Online Curriculum Planning Behavior of Teachers

Keith E. Maull[1], Manuel Gerardo Saldivar[2,3] and Tamara Sumner[1,3]
{keith.maull, manuel.saldivar, tamara.sumner}@colorado.edu
[1]Computer Science Department, University of Colorado, Boulder
[2]School of Education, University of Colorado, Boulder
[3]Institute of Cognitive Science, University of Colorado, Boulder

Abstract. Curriculum planning is perhaps one of the most important tasks teachers must perform before instruction. While this task is facilitated by a wealth of existing online tools and resources, teachers are increasingly overwhelmed with finding, adapting and aligning relevant resources that support them in their planning. Consequently, ripe research opportunities exist to study and understand online planning behavior in order to more generally characterize planning behavior. In this paper, we introduce a web-based curriculum planning tool and study its use by middle and high school Earth science teachers. We examine the web analytics component of the tool and apply clustering algorithms to model and discover patterns of the use within the system. Our initial results provide insights into the use of the tool over time and indicate teachers are engaging in behavior that show affinity for the use of interactive digital resources as well as social sharing behaviors. These results show tremendous promise in developing teacher-centric analysis techniques to improve planning technologies and techniques to study online curriculum planning patterns.

1 Introduction

A large body of research suggests that teachers are the single most important influence on students' academic achievement [1]. Thus, helping teachers do their jobs better should lead to improved student outcomes. One of the specific pedagogical techniques now being demanded of many K-12 teachers is *differentiated instruction.* Differentiated instruction involves the customization of curriculum and teaching practices to better foster student understanding of course material [17]. An example of differentiated instruction would be a teacher who uses animations and graphic images to impart a science concept to his students because he has discovered his largely immigrant population of students struggles with English reading comprehension. Given the increasing racial, ethnic, and linguistic diversity in American K-12 schools, differentiated instruction is becoming more important than ever precisely because a 'one size fits all' approach to teaching cannot reach all of today's diverse student body.

At the same time that differentiated instruction is being stressed in K-12, the Internet is changing the educational landscape to a degree not seen since the introduction of personal computers in the 1980s [11]. As Internet connectivity becomes available in more and more classrooms, cloud-based applications and databases have become increasingly important both for teachers' instructional practices and students' learning [3]. Our study lay at the intersection of these two important trends – the drive to differentiate instruction and the widening availability of the Internet in educational contexts.

One of the challenges of differentiating instruction is developing supplementary educational materials that target specific students' learning needs [17]. That is, given a common curriculum, *this* group of students may require additional visual aids to help them grasp a concept while *that* group of students might benefit most from an extra hands-on activity that reinforces a lesson. The Internet provides a portal to a nearly infinite set of digital resources that could help teachers in their differentiation of instruction, but the unmanaged nature of the Internet places the burden of filtering and evaluating digital resources on teachers, adding to their already significant workload. If this filtering and evaluation process could be at least partially automated, teachers would be able to focus on teaching rather than on preparing to teach.

In this paper, we examine two research dimensions related to the process of differentiating instruction: (1) learning science for online curriculum planning and (2) educational informatics for application usage pattern detection and analysis. Since curriculum planning is a task which every teacher must perform before instruction, we begin by describing a Web-based application that is designed to help teachers review curricular objectives, locate relevant supplementary digital resources, and develop differentiated instructional plans that connect their curricular goals and digital materials with classroom activities and assessments.

The first research dimension is framed by a few key research questions : (1) What are the behavioral components of Web-based curricular planning? (2) How do these behaviors play out as patterns in online curricular planning? (3) What techniques are used to measure the effectiveness of curriculum planning behavior? (4) How can or do online tools and resources shape curriculum planning behavior? (5) How do online tools and resources impact curriculum planning outcomes? Not all of these questions will be covered within the scope of this study.

