AAT – A Tool for Accessing and Analysing Students' Behaviour Data in Learning Systems

Sabine Graf
School of Computing and Information Systems,
Athabasca University, Canada
+1 (780) 752-6836

sabineg@athabascau.ca

Cindy Ives, Nazim Rahman
Centre for Learning Design and Development,
Athabasca University, Canada
+1 (780) 675-6957

cindyi@athabascau.ca,
nrahman@athabascau.ca

Arnold Ferri
Project Management Office,
Athabasca University, Canada
+1 (780) 421-5866

aferri@athabascau.ca

ABSTRACT

In online learning environments, teachers and course designers often get little feedback about how students actually interact with and learn in online courses. Most of the learning systems used by educational institutions store comprehensive log data associated with students' behaviours and actions. However, these systems typically reveal or report on very general and limited information based on this data. In order to provide teachers and course designers with more detailed and meaningful information about students' behaviour and their use of learning resources within online courses, an analytics tool has been developed. The tool incorporates functionality to access and analyse data related to students' behaviours in learning systems. This tool can provide valuable information about students' learning processes allowing the identification of difficult or inappropriate learning material, and can therefore significantly contribute to the design of improved student support activities and resources.

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Types of Systems – *decision support.*

K.3.m [**Computers and Education**]: Miscellaneous.

General Terms

Design, Human Factors.

Keywords

Academic analytics, learning systems, data extraction and analysis.

1. INTRODUCTION

Educational institutions need ways of responding to internal and external pressures for accountability. Designers of instruction need feedback about how successfully the teaching materials and learning activities support student success. Academic analytics is a relatively recent response to both these requirements in higher education. According to EDUCAUSE, "analytics marries large data sets, statistical techniques, and predictive modeling" [1] to

better understand the wealth of data produced by interactions and transactions in organizational systems with a view to informing action. Campbell et al. [1] describe how the techniques and tools of early institutional-level analytics efforts in administrative areas such as enrolment management and fundraising have evolved to include analyses of factors that support student learning and success. Administrative systems, registration systems and learning management systems (LMS) together provide large amounts of data that can be combined to provide understanding of student engagement and performance. Advanced data analysis skills, integrated information systems and multifunctional collaborative teams are necessary to extract and interpret the evidence for decision making in various academic areas. For example, LMS data reveal student effort measures through participation in discussions, time on task, quiz results, and log files that register click patterns. These and other data can be combined to inform the development of a variety of interventions – from simple early alert systems to customized learning environments to personal learning plans. Some of the factors for success of academic analytic projects include careful attention to ethical and privacy issues, stewardship for systemic implementation plans, attention to faculty perspectives, broad-based collaboration and information sharing, and adequate resources and skills.

Universities that have reported on their academic analytic projects include McGill University in Canada [2], University of Fairfax in the United States [3], and the University of Cordoba in Spain [4]. In 2003, a cross-functional team reporting to McGill's Chief Information Officer used the data from one LMS to understand student and faculty preferences in order to establish criteria for the choice of a replacement LMS [2]. In 2009, a team in the Office of the Dean of Academic Affairs at Fairfax used the community of inquiry model [5] to interpret patterns of faculty interactions with students that were extracted from the enterprise course management system with goals of assessing the relationships between student satisfaction and instructor involvement, and of preparing for a long-term trend analysis project. In 2008, researchers at Cordoba reported on their data mining project in Moodle [6], describing the "emerging discipline" as a way of combining complex student usage data and applying the results to elearning problems such as assessing students' performance, adapting courses based on learning behaviours, evaluating learning materials and courses, providing feedback to instructors, administrators and students, and identifying at-risk students [4]. They described the use of statistics, visualization, clustering, classification, and association rule mining in an iterative "continuous empirical evaluation approach" to course development for online learning. Further, Morris and Finegan [7] suggested using tools available to track student behaviour, collecting aggregate data over time to

understand "course norms" for success and early indicators for failure to persist. They described efforts to predict student retention through analysis of relationships among academic and demographic data types (achievement, locus of control, financial assistance, attitude, motivation), and identified other relevant success factors such as instructor presence and multiple roles, feedback, explicit learning support, and technical assistance available.

