students’ behaviours are saved within the web-based Distance-Learning. Some models of data mining make automatic inferences through the

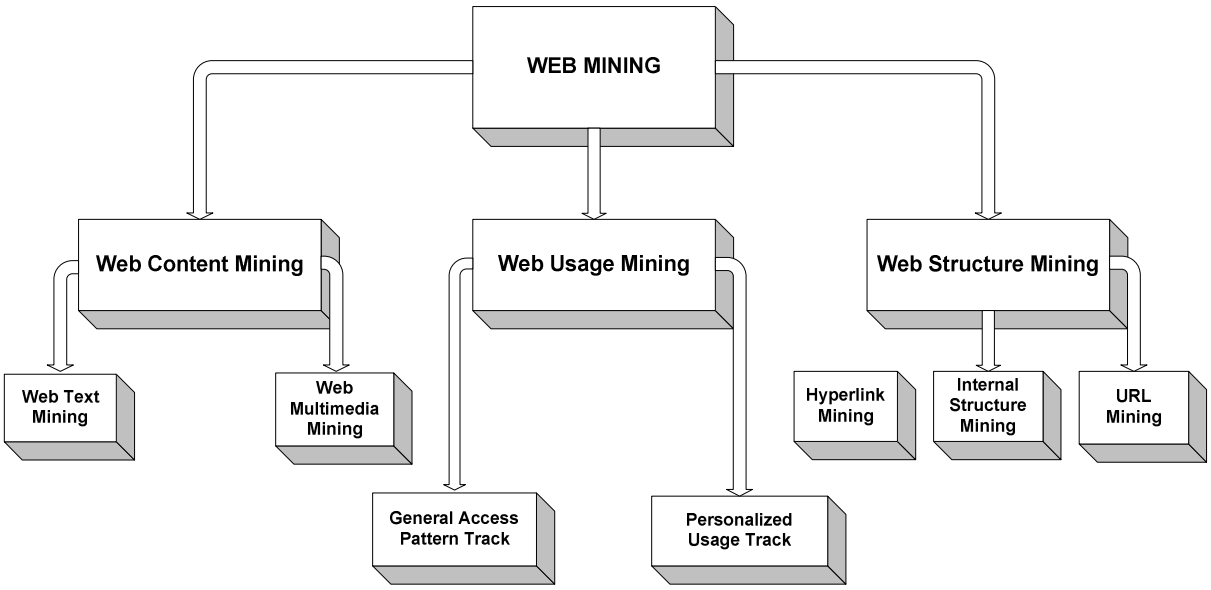

Figure 2. Classification of web mining (Qu et al., 2009).

data. However, the targets shoud definitely be specified in order to provide data for these models and assess the results (Savaş and Arıcı, 2009).

Web mining, used on the distance education platform, could be analysed in two sections:

1. Data analysis directed at analysing the sources
2. Data mining based on the user registrations

Lesson materials, lesson plans, exercises, education videos, written documents and other educational tools within the type of web mining, which is obtained through analysing the sources, are used as data sources. After cleaning the noisy sections within the data, data are prepared by processes such as tagging and automatic defining of passage. Using the phases of pattern recognition and analysis, relations between the data and any kind of unexpected results are reduced.

At the phase of web mining, which is based on the user registrations, process follow-up journals, which are revealed with the interaction of users, are primarily used. Using these information, the habits of the users and what kind of pages are most frequently used by what sort of users could be determined. Then, these data are transformed into a pattern. The connections between the data are determined.

The information acquired from the habits of the users and content analysis could later be used in showing recommendations for users with similar characteristics. The following subjects come into prominence at this phase: the subject that would attract the user's attention might be used in suggesting pages having similar characteristics with the content of the training, being given (Qu et al., 2009). The diagram showing the classification of Web mining can be seen in Figure 2.

