

# **Data Mining for the Analysis of Content Interaction in Web-based Learning and Training Systems**

Claus Pahl

*Dublin City University, School of Computing, Dublin 9, Ireland*

## **Abstract**

Data mining is based on the detection and extraction of knowledge from a database. Web usage mining is a special form of data mining to address behaviour in Web sites. Web usage mining shall be proposed as the central, non-intrusive and objective evaluation technique for the analysis of the usage of Web-based learning and training systems and in particular of the interaction with educational content in these systems. We will introduce and illustrate a number of mining techniques that were developed specifically for the educational context. The behaviour of learners in learning and training technology systems, in particular when a variety of interactive learning and training features is offered, needs to be analysed and evaluated in order to show the effectiveness and to improve, if necessary, the instructional design. We look at techniques to determine the goals of learning sessions, the detailed interaction with content, and the changing of learning behaviour over time.

## **1 Introduction**

The use of computer-supported learning and training environments to enhance or replace traditional forms of learning and training has increased over the past years. In particular the World-Wide Web has gained the status as the predominant platform for these environments. First generation Web-based educational environments succeeded due to their advantage of easy access to educational resources. Recently, the focus has been on supporting a wider range of educational activities, thus enhancing the learning experience for the user through improved interactivity and engagement of the learner. Traditional knowledge-based learning is complemented by skills-oriented training activities.

In Web-based environments, the interaction between learners and content supporting knowledge acquisition and skills training is central [1,2]. On the technical level, interaction is a reflection of learning activities and strategies. The evaluation of learning and training behaviour needs to be based on the analysis of content interaction in these environments.

Learning behaviour in learning and training environments, however, is currently not well understood. In contrast to traditional classroom-based learning and training, the learning strategies and behaviour are more determined by the learner's own decisions how to organise learning and training [3]. Additionally, often several educational features are available at the same time, allowing competent learners to choose their own approach of combining resources and features. Consequently, the analysis and evaluation of learning and training behaviour is of central importance [4]. A general understanding of effective and preferred learning styles and behaviour is required for authors and instructional designers to improve the design of learning content. Instructors require feedback on usage to improve the delivery of Web-based educational resources. Supporting both summative and formative evaluation is therefore our aim, whereby, due to the novelty of the application, formative evaluation integrated in an incremental design and development process is the more crucial aspect.

A framework for the analysis and evaluation of learning and training behaviour and interaction in Web-based educational systems needs to support a variety of techniques:

- the detection and discovery of learning and training interaction from sources such as Web access logs,
- the explicit capture and representation of interaction behaviour abstracted from the interaction and access requests recorded in the logs,
- the analysis and interpretation of behaviour within the educational context using an analytic model.

Traditionally, direct observation and surveys are used to determine the learning behaviour in classroom-based learning and training, but with the emergence of computer-supported, and in particular Web-supported learning and training environments, there is now another option. Learners leave traces of their activities and behaviour in computer supported systems. In Web-based systems, access logs are automatically generated by Web servers that handle user requests. Data mining [5,6,7], in the application to behaviour in Web-based environments called Web usage mining, can here be deployed to make this latent knowledge explicit.

Data mining is about the discovery, extraction, and analysis of data from large databases. Data mining aims to make implicitly represented information explicit. The special form of Web usage mining aims to extract behaviour in Web sites from access logs. Web usage mining is an observation-oriented evaluation technique suitable for learning behaviour analysis [8,9,10,11]. Despite some limitations, it offers a non-intrusive form of observation that can contribute substantially to reliable and accurate evaluation results for educational applications.

Our objectives here are to introduce central Web usage mining techniques and to illustrate their benefits using a case study. While Web usage mining has been applied in various domains, we will focus here on techniques that are specific to the educational context and content interaction in particular.

## **2 Interaction and Behaviour**

### **2.1 Learning and Training Interaction**

Supporting design and formative evaluation is the main goal behind the deployment of Web usage mining. The design of learning and training systems requires a suitable methodology. An abstract model of learning and training can form part of the foundations for the instructional design that can also act as an analytic model for evaluations. In particular, the notion of interaction plays a central role in such a model.