These projects and others are also informed by the notion of formative evaluation of courses, which is central to instructional design theories and models [8]. Formative evaluation, a methodology for improving the effectiveness and efficiency of instructional resources [9, 10], is a key phase in most instructional systems design models, which are typically used in the design of online learning environments [9]. Reigeluth and Carr-Chellman [11] proposed that in the information age, design of learner-centred instructional strategies for student success depends on information about learning behaviours and activities; the gathering of this information is facilitated by the technology of the LMS. Baker and Herman [12] provided distance learning guidelines that emphasize adaptation of materials based on students' needs. Activities that enhance motivation and social support lead to student engagement and help them sustain their commitment and persistence. Thus, formative evaluation leads to data-driven design decisions about suggested improvements, whether directly from students to the developer or through system data. They recommended focusing on specific elements of the learning environment to identify weaknesses and recommend revisions.

Building on both lines of thinking above, Athabasca University (AU) – Canada's open, distance university – intends to use data from its enterprise LMS (Moodle) to enhance its online courses through learning interventions and other support mechanisms. By tracking student activities in courses and then linking those data to other data sources we will better understand the effectiveness of the learning environment. This understanding will lead to revisions to course materials and structure that should enhance student learning, motivation and/or persistence. The project is building on past work on student persistence and completion rates at AU, and on studies from other universities on the success of interventions to decrease course dropout, student engagement analysis, and learning design improvements. This is one of several institutional projects designed to support AU's capacity for teaching and learning using digital technologies. Called Moodle Analytics, it involves cross-functional teams of faculty and staff with special expertise in learning design, analytical approaches, institutional research, information technology applications and programming.

In this paper, the Academic Analytics Tool (AAT) developed within the Moodle Analytics project is introduced. First, its objectives and the design decisions in developing the tool are explained. Subsequently, the architecture of the tool and its functionalities are presented. The last section summarises the paper and discusses the potential contributions of the tool to improve academic course design.

2. OBJECTIVES AND DESIGN DECISIONS FOR THE ACADEMIC ANALYTICS TOOL

The Academic Analytics Tool (AAT) is a software application that allows users to access and analyse student behaviour data in learning systems; it enables users to extract detailed information about how students interact with and learn from online courses in a learning system, to analyse the extracted data, and to store the results in a database and/or CSV/HTML files. AAT is primarily developed for learning designers who want to get feedback about how students use and learn in courses, but it can also be used by teachers. In order to use AAT, users need to know the courses they aim at investigating well in order to interpret the results correctly.

While several prototype tools exist that extract particular data from a learning system's database and analyse these data in different contexts [e.g., 13, 14], AAT allows users to decide and specify what data they are interested in and what analysis they want to perform with this data. Furthermore, the data and information that can be extracted and analysed through AAT go far beyond the statistics and activity reports provided in some LMSs, which show limited information predefined by their developers (e.g., information about when a student logged in the last time or accessed a certain activity). Instead, AAT provides comprehensive and customized information to its users, allowing them not only to select from predefined types of information but also to specify what information they are interested in.

Furthermore, most LMS statistics and activity reports are only based on the data from individual courses rather than from a set of courses hosted in a learning system. Similarly, most prototype tools aim at analysing data from one particular course. AAT is designed for academic analytics in educational institutions and therefore aims at flexibility with respect to the choice of courses, allowing, for example, the capture and analysis of data from all courses offered by the educational institution, courses of one or more departments/centres, a single course or a purposefully chosen combination of courses. Furthermore, distinctions can be made between the level of courses (i.e., undergraduate courses, graduate courses, 200-level courses, etc.).

Another objective and design decision was to develop the tool in a generic way so that it could be applicable for different learning systems. Therefore, AAT can be used independently of the learning system used by the educational institution. Furthermore, most educational institutions use LMSs such as Moodle [6], Sakai [15], and Desire2Learn [16], systems which are frequently updated with new versions released regularly. By making the tool applicable for different learning systems, updates to newer versions of the same learning system can be handled easily.

In addition, the tool aims at being easily extendable, for example, with respect to adding sophisticated analysis techniques such as artificial intelligence algorithms, different data sources such as data about students' demographics, marks, etc., and any other kind of functionality that users require to conduct effective academic analyses.

From a technical point of view, the tool is implemented as a web application using PHP as programming language.

3. ARCHITECTURE OF AAT

The architecture of the tool is based on the architecture of DeLeS [14], a tool for identifying learning styles from the behaviour of students in online courses. While DeLeS also aims at being applicable for different learning systems, several extensions in the architecture have been made for AAT in order to fulfil all the objectives described in the previous section.