TEXT MINING

Text mining could be defined as a type of data mining based on a large scale of information, saved by the user during interaction. Being similar with the data mining, text mining aims to find the interesting patterns hidden within the information. Besides, data sources are found as document archives. Being different from data mining, information is found not as regular but as a complicated text (Poteet and Kao, 2007).

Text mining uses a part of the methods, developed by data mining. Therefore text mining and data mining show many similar characteristics. For example, both systems have a phase of pretreating the data; they both have their own pattern recognition algorithms, presentation layer elements and visualisation tools. Moreover, text mining inherited some methods which were previously explored in the data mining as well.

Since data are accepted to be gathered as configured in the data mining, the large part of the proceeding part focuses on two phases; normalizing and correcting the data and producing large scale table connections. On the other hand, text mining focuses on the phases of defining the data and inference more. These pretreat phases are consisted of phases in which unstructured data are rendered into the required format by being transformed into more regular document. Such kind of a decomposition is generally not seen in data mining systems. Moreover, since the acquired information is consisted of texts in the format of native language, text mining focuses on processing native language, which is different compared

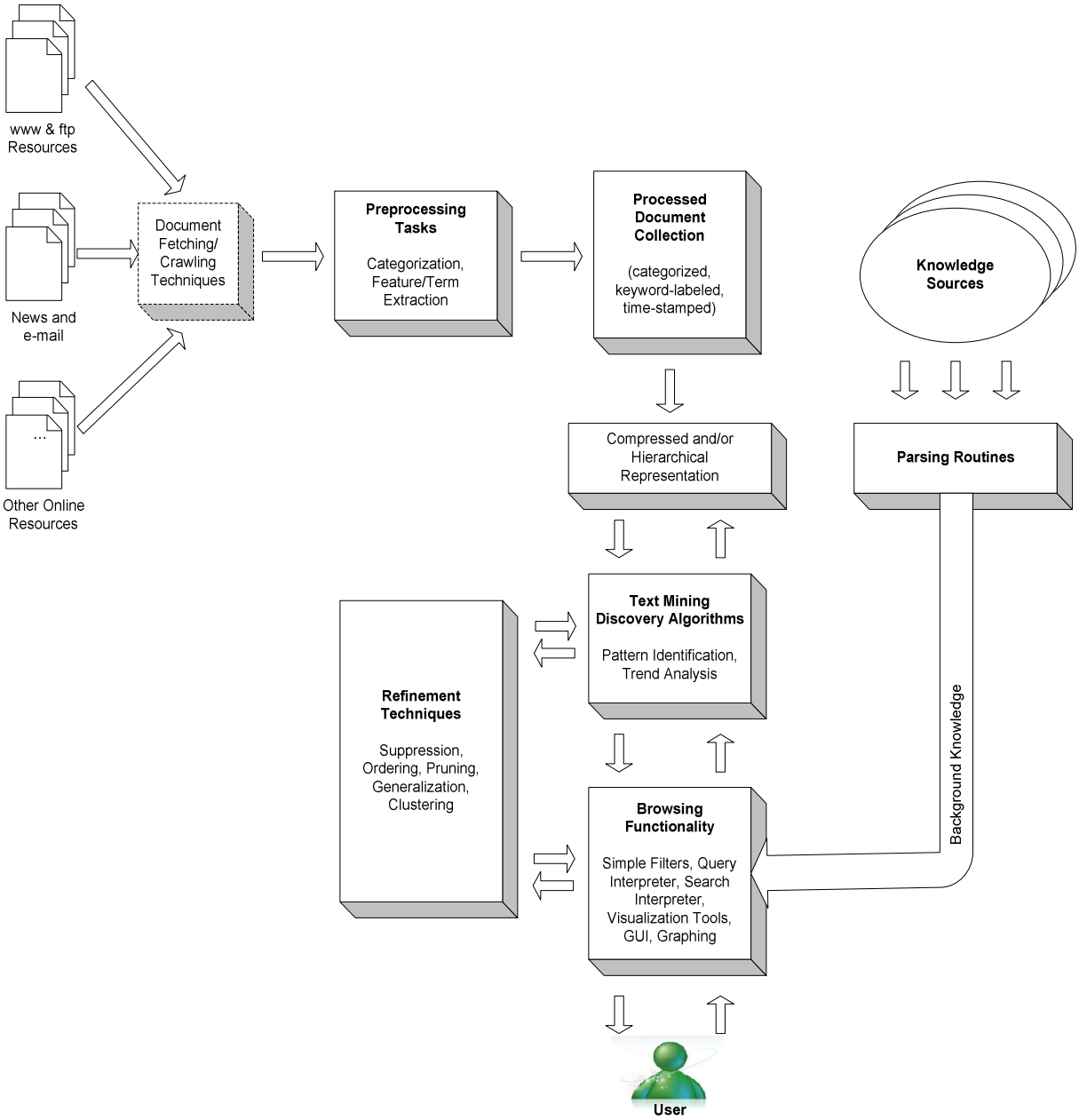

Figure 3. Phases in an advanced text mining system (Feldman and Sanger, 2007).

to other computer sciences. The most attention grabbing points concerning the text mining are probably information retrieving, information gathering and stack-based linguistic calculation processes (Feldman and Sanger, 2007). In Figure 3, the diagram of phases in an an advanced text mining system is shown.

Using text mining in education

Using data mining in education shows some relative difficulties compared to other disciplines. In the distance education systems, it is possible to determine and store the data types in accordance with the online access information and demographic records of the students. Since computers are used as interface on different online education models, it is quite easy to record the information. However, it is hard collecting the written data and transfering them to computer in the face-to-face education.

Data mining is used in various dimensions in education. It could be used in analysing the students' behaviours and making prospective estimations (El-Halees, 2008). It is possible to use the web mining in analizing the behaviours