Interaction can be characterised in different ways. A common classification is by role [12]: learner-content interaction, learner-instructor interaction, and learner-learner interaction. Learner-content interaction is at the core, according to Ohl [13]. In particular behaviour in Web-based educational environments is often an expression of learner-content interaction, as the Web has been predominantly used to provide access to learning resources. In the Web context, the notion of interaction is, however, overloaded. It has a meaning in the context of education, but also as part of the computer environment. These different views can be reconciled through an abstract interaction model [14].

### **2.2 Implementing Interaction**

Learning and training has to be mapped to, or implemented by, a learning technology system [15]. On the technology side, two aspects have to be considered in particular: the human-computer interface from a more general perspective [16] and educational media and services in particular [17] as components of the learning technology system. These two aspects add two additional layers to the learning and training interaction, with different notions of interaction in all three of them. The interaction between the learner and the system (which provides content and educational services) needs to be designed and implemented.

- The learner needs to be characterised in terms her or his behaviour at the human-computer interface, i.e. in terms of cognitive aspects, the learning goals and tasks, and linguistic aspects.
- The system needs to be described in terms of a technically oriented interaction language for the educational services implementation and media.

As we can see, the notion of interaction is central – but interpreted differently in different contexts. The interaction of a learner in educational systems can be traced on the implementation level, translated into human-computer interactions, and analysed in the context of learning and training interaction. This

characterises the layered abstract interaction model for our design and evaluation approach based on Web usage mining.

### 2.3 An Abstract Model of Content Interaction

Interaction is a term that is used in different contexts. Different meanings of interaction can be distinguished at levels of abstraction for interaction and behaviour:

- *Learning and training interaction* refers to interaction between learners, instructors, and content in the context of education.
- The *human-computer interface* aspect relates human activities to the software and interface features of the computer system.
- *Multimedia and service interaction* deals with processes and formats of interaction at the technical level.

These three layers are the essential elements of an abstract conceptual model that defines and structures notions of interaction.

In instructional design, we need to relate learning and training interaction to system-level interaction at the multimedia and services interface [18]. In the Web mining process we need to look in the other direction: system-level interaction activities extracted from the Web logs need to be related to and interpreted in the learning and training context.

The formative evaluation is our major objective. Due to the novelty of Web-based learning and training, development methodologies focusing on interaction and supported by data mining are sought. The support of incremental instructional design and its implementation is, however, only one goal. Web usage mining is a tool that allows the constant monitoring of learners in an educational technology system. For instance, the approach can be used to identify weak learners through their behaviour. These are, for instance, often characterised by erratic behaviour.

### 2.4 The Interactive Database Learning Environment IDLE

In order to illustrate the education-specific mining techniques and the benefits of these techniques for the analysis and evaluation of learning and training interaction, we refer to the *Interactive Database Learning Environment IDLE* [19]. IDLE is a support environment for a 3rd-level undergraduate computing course – an introduction to databases. IDLE can be characterised as follows in terms of its educational objectives:

- IDLE implements a virtual apprenticeship approach, considering the learner as a (virtual) apprentice who has to be trained to self-reliantly perform subject-specific tasks under the guidance of a (virtual) master.
- IDLE aims to seamlessly integrate various educational services (lectures, tutorials, labs) in one environment and to enhance the practical aspects by a realistic learning and training environment.

A virtual lecture component in IDLE is based on a synchronised delivery of audio and visual material. IDLE tutorial services make use of animation and simulation techniques in order to demonstrate practical aspects of the course

content. A lab feature aims to support the practical coursework (consisting of database development and programming) within the learning environment. An especially interesting component for Web mining applications are the tutorial and lab training services. These aim at skills training by supporting a range of active learning features. Learners interact with IDLE to learn about and train graphical database design and database programming. IDLE plays here the role of the virtual master interacting with a virtual apprentice.

The IDLE environment has been developed over a period of several years. As a result of the novelty of the technology, an incremental, prototyping-based development approach was taken. System prototypes were developed, used in the course, and evaluated; the formative evaluation results were integrated in consecutive improvement and extension steps. Web usage mining has played an essential role in this process. It allowed the determination and evaluation of learner behaviour in order to improve the system and the instructional design. In addition, the possibility to constantly monitor the class has helped the instructor in running the course.