Figure 1 shows the architecture of AAT. AAT uses input data from one or several databases of a learning system, extracts and analyses the data that are specified by users, and stores these data within the Academic Analytics database or outputs CSV/HTML files with the results.

In order to fulfil the objectives described in the previous section, four design elements have been used: a framework of types of learning objects, patterns, templates, and profiles. In the following paragraphs, these elements are described.

AAT is based on the assumption that each course consists of learning objects, which are digital resources that students interact with and learn from. Learning objects can be, for example, learning material, forum postings, questions of a quiz, the outline of the course as well as video and audio files. Since AAT mainly focuses on analysing the behaviour of students in relation to such learning objects, the consideration of these learning objects is of particular importance.

Learning objects have an inherent pedagogical purpose. However, learning objects of the same type can be used for different pedagogical purposes. For example, quizzes can be used for training or testing, and forums can be used for discussions or announcements. An analytics investigation on two learning objects of the same type used for different pedagogical purposes could lead to erroneous interpretations of results. For example, when analysing students' participation in discussion forums, including forums for announcements would lead to aberrant results.

Therefore, a *framework of types of learning objects* has been introduced that distinguishes between general types of learning objects and pedagogical types of learning objects. General types of learning objects refer to types of learning objects without regard to their pedagogical use (e.g., quiz, forum, resource). Each general type of learning object can be related to one or more pedagogical types of learning objects, which refer to a type of learning object associated with its pedagogical use or educational purpose (e.g., a quiz that is graded and a quiz that can be performed as self-assessment; a forum for announcements and a forum for discussions). By distinguishing between general types

and pedagogical types of learning objects, mixing data that are based on learning objects with different pedagogical purposes can be avoided and misinterpretations due to such a mix of data can be prevented.

Patterns are based on types of learning objects and specify what data the user is interested in and therefore, what data should be extracted from the database(s). A pattern can be a query that extracts specific data, or a formula supported by a query where the tool performs calculations on extracted data. Patterns can be, for example, the average amount of time each student spent on quizzes, the number of times a discussion forum has been visited by students, etc.

Templates aim at making the tool applicable for different learning systems and can be seen as the interface between the tool and the databases. While patterns specify what data should be extracted from a database, templates specify where (i.e. what tables and columns) the respective data resides within the database of a particular learning system, considering the version of the system (e.g., Moodle 2.0). Different templates are developed for different learning systems (and different versions) and are then used for extracting respective data from the database of these learning systems.

A *profile* can be seen as an experiment for extracting and analysing particular information. In a profile, a user specifies which learning system is used (through selecting a template), how to connect to the data (through selecting and setting up database connections), which courses, learning objects and time spans should be investigated (through selecting the data set), and which data the user is interested in (through selecting patterns). AAT guides the user through this specification process. Once the profile is created, it can be used to extract and analyse the specified information.

4. FUNCTIONALITIES OF AAT

AAT is an easy-to-use and powerful tool that allows users to study student behaviour in online courses. It allows users to execute predefined and customized queries against any learning system that stores its data in an SQL accessible database. Users
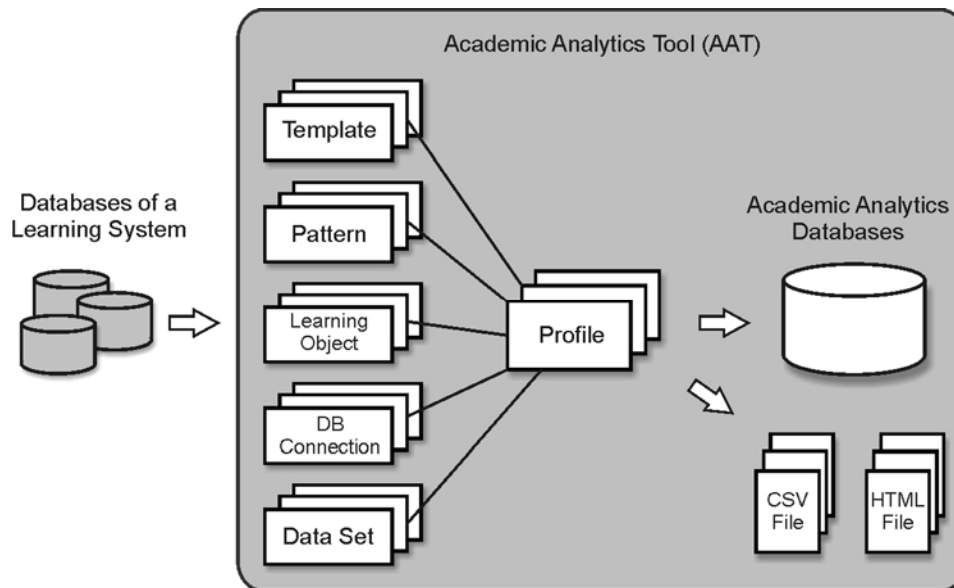