of students by benefiting from entrance records of students into online education systems (Merceron and Yacef, 2005). Using the data mining in higher education, the students to be graduated could accurately be determined at the chance of 85%. By this way, students who are in critical conditions could be given academical support (Luan, 2006). When students drop out of school, it becomes an undesired situation since it makes a bad impression for universities.

Universities with a high rate of failing should come up with a decline in tuitions and lowness of morale.

Besides, compared to the graduates, the unsuccessful students will have less earning in their future lives. Using the methods of data mining and constantly following the success conditions of students, various supports and solution suggestions could be presented for students taking a turn for the worse (Chong et al., 2007). Considering the students as customers in education, their satisfaction is of the primary importance for educational institutions. Therefore, using the methods of data mining, surveys directed at determining the students' satisfaction could be made and filling open-ended satisfaction forms, analysis could be performed. Analysis results can also become an indicator for the success of the educational institution (Thomas and Galambos, 2004).

CONCLUSION

In recent years, data mining has begun to be used in various areas. Especially in the century, which is referred to as "Information age" and in which we live, there is an excessive amount of information increase. In order to analyse the large scale information, data mining is supposed to be used. There are many developed data processing methods which could be used according to the type of the data we have and various tools used in the computer environment.

While analysis of the small scale information at the phase of student follow-up and assessment in the distance education which is a contemporary branch of education could be carried out by the human brain, the processing of the information stacks which are produced as a result of the participation of crowds who are supposed to be in a distance education to literally be structured in accordance with modern education models could only be possible through using computers. The information available will constantly increase through recording the behaviours of students at every phase, especially like portfolio assessment. Similarly, the processing of the open-ended questions used in face-to-face education and data acquired as a result of a survey which is applied to crowds could only be enabled by computer. The analysis of written data which are obtained as a result of open-ended questions and intercourses can be taken easily and rapidly through the methods of text mining.

REFERENCES

Antalyalı ÖL (2004). Perception Of Distance Education And Teachability Of Operations Research Course With Distance Education. Isparta: M.Sc. Thesis, p. 19.