### **3 Data and Web Usage Mining**

The aim of *data mining* is the discovery and extraction of latent knowledge from a database [5]. This knowledge is classified into rules and patterns in order to support analysis and decision making processes. Classical uses of data mining include decision support systems for the business context and analysis tools for scientific applications. It can be used in a predictive (decision making), generative (create new or improved designs), and explanatory (scientific analysis) style.

#### **3.1 Web Usage Mining in the Educational Context**

*Web usage mining* is a specific form of data mining, focusing on the analysis of user behaviour in Web-based systems, see Fig. 1. The database in the Web context is the access log created by Web servers. A Web log entry includes the requestor (or the respective URL), the requested resource, and the date and time of the request. If the content and purpose of the resource is known, we can deduce the corresponding activities from the log entries. For instance, audio files and some HTML-pages in the IDLE system support lecture participation as the activity. Web usage mining is different from data mining in general since only activities and behaviour are looked at. Web usage mining has been used widely in e-commerce application to monitor and analyse shopping behaviour [5,6,7]. Customer relationship management is the ultimate objective in the e-commerce context, which is achieved through optimisation and personalisation of the shopping processes based on usage mining results.

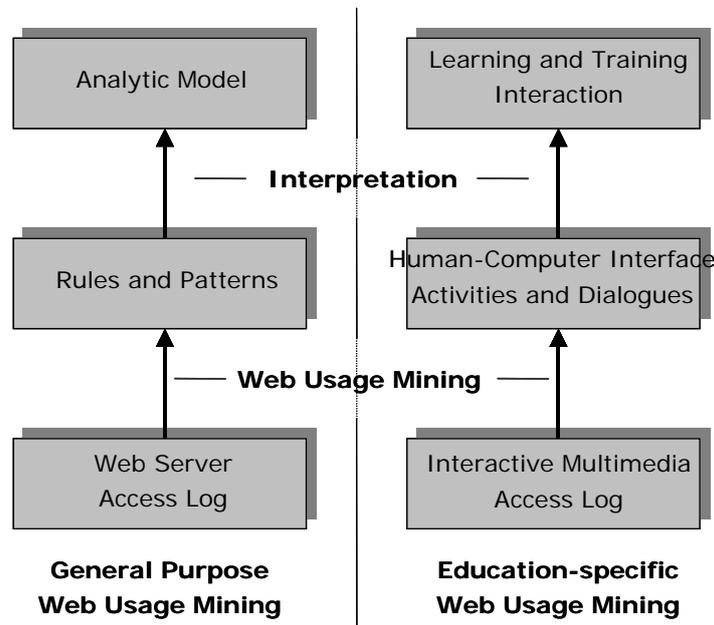


Figure 1: Web usage mining in the educational context.

We propose Web usage mining here as a similar tool to design and manage the relationship between the learner and the educational environment, aiming at an improved learning experience. We can classify data mining techniques [20,21,22,23,24] in general into three broad categories that are of relevance to the educational context [8,9,10,11]: basic statistics (of usage), static relationships, and dynamic relationships.

### 3.2 Data and Web Mining Techniques

Basic quantitative usage attributes are often looked at first, although *usage statistics* are usually not considered as data mining techniques. However, they often form the starting point of evaluations. For Web-based systems, usage is captured in simple measures such as total number of visits, number of visits per page, and so forth. Tracking features of most learning technology systems and most Web log analysers are based on these measures.

*Static relationships* refer to relationships between objects of interest at one moment in time or within a given period of time.

- *Classification* and *prediction* are related techniques. Classification predicts class labels, whereas prediction predicts continuous-valued functions. A model is used to analyse a sample. The result of this learning step is then applied. Regression is a typical form of prediction.

- *Clustering* groups mutually similar data items. In contrast to classification, the class labels are not pre-given. The learning process is called unsupervised here. Pattern recognition is a typical example.
- *Association rules* are interesting relationships discovered among the set of data items. A typical example is purchasing analysis, which can identify item pairs frequently purchased together.

*Dynamic relationships* refer to changes or changing patterns in the database under consideration.

- *Sequential pattern* analysis is applied if events are captured in a database over a period of time. Frequently occurring patterns of events are extracted. Web usage or sales transaction patterns are typical examples.
- *Time series*, which refers to the analysis of the variance of patterns and rules over time, are important since they allow an analyst to evaluate changing and varying behaviour.