Figure 1. Architecture of AAT

can also chain together queries to make more sophisticated compound queries. More importantly it allows users to progressively improve the analytical capabilities of the tool with a simple to use graphical user interface (GUI). Figure 2 shows a screenshot of AAT, demonstrating the first step in creating such queries.

During the installation of AAT, the administrator specifies database connectivity information and selects a suitable template for the LMS. Based on this information, AAT automatically finds courses and learning objects and makes predefined patterns available to the users. After the installation process, the users can change the selected settings, such as changing the template and adding/removing databases using the GUI.

In the following sections, the main functionalities of AAT are explained.

4.1 Profiles

Users of AAT perform analytical investigations by creating and executing profiles. To create a profile, the user needs to choose a data set (courses and learning objects) and a set of analytics operations to be performed on the data set. Analytics results are generated when a profile is executed. These results can be stored in the Academic Analytics database, displayed on screen, and saved as HTML and/or CSV files.

4.2 Choosing a Data Set

Before users run an analytics query, they need to be able to precisely define the data set they wish to analyse. Using a GUI, users can select the data set they are interested in analysing from the identified pool of courses and their associated learning objects. Functionality for selecting groups of courses is also provided.

Since online courses are not restricted by time constraints, some universities, such as AU, use a continuous enrolment model. In order to make AAT applicable for courses with semester-based enrolment as well as courses with continuous enrolment, AAT allows users to specify the exact periods of time they wish to analyse.

4.3 Choosing Analytics Operations

Once a data set has been specified, users need to define the analytical operations they wish to perform on the data. They can choose from an extensive set of predefined patterns (e.g., overall activities of students in a course, the number of visits of particular types of learning objects, the amount of time spent on particular types of learning objects). Furthermore, users have the option of creating their own custom patterns. The ability to create custom patterns allows users to get answers to questions they need to ask. Multiple patterns can be applied to a data set. Patterns can be chained (i.e., the output of one pattern can be used as input into another pattern). Thus, powerful and complex queries can be constructed incrementally and progressively. The entire process of creating and chaining patterns is performed using a simple GUI, where users can either use an SQL editor that guides them step by step through the process or users can directly input SQL queries.

For example, if a user wishes to identify quiz questions that are difficult to answer for students, he/she can build a pattern that extracts data about the average performance of students on questions within quizzes. On top of this pattern, the user can create another pattern that outputs all quiz questions where the average performance of students is lower than, for example, 70%. Using the results of this pattern, the user can create another pattern that investigates the question types (e.g., multiple choice, true/false, matching, etc.) of the questions that were difficult to answer for students and output a distribution of these types. Furthermore, a user can investigate the learning material that is associated with the questions that were difficult to answer and can, for example, create a pattern that looks into the time students spent on this learning material and compare this time with the average time students spent on all learning materials.
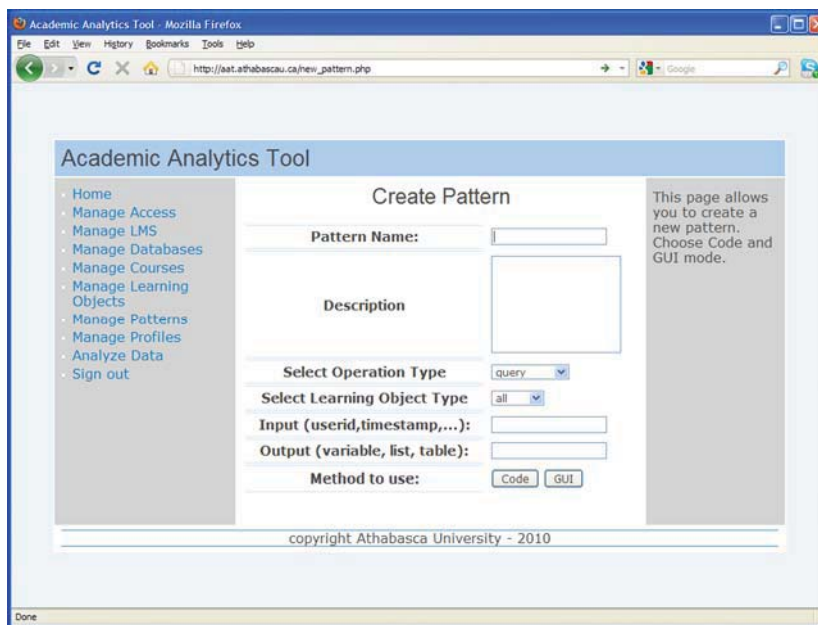