With an application context to study the learning science component of curriculum planning, our second research interest is focused on developing and applying tools and techniques for observing and classifying teachers' online behavior in educational applications. This research offers a unique view into the online usage patterns and behaviors of educators by examining the use of a curriculum planning application in the web mining and analytics context. Specifically, we observe and analyze the use of system features and functionality as well as commonly used resources available within the system to gather a more complete understanding of system use. Using web analytics and clustering algorithms, we develop the education informatics dimension of this research, framed by the following questions : (1) What computational tools can be used to discover and model the online behavior patterns of teachers engaged in curriculum planning? (2) What computational tools can be used to predict the online behavior of teachers once this behavior is modeled? (3) What techniques can be used to maximize teacher's use of online tools and resources?

2 Research Context

While much of the online learning science research to date tends to focus on students' behaviors vis-á-vis computer-assisted learning [12], little research has been done to understand teacher behavior in online curriculum planning tasks. The online behaviors of teachers performing planning tasks online may hold useful clues to the development of applications that not only improve student outcomes, but also teacher outcomes, particularly as they relate to improving teacher access to and use of digital materials within the classroom instruction and learning context. This research presupposes that teachers' online activities are important in their own right and worthy of study.

Two broad contexts are applied to the research : the user context and application context. The focal point of the user context, described in more detail below, is middle and high school Earth science teachers. The corresponding application context is a web-based curriculum planning application called the Curriculum Customization Service (CCS) [15], designed to support those same middle and high school teachers in curriculum planning tasks. The application was designed to provide general support for accessing, searching, browsing, storing, sharing and reviewing curriculum goals, objectives, guidelines, materials and resources for the middle and high school Earth science curriculum.

2.1 User Context

The users of the CCS are Earth science teachers at the 6th and 9th grade levels within a medium sized urban school district. Nearly 120 sixth and ninth grade Earth science teachers within the school district were invited (but not required) to use the CCS program during the 2009-2010 school year. In July 2009, a four-hour face-to-face training session (and a Web-based teleconference for those who could not attend) was carried out to demonstrate the features and use of the CCS. User account access to the CCS was provided during the remainder of the summer so that teachers could further acquaint themselves with the tool and learn more about its features before the start of the semester. Users represent a wide cross section of experience, planning skill, technological ability and interests, with user teaching experience spanned from first year teacher to more than 30 years of experience. No information was collected before the study window about who would choose to participate or who would plan to fully use the tool during the semester, though a survey was given to teachers before the semester to probe their familiarity with technology within the classroom and initial perceptions of the CCS tool during the training sessions. The survey also provided further insights about the teachers' technology-related skills and perceptions, as well as demographic profile information.

2.2 Application Context

The CCS application is a Web-based application that provides access to the specific content of the 6th and 9th grade Earth science curriculum of the participating school district (see Figure 1). The application has several unique features that target teacher-centric behaviors around curriculum planning and customization, specifically planning, storing and sharing, and searching. First, the CCS encompasses the entirety of the district-wide Earth science curriculum – that is it contains all the curriculum units, goals and objectives for the district-wide 6th and 9th grade Earth science curriculum.

Traditional curriculum materials that were typically provided in paper form are instantly accessible digitally through the CCS. Within a single environment, teachers also can access critical planning artifacts : all of the Earth science materials for their lesson planning, including digital access to publisher materials (e.g. textbook content, assessments), curriculum learning goals and objectives, as well as relevant standards and concept hierarchies. Second, the CCS provides instant access to a number of digital resources that are automatically coupled with relevant curriculum unit topics. Teachers accessing unit-level goals have access to pre-fetched digital resources that match the topics of the unit. This feature displays the most relevant and useful materials for quick access, circumventing the need for custom (and potentially time-consuming) searches to find the same materials. Third, a customizable space to save resources of interest, called "*My Stuff*," is provided to encourage teachers to privately save what they may consider to be useful resources. This feature also has the ability to make available to other CCS users the resources a teacher may think his or her colleagues might also find useful. Therefore, within the same environment as planning tasks, teachers can engage fully in customization activities : finding and saving materials of value, sharing and contributing those materials to the participating teachers at large, rating, tagging and even uploading arbitrary materials they wish to store for later use within the planning environment or share with other users. Finally, since searching for contextualized materials is a difficult task within generic search engines (e.g. Google, Yahoo!, Bing), a customized search engine was provided within the application that returns concept-relevant digital resources obtained from search results against the Digital Library for Earth Systems Education (DLESE), a high quality and highly regarded Earth systems digital library [16].