The implementation of data mining techniques through software tools can be distinguished into three phases – cleaning, extraction, and interpretation. During cleaning, irrelevant entries such as images can be removed. Extraction is based on the mining techniques, such as the ones we have introduced above, to determine patterns and rules. Finally, the results from the extraction phase have to be interpreted in the context of the application domain.

Despite its potential and benefits, Web usage mining has also some limitations [6,7]. These are, for instance, connected to the fact that not necessary all interactions are recorded in Web logs. Web browsers use caching as a mechanism to avoid network traffic as a consequence of repeated loading of pages. In this case, the usage picture provided through data mining would be distorted. However, problems of this kind can be circumvented if pages are created dynamically, which is standard practice in commercial Web sites.

### 3.3 Education-specific Web Usage Mining

A number of aspects need to be considered if data mining shall be deployed in the educational context. The mining focus is on learning behaviour. Some of the presented mining techniques for standard case are not targeted enough for the educational context to extract meaningful results. Some domain-specific techniques – mainly variations of the standard techniques – shall therefore be introduced. Learning and training interaction is the context in which the mining results have to be interpreted – this shall be captured through an analytic model of interaction. As a consequence of these considerations, we have developed the following education-specific techniques:

- *Session statistics* are based on simple quantitative measurements – the purpose is a quantitative overview of the usage.
- *Session classification* is based on the classification technique – the purpose is learning goal identification.
- *Behavioural patterns* extend and generalise the sequential pattern technique – the purpose is activity identification.
- *Time series* are based on the time series technique – their purpose in the educational context is strategy identification.

We will illustrate these techniques in the next sections using the case study.

Languages play a central role in the mining and interpretation process. The interaction behaviour has to be captured. While Web logs are an expression of behaviour, reflecting the interaction between the learner and content resources, a more abstract language is required to capture these interactions closer to the context of the interpretation and the interaction model. An intermediate level is the multimedia interface and service interaction protocol. Interactivity in the learning context can be expressed through a learning activity language that captures navigation behaviour and allows the interpretation in a learning and training interaction model. The language needs to capture the interaction topology consisting of nodes and arcs between nodes:

- *Nodes* represent the system's *resources* ranging from static content objects to active services. Nodes are named resources, typed for instance based on the activities and topics they support.
- *Arcs* between the nodes represent the user's *activities* – navigation and selection of activities. Activity combinations such as sequences, choices, concurrent activities, and iterations can be expressed.

Given the nodes (resources) *Exercise1*, *Exercise2*, and *Exercise3*, the expression

$$( Exercise1 ; ( Exercise2 | Exercise3 ) )^*$$

means that repeatedly (iteration operator ‘\*’) *Exercise1* is addressed before (sequence operator ‘;’) either (choice operator ‘|’) *Exercise2* or *Exercise3* is addressed. This language can be a tool for design and implementation.

## 4 Session Statistics

A *session* is a central notion in computer-mediated learning and training. A session is technically a sequence of Web log entries that reflects the interaction behaviour of a learner in a period of active study.

Table 1: Access statistics for IDLE Chapter 1 material for a one-week period.

Resource	Number of Requests
ch1-lectov.html	396
ch1-lect1.html	224
ch1-lect2.html	218
ch1-lect3.html	207
ch1-lect4.html	198

*Session statistics* is about basic quantitative measures that can help to answer questions about *which resources are used* and the *investment of time* for a given learning activity, often based on sessions as the basic unit. Any of the results can

be compared against the expectations of the instructional designer. Explicitly formulated expectations form part of the analytic model. There are other statistical measures that might result in useful insights. The total number of requests by interval or total numbers ranked by resource provide relevant information based on simple measures. Table 1, for instance, shows that about half the students that have looked at the chapter overview pages also looked at the chapter content and the number actually finishing the chapter is lower than the number starting the chapter.

For instance, a student submitted in average 239 requests for resources to the IDLE system per term (after the cleaning, i.e. irrelevant resources requests were filtered out). Looking at the ranked total number of requests per resource, the course notes ranked first (initially a surprising result, but a more detailed later analysis showed that students use the course notes online during each session), followed by interactive lab features. A high total number of requests for interactive lab features, where students submit their solutions to given problems, shows a high investment of time for this part. These measures, however, give more an idea about ‘what’ resources are used than ‘how’ they are used.