Figure 2. Creating Patterns/Queries in AAT

4.4 Pedagogical Types of Learning Objects

While general types of learning objects (e.g., forum, quiz) are identified automatically by AAT when database connection and learning system information is available, specifying the pedagogical purpose of learning objects requires the user's intervention. To address the issue of pedagogical purpose, AAT allows users to define pedagogical purposes for each general type of learning object, using user-defined controlled vocabulary. Controlled vocabulary schemes mandate the use of pre-selected terms which have predefined definitions. Subsequently, users are supported by AAT to annotate learning objects through a semi-automatic approach, using the defined pedagogical purposes. It is up to the user to interpret the meaning of the pedagogical purpose and thus, it is important for the user to consider the meaning of the pedagogical purpose when defining and/or using a pattern in order to perform data extraction and statistical analysis in alignment with that interpretation.

Using the example given in the previous section, a user can define two pedagogical purposes for quizzes to distinguish between marked quizzes and self-assessment quizzes. Using only marked quizzes to analyse distributions of question types that are difficult to answer for students and learning material that is associated with these difficult questions, will result in more accurate understandings, since some students might not take self-assessment quizzes as seriously as marked quizzes and may choose to take the quiz before reading the learning material.

4.5 Working with Databases

AAT requires read-only access to the database(s) of the learning system it is to analyse. Therefore, the user or administrator must provide database connectivity information. AAT can connect to multiple instances of a learning system's database. In addition, AAT allows users to perform analytics on data from several different instances of a database (of the same learning system) simultaneously. Since it is not uncommon for universities to distribute course data across several databases, AAT is capable of working with such complex database configurations.

4.6 Working with Different Learning Systems

There are many different learning systems available and new versions of learning systems are introduced frequently over time. To make AAT applicable for different learning systems as well as to allow upward compatibility with future versions of learning systems, AAT uses templates to define how to find specific pieces of information from a specific version of a learning system. AAT comes pre-packaged with templates for several learning systems / versions of learning systems. Therefore, a user simply needs to select the right template in order to specify the learning system in AAT. If a template is not available, for example, for a newly released version of a learning system, administrators of the AAT instance, who know the database of the new learning system well, can create new templates. These new templates can then be shared within the community and made available to administrators of AAT systems.

4.7 Extending the Tool

AAT has been created in a modular fashion and many design features have been inspired from the content management system Drupal [17]. Administrators have the option of coding new modules to extend functionality of AAT. Writing AAT modules is as easy as writing Drupal modules.

4.8 Other Features

In addition to the above-mentioned features, AAT provides strong data security and access control features, support for Smarty templates, embedded help files, a SQL editor, a GUI SQL query generator, history, and backup and recovery features. Furthermore, the designers of the tool have made every effort to make it user-friendly and user-centred without compromising the design principles or its functionality.

5. CONCLUSIONS

In this paper, the Academic Analytics Tool (AAT) is introduced. AAT is a powerful and easy-to-use tool designed to allow users to perform simple and complex analytical queries on students' behaviour in online courses. In the following paragraphs, the possible benefits this tool can bring to educational institutions are described, discussing the plans of using the tool at Athabasca University (AU).