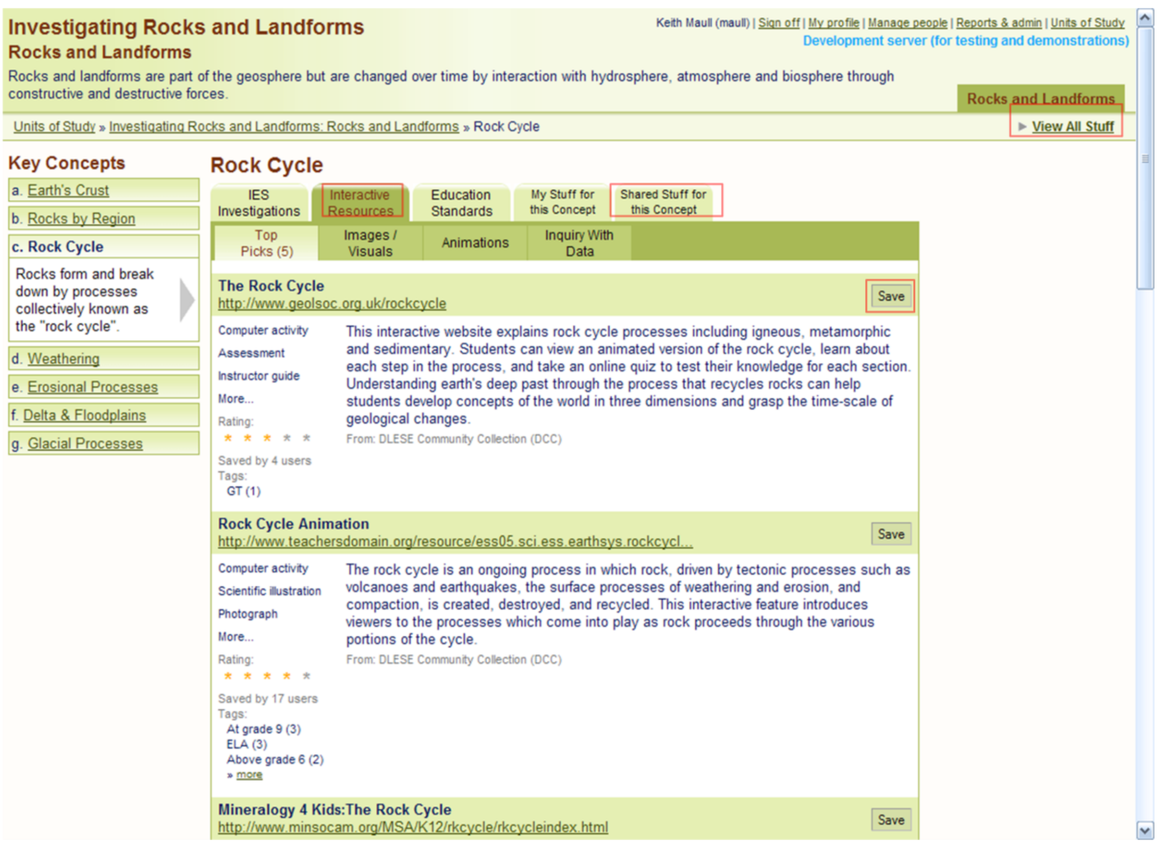

Figure 1: The Curriculum Customization Service User Interface
(A few sample features are in highlighted in red)

The CCS provides a single entry point for teachers to efficiently develop and execute their most important curriculum planning and customization tasks, and as such provides a powerful research platform for several key reasons : (1) it is bounded and constrained to the most frequently used curriculum planning task space used by teachers, thus providing a single place to examine teachers' behavior around this task, (2) the user pool is constrained to a diverse but specific group of users, middle and high school Earth science teachers, so relationships between these users will be more amenable to analysis, (3) the subject area of analysis is narrowed to a middle and high school Earth science curriculum focus, thus amplifying inter-user and inter-group comparison and analysis, and (4) the nature of web-based application activity analysis is well understood and supported by techniques that are robust and effective.