## 5 Session Classification

The objective of *session classification* is to *extract and identify the main learning goals and higher-level learning tasks* from a session log [25]. Typically, a learner focuses on one or two main activities in a session. Using a *classification* approach we can identify the main learning objectives by looking at the predominant types of interaction with the system.

Table 2: Resource and activity classification.

Class	URLs	Activity
Lectures	{ch1-lectov.html, ch1-lect1.html, ch1-lect2.html, ... }	attending virtual lectures
Tutorials	{ch3-anim1.html, ch5-anim1.html, ... }	participating in a virtual tutorial
Labs	{ch6-sql1.html, ch6-sql2.html, ... }	practising and training in a virtual lab
Downloads	{CourseNotes.pdf, ProjectSpec.pdf, ... }	downloading resources
Look Up	{Schedule.html, Results.html, ... }	look up of course-related information

The media resources of a course Web site can be classified, i.e. a number of classes  $C_1, \dots, C_N$  are created where each class  $C_i$  is a set  $\{U_{i1}, \dots, U_{iM}\}$  of URLs. This corresponds to the nodes and their types of our learning activity language. Each class  $C_i$  is associated to a type of system-level interaction that facilitates a particular knowledge-level learning activity, such as attending a

virtual lecture or working on lab exercises. If a learner spends substantial session time on a particular activity or submits a high number of requests for these resources and activities, then this activity is a manifestation of a particular goal. The requests of pages of the individual classes are counted. In the following example the class names such as *Lectures* or *Tutorials* represent the activities connected to the resources, see Table 2.

Table 3: Session classification.

Class	Percentage of Requests
Lectures	33
Labs	26
Tutorials	21
Download	12
Look Up	8

For each session, a ranking  $C_{i1} \leq \dots \leq C_{iN}$  of main learning goal(s) represented by learning activity classes is produced based on the number of requests for each class, which gives us some insight into the students' learning goals and their implementation. The classification for the sample activities results in

$$Lectures \leq Labs \leq Tutorials \leq Downloads \leq Look Up$$

as the ranking based on Table 3.

This session classification can be generalised for all sessions of a learner or for groups of learners. It gives an instructor a high-level overview of the system usage.

The technique can be used in an iterative evaluation process. Initial classifications might turn out too unspecific and can be refined into more fine-granular classifications, identifying more specific activities, tasks, and goals; thus providing more detailed and meaningful analysis results. While our distinction in our example was by learning activity type, we found it useful in a refined classification to look at the topics as a subclassification, e.g. by chapter resulting in classes such as *LecturesCh1*, *LecturesCh2*, etc. or *TutorialCh1*, *TutorialsCh2*, etc. This is a typical form of a drill-down approach to data mining usage, which gives a more detailed picture of learner activity.

## 6 Behavioural Patterns

The goal identification through session classification that we looked at in the previous section is a tool on an abstract level that ignores the time dimension, i.e. the sequencing of different activities is ignored. Often, however, a closer look at interactions at a lower, fine-granular level is necessary in order to investigate learning activities in detail.

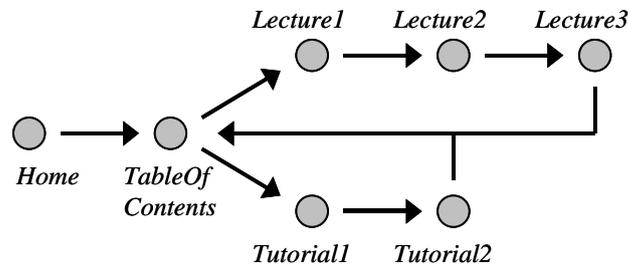


Figure 2: A behavioural pattern (visualised as a graph).

The objective of the *behavioural pattern mining* technique is to *extract behavioural interaction patterns* from the log file. Irrelevant activities – students might lookup other pages, even leave the system temporarily – can be discarded. The filtered sequences are candidates for *sequential patterns*. In order to find out what patterns learners follow, the sequences are subjected to some threshold control – another filter to discard too uncommon ones.