Beaudoin M (1990). The Instructor's changing role in distance education. The Am. J. Distance Edu., 4(2): 21-29.

Berry MJ, Linoff GS (2004). Data mining techniques. Canada: Wiley Publishing, Inc., p. 468

Chong YH, Digangi S, Pennel AJ (2007). A data-mining approach to differentiate predictors of retention. Educause Southwest Conference, Austin, TX.

Çalış H (2007). Fault Detection In Induction Motors Using Data Mining. M.Sc. Thesis, p. 16.

Daş R, Türkoğlu İ (2009). Extraction Of Interesting Patterns Throught Association Rule Mining for Improvement of Web Usability. J. Elect. Elect. Eng., 9(2): 1037-1046.

Daş R, Türkoğlu İ, Poyraz M (2007). Analyzing of the user access logs of a website using web usage mining method: Example Of Firat University. e-J. New World Sci. Acad., 3(2): 310-320.

Daş R, Varol A (2001). Overview of distance education application today, The National IT-Multimedia Conference, Elazığ, pp. 53-62.

Dolgun ÖM, Özdemir TG, Doruk O (2009). Unstructured data analysis in data mining: Text and web mining, J. Stat., 2: 49-52.

El-Halees A (2008). Umm-Al Qura University: http://eref.uqu.edu.sa/files/eref2/folder6 /f158.pdf, (01.17.2010)

Feldman R, Sanger J (2007). The text mining handbook. New york: Cambridge University Press, p.17.

Gürol M, Sevindik T (2004). Dimension of distance education technology. XIII. National Science Education Congress, Malatya: İnönü University, Faculty of Education.

İşman A (2008-a). Distance Education. Ankara: Pegem Akademi, p. 114.

Kaya Z, Erden O, Çakır H, Bağırsakçı BN (2004). Web-based presentations preparation of distance education require units in course of the foundations of distance education. Turk. Online J. Edu. Technol., 3(3): 165-175.

Luan J (2006). Data mining applications in higher education. spss.com: http://www.spss.com/home_page/wp. 114, (01.17. 2010).

Merceron A, Yacef K (2005). Beuth Hochschule Für Berlin. (01 18, 2010), Berlin University: http://public.tfh-berlin.de/~merceron/ publi.html.

Monolescu D, Schifter C, Greenwood L (2004). The Distance Education Evolution: Issues and Case Studies. London: Idea Group Inc., p. 164.

Olson LD, Delen D (2008). Advanced Data Mining Techniques. Springer, p. 10.

Oral B (2004). The Internet Usage of Teachers Candidates. XIII. National Science Education Congress, Malatya: İnönü University, Faculty of Education.

Poteet SR, Kao A (2007). Natural Language Processing and Text Mining. London. Springer, p. 6.

Qu Y, Zhong L, Zou H, Wang C (2009). Research about the application of web mining in distance education platform. International Conference on Scalable Computing and Communications. Comput. Soc., pp. 508-513.

Rud OP (2001). Data Mining Cookbook. New York: Wiley Publishing.

Savaş SS, Arıcı N (2009). Investigation Of The Effects Of Two Different Instructional Models In Web-Based Distance Education On The Student Achievement. IATS'09.

Schlosser CA, Anderson ML(1994). Distance education: review of the literature. EDRS, pp. 2-5.

Thomas EH, Galambos N (2004). What Satisfies Students? Mining Student-Opinion Data with Regression and Decision Tree Analysis. Research in Higher Education, Springer, pp. 251-269.

Witten IH, Frank E (2005). Data Mining. San Francisco: Morgan Kaufmann Publishers.
World Internet Usage Statistics and World Population stats. (9 30, 2009). World Internet Usage Statistics: http://www.internetworldstats.com/stats.htm, (01.10. 2010).

Yıldırım P, Uludağ M, Görür A (2008). Data Mining in Hospital Information Systems. Acad. Infor. Congress, pp. 431-432.