3 Methodology

The goal of our methodology is to understand and prepare the required data to run experiments that allow us to detect patterns of use for further evaluation. The research methodology of this initial study has four components : (1) select and prepare an initial data set, (2) select an experimental feature set, (3) perform clustering experiments, (4) analyze experimental results. Our research draws on widely understood techniques for web server log analysis [11] of the CCS application. While it is not difficult to determine system use frequency, such as commonly visited pages or resources, frequent use alone provides a narrow view of actual use. System components are often used in concert with one another, thus providing different views of system use that may reveal unexpected or unusual relationships. We therefore turn to clustering algorithms to help build connections among the CCS system components that make up the experimental feature set.

3.1 Data Set and Data Filtering

Data for the initial set of experiments in this study was derived from the server logs of the core CCS system. The CCS system has a number of unique client-side scripting features that are not usually captured in server logs because there is typically no direct server interaction during such activity. This client-side activity was capture by instrumentation that stored such activity as if it were direct server interaction. All users of the system were given unique login IDs that were captured during session activity over a 16 week period from August 2009 to December 2009.

Allowing for the broad use of the CCS, session duration was defined to accommodate a complete working day of activity. Sessions longer than 10 hours or shorter than 30 seconds with fewer than 4 actions were eliminated from the data set. While the minimum session length may seem short, there were some sessions detected that lasted for abbreviated periods of time but nonetheless showed meaningful activity. Logging in to the system to access a singular, specific resource is one such example. Other filtering activities involved removing data with user IDs of non-teacher users not part of the research study.

3.2 Experimental Features

Feature selection for data mining is a difficult task and requires many considerations to be both valuable and effective [12]. For the initial research experiments, the user session was abstracted and examined for features to be used in the clustering experiments detailed in section 5. In all, 27 features were selected for our initial feature set.

Web sessions are usually seen as a temporal ordered stream of clicks (click stream). However, for our initial research these sessions were simplified into a larger grained *bag of clicks*. Instead of examining the transitions from one click to another, we assign and examine click types within the system for a given session. This *bag of clicks* model simplifies log processing and feature selection and borrows similar intuitions found in *bag of words* models in linguistic processing, where word frequency is used to analyze document content, sometimes arriving at similar conclusions as temporal, structural or semantic analysis with far less complexity.

Each client-side UI click has one of 5 associated *click types* which our initial experiments exploit as a feature. Furthermore, since click stream data is ultimately important in understanding session behavior, the *number of clicks* within a session was also selected as a feature. It would not be unreasonable to expect sessions with high click counts to represent sessions of interest. In addition to the number of clicks, *session duration* or the amount of time (in seconds) a user spends interacting with the system, was chosen as a feature.

Other data was also selected as part of the initial feature set selection process. Since the CCS is a highly visual and largely client-side interaction environment, commonly used visual controls of the environment were natural targets for the experimental feature set. This visual features set was narrowed to the *top 20 visual features* of the interface, which was computed by flattening the visual hierarchy of the interface controls and examining the global use frequencies of each of these visual components.

User activity can be measured in many ways and for the purpose of this study, we make note of several server log data. First, *total actions analyzed* provides an indication of how many discrete actions were logged from the server, *discrete sessions* details the total number of sessions in an interval (monthly, weekly, etc.), and *unique user accounts* shows the number of unique user accounts in a period. The total number of actions over the 16 week data collection period was 17,527 in 1,370 total sessions over 82 unique user accounts, detailed in Table 1.

Table 1 : Monthly data summary

Month	***Total Actions Analyzed***	***Discrete Sessions***	***Unique User Accounts***
September	7,371	526	82
October	3,626	331	65
November	4,698	337	60
December	1,832	176	50

Table 2 : Cluster ranks and relative size (%)

Rank	***K12***	***EM12***	***EM****
1	58%	47%	37%
2	14%	17%	26%
3	8%	11%	11%
4	8%	5%	10%
5	5%	5%	–
6	4%	–	–

4 Algorithms and Feature Analysis

Clustering algorithms are commonly used to