Behavioural patterns encompass more than sequences – learners repeat elements, choose between options, or work on several course elements in parallel. A model of the course topology – navigation infrastructure and interactive elements abstracted by nodes and arcs – underlies behavioural patterns. The navigation along these topologies can be expressed using the *learning activity language* – in particular the expressions on the connections of the topology. A *behavioural pattern* is an expression of a learning activity language that describes potential or actual learning as interactions with an educational multimedia system.

$$(Lecture1 | Tutorial1^*); (Lecture2 | Tutorial2^*); \dots$$

In this example, *Lecture1* and *Tutorial1* are activities. The expression specifies that the learner can use either lectures or tutorials (the | -operator). The tutorial might be attended repeatedly (the \* -operator), before the next lesson is looked at (the ; -operator). These expressions can also be represented in a graphical form, see Figure 2, which visualises

$$Home ; ( TableOfContents ; ( Lecture1 ; Lecture2 ; Lecture3 | Tutorial1 ; Tutorial2 ) )^*$$

in the learning activity language.

The learning activity language we used here to describe these patterns is an integral part of the analysis model that is used to interpret mining results. Therefore, we need to relate these behavioural patterns with the sequential patterns extracted from the Web log. As explained earlier on, activities can be

associated with the transitions between the nodes (URLs) reflected in the log. The behavioural pattern expressions can be a reflection of the instructor's intended use (a design instrument) or an abstraction of the learner's behaviour (an evaluation instrument) gained through Web usage mining. Two applications can be distinguished:

- Verification of expected behaviour: The aim is to compare abstract behavioural patterns specified by the instructor with the actual sequential patterns of learner-content interactions. This only needs sequential pattern extraction to be implemented as a mining technique. For instance, we found that the expected behavioural pattern for a specific lab feature (specified by the instructor) has been met in 85% of all lab sessions (extracted from the Web logs as sequential patterns). Sequential patterns are compared with behavioural patterns by checking for instance the choice and iterations that were allowed according to the behavioural pattern. We used a simulation relation here to determine matching between behavioural and sequential patterns [9]. In order for a sequential pattern to satisfy (or simulate) a behavioural pattern, the following sequential pattern expressions are permitted:

- actual repetitions  $P ; \dots ; P$  are allowed if  $P^*$  is specified,
- choices  $P1$  or  $P2$  can be made if  $P1 | P2$  is specified.

Sequences and parallel uses of resources have to be followed, if required. Alternative definitions that would loosen the simulation constraint could also allow deviations from the required path.

- Extraction of actual behaviour: Actual behaviour is extracted in form of behavioural patterns. The difficulty with this approach is that there is not a unique solution. In particular iterations and concurrent use is difficult to determine, even if the overall topology with its navigation links is known. The extraction of the pattern is the first step; the second is the determination of the support of the pattern by the class. Two patterns can be compared if the distance between them is calculated; this is based on the sequence alignment method [22] where the comparison is based on the extent of deviation from the joint path.

While this technique offers interesting results, the potential has not been fully exploited; more research is required here. More advanced results can be achieved if for example the time spent on each activity and other properties are included in the evaluation [26]. The time spent on an activity can tell us about the actual usage of the resource (and not only the presumed activity based on the intended usage).

## 7 Time Series

*Time series* are sequences of measurements over a period of time. These measurements can include results from any of the mining techniques presented so far. The purpose is here the *detection of change in learning behaviour*, which is often a reflection of the overall learning strategy over the duration of a course. This is important for two reasons.

- Firstly, change might be intended by the instructional designer and the actual occurrence of change needs to be verified. An example for this case is an evaluation of scaffolding features through behavioural pattern analysis. Fading use of scaffolds – features that support students in becoming self-reliant and competent in a topic – is an essential characteristic that is expected to happen in an effective scaffolding implementation.
- Secondly, unexpected changes need to be detected. Web mining-based time series evaluations allow the detection and constant monitoring of student activity changes. Apart from behaviour change, changes of learning strategies have also been observed in the case study. Early patterns often show single-goal use, but later patterns show a concurrent, integrated usage of different educational services. Time series of usage patterns can illustrate the evolution of student learning from Web logs. This was initially an unexpected change in our case study.