The data that can be retrieved through the use of AAT on how students are currently using the learning objects in their AU courses will be evidence for the formative evaluation of those courses. The data will be analysed as part of our regular course revision process. Combined with students' evaluations of the courses and professor and tutor recommendations for changes, these data will inform the work of our learning designers, who with subject matter experts will adapt and extend resources that are generating successful learning and revise materials that are of less direct value to students. Engaging learning objects will be shared across disciplines as appropriate, generating interest in new pedagogical approaches within the academy. Once it is possible to integrate data from administrative systems with data from the learning system, we will be well positioned to identify factors affecting student success. The infrastructure to be established will facilitate the extraction and transformation of data required for improved operational reporting across a number of different aspects of the teaching and learning environment at AU. Eventually it will be possible to generate automated interventions to enhance student retention, motivation and/or learning, and to generate customized dashboards for sharing progress information with tutors and students, thereby meeting institutional goals of quality and access.

The importance of analysing student activities in Moodle has increased over the last couple of years as we have moved from simple course conversion to complete course re-design for the online environment. The direct value of the results of our analyses can be understood in two ways. We will have data about which learning activities students are completing and which ones they are not. This will build on course evaluation data and help inform improvements to individual courses. It will also allow us to evaluate and revise the standards we are setting for excellence in AU online courses.

Future work will deal with conducting a study where learning designers will test AAT with respect to its usability and usefulness. Furthermore, we plan to release AAT as an open source product in order to allow other educational institutions to benefit from AAT as well. In addition, the indirect value of the project will be realized in a methodology for further activities in academic analytics.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Campbell, J.P., DeBlois, P.B., and Oblinger, D.G. 2007. Academic Analytics: A New Tool for a New Era. *EDUCAUSE Review* 42, 4 (Jul./Aug. 2007), 40-57.

[2] Finkelstein, A.B.A., Masi, A.C., and Winer, L.R. 2004. My LMS Gets 1,000,000 Hits a Day: Supporting Your Strategic IT Decisions with Log Analysis Data from your LMS. *Presentation at the EDUCAUSE Conference*, Denver CO, October, 2004

[3] Orcutt, J.M. 2010. Using Enterprise Reporting to Assess Instructor Involvement in Online Classes. *Presentation at Pearson CiTE Conference*, Denver CO, April, 2010.

[4] Romero, C., Ventura. S., and Garcia, E. 2008. Data Mining in Course Management Systems: Moodle Case Study and Tutorial. *Computers and Education* 51 (2008), 368-384.

[5] Garrison, D. R., Anderson, T., and Archer, W. 1999. Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, 2, 2-3 (Spring 1999), 87-105.

[6] Moodle, 2011. http://moodle.org/ (accessed on July 31, 2011).

[7] Morris, L.V. and Finegan, C.L. 2008. Best Practices in Predicting and Encouraging Student Persistence and Achievement Online. *Journal of College Student Retention* 10, 1 (2008), 55-64.

[8] Reigeluth, C.M. and Frick, T.W. 1999. Formative Research: A Methodology for Creating and Improving Design Theories. In *Instructional Design Theories and Models*, Reigeluth, C.M., Ed., Vol. I, Lawrence Erlbaum Associates, Mahwah, NJ, 633-651.

[9] Smith, P.L. and Ragan, T.J. 1999. *Instructional Design* (2nd ed.). Merrill, Upper Saddle River NJ.

[10] Tessmer, M. 1998. *Planning and Conducting Formative Evaluations: Improving the Quality of Education and Training*. Kogan Page, London.

[11] Reigeluth, C.M. and Carr-Chellman, A.A (eds.). 2009. *Instructional-design Theories and Models: Volume III. Building a Common Knowledge Base*. Routledge, New York.

[12] Baker, E.L. and Herman, J.L. 2003. A Distributed Evaluation Model. In *Evaluating Educational Technology,* Gaertel, G., Means, B., Eds., Teachers College Press, New York, 95-119.

[13] Mazza, R. and Milani, C. 2008. Exploring usage analysis in learning systems: gaining insights from visualisations. In *Proceedings of the AIED Workshop on Usage Analysis in Learning Systems* at the International Conference on Artificial Intelligence in Education (AIED 2005), Springer.

[14] Graf, S., Kinshuk, and Liu, T.-C. 2009. Supporting Teachers in Identifying Students' Learning Styles in Learning Management Systems: An Automatic Student Modelling Approach. *Educational Technology & Society* 12, 4 (Oct. 2009), 3-14.

[15] Sakai, 2011. http://www.sakaiproject.org/portal (accessed on July 31, 2011).

[16] Desire2Learn, 2011. http://www.desire2learn.com/ (accessed on July 31, 2011).

[17] Drupal, 2011. http://drupal.org/ (accessed on July 31, 2011).