Table 4: Time series (by week) based on the results of session classification.

	<b>Lecture</b>	<b>Tutorial</b>	<b>Lab</b>	<b>Download</b>	<b>Look Up</b>
Week 1	35	22	13	19	11
Week 2	29	22	33	9	7
Week 3	20	27	40	5	8
Week 4	42	33	10	9	6

The example in Table 4 illustrates the change in behavioural patterns over a period of four weeks (the first four weeks of term). Initially, learners have predominantly focused on online lectures and have downloaded course material. Over the next weeks, the tutorial and lab elements have been used more frequently – which in the case study reflects the start of the practicals and courseware in weeks two and three of the term.

## 8 Conclusions

Both the design and the evaluation of Web-based learning and training technology systems poses problems due to the novelty of these systems, in particular when different learning and training activities are integrated in highly interactive environments. In this context, the learning behaviour of learners and their interaction with content in particular are central for both design and evaluation activities.

Web usage mining geared towards the specifics of the educational domain can provide access to latent knowledge hidden in access logs for Web-based systems. We have used Web usage mining for different purposes:

- In an *explanatory* style to understand student learning in a novel environment. Mining techniques have been used to clarify the understanding of the learner's goals and task hierarchies. Mining techniques have been

used to investigate the sequential and concurrent use of features. This analysis formed the starting point for the design of a multi-feature learning environment.

- In a *predictive* style to confirm expectations and validate designs. Learning styles and strategies and their change over time were analysed. An example is the design of lab and scaffolding usage support, which has been validated through predicted learning behaviour based on mining data.
- In a *generative* style to improve the design. The instructor's expectations of student learning behaviour were expressed in a learning activity language and compared with actual behaviour. The navigation features and the course topology have been gradually improved using this technique.

The explanatory and generative use of Web usage mining techniques supports the incremental development of Web-based learning technology systems. We have focussed our discussion here on three central usage mining techniques for learning behaviour extraction and analysis. Other applications of Web usage mining in this context are possible. For instance, learner types – or learning styles – could be identified. Kolb's Learning Style Inventory is an example of a learner classification, which suggests concrete experiences, reflective observation, analytic conceptualisation, and active experimentation as four activity dimensions. Based on the preferred activity type, learner types are identified. We could relate these activity types to the activities supported by the case study features: tutorials provide concrete experiences, labs support active experimentation. A classification of learner activities – see session classification – results in a ranking of activities. Variations from the standard pattern can be interpreted as expressions of a particular learning style.

In addition to more substantial analyses that were carried out at the end of term, we also used mining as a tool to constantly monitor students throughout the term, which created a form of immediate feedback for the instructor. While Web usage mining provides useful insights into learner behaviour, it should be combined with other evaluation methods such as surveys to broaden the evaluation.

Education-specific mining techniques can help us to improve the instructional design and also to confirm delivery-related decisions that were made. It has supported instructors in our case study in providing quality instruction and in achieving a better learning experience for learners. In a wider sense, it allows the instructor or author to stay in touch with the learners' activities and to maintain the relationship by reacting to unforeseen events and necessary changes. Mining results can act as a medium of communication between the actors involved in Web-based learning and training systems.

## **Acknowledgements**

The author would like to thank the School of Computing and the Teaching and Learning Fund at Dublin City University for their support. The author is also greatly indebted to Dave Donnellan and Lei Xu for their work on the mining techniques described here.

## References

- [1] Sims, R. Interactive learning as "emerging" technology: A reassessment of interactive and instructional design strategies. *Australian Journal of Educational Technology*, 13(1), pp. 68-84, 1997.
- [2] Weston, T.J. & Barker, L. Designing, Implementing, and Evaluating Web-Based Learning Modules for University Students. *Educational Technology*, 41(4), pp. 15-22, 2001.
- [3] Northrup, P. A Framework for Designing Interactivity into Web-based Instruction. *Educational Technology*, 41(2), pp. 31-39, 2001.
- [4] Oliver, M. (ed.) Special Issue on Evaluation of Learning Technology. *Educational Technology & Society*, 3(4), 2000.
- [5] Chang, G., Healey, M.J., McHugh, J.A.M., & Wang, J.T.L. (2001). *Mining the World Wide Web - An Information Search Approach*. Kluwer Academic Publishers: Boston, 2001.
- [6] Scime, A. *Web Mining: Applications and Techniques*. Idea Group Inc.: Hershey, 2004.
- [7] Spiliopoulou, M. Web Usage Mining for Web Site Evaluation. *Communications of the ACM*, Vol. 43(8), pp. 127-134, 2000.
- [8] Zaïane, O.R. Web Usage Mining for a Better Web-based Learning Environment. *Proc. Conference on Advanced Technology for Education CATE'2001*, pp. 60-64, 2001.
- [9] Pahl, C. & Donnellan, D. Data Mining for the Evaluation of Web-based Teaching and Learning Environments. *Proc. E-Learn 2002 World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*. AACE, 2002.
- [10] Zaïane, O.R. & Luo, J. Towards Evaluating Learners' Behaviour in a Web-Based Distance Learning Environment. *Proc. International Conference on Advanced Learning Technologies ICALT'01*. IEEE, 2001.
- [11] Xu, L., Pahl, C., & Donnellan, D. An evaluation Technique for Content Interaction in Web-based Teaching and Learning Environments. In *Proceedings International Conference on Advanced Learning Technologies ICALT 2003*. IEEE Press, 2003.
- [12] Moore, M.G. Three types of interaction. *The American Journal of Distance Education*, 3(2), pp. 1-6, 1992.
- [13] Ohl, T.M. An Interaction-Centric Learning Model. *Journal of Educational Multimedia and Hypermedia*, 10(4), pp. 311-332, 2001.
- [14] Pahl, C. A Conceptual Architecture for Interactive Educational Multimedia. In J. Ma (ed). *Web-based Intelligent e-Learning Systems: Technologies and Applications*. Idea Group Inc.: Hershey, 2005. (in press)
- [15] Jonassen, D.H., & Mandl, H. (eds) *Designing Hypermedia for Learning*. Springer-Verlag: Berlin, 1990.
- [16] Dix, A., Finlay, J., Abowd, G., & Beale, R. *Human-Computer Interaction*. Prentice Hall: London, 1993.
- [17] Elsom-Cook, M. *Principles of Interactive Multimedia*. McGraw-Hill: London, 2001.

- [18] Hiram, A. The Design and Sequencing of e-Learning Interactions: A Grounded Approach. *International Journal on E-Learning*, 1(1), 2002.
- [19] Pahl, C., Barrett, R., & Kenny, C. Supporting Active Database Learning and Training through Interactive Multimedia. *Proc. International Conference on Innovation and Technology in Computer Science Education ITiCSE'04*. ACM, 2004.
- [20] Agrawal, R. & Srikant, R. Mining Sequential Patterns. Proc. 11<sup>th</sup> International Conference on Data Engineering ICDE, Taipei, Taiwan, 1995.
- [21] Cooley, R., Tan, P.-N., & Srivastava, J. Discovery of Interesting Usage Patterns from Web Data. *Proc. Web-based Knowledge Discovery and Data Mining WEBKDD'99*, pp. 163-182. London: Springer-Verlag, 1999.
- [22] Hay, B., Wets, G., & Vanhoof, K. Clustering navigation patterns on a website using a Sequence Alignment Method. *Proc. 7<sup>th</sup> International Joint Conference on Artificial Intelligence IJCAI'2001*, 2001.
- [23] Mullier, D., Hobbs, D., & Moore, D. Identifying and Using Hypermedia Browsing Patterns. *Journal of Educational Multimedia and Hypermedia*, 11(1), pp. 31-50, 2002.
- [24] Srivastava, J., Cooley, R., Deshpande, M., & Tan, P.-N. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations* 1(2), pp. 12-23, 2000.
- [25] Donnellan, D. *User Session Classification Tool for the Analysis of Web Server Logs*. M.Sc. Dissertation. Dublin City University. School of Computing, 2002.
- [26] Xu, L. *Analysis of Behavioural Patterns using Web Mining*. M.Sc. Dissertation. Dublin City University, School of Computing, 2003.
- [27] Pahl, C. An Evaluation of Scaffolding for Virtual Interactive Tutorials. Proc. 7th E-Learn 2002 Conference, Montreal, Canada. AACE. 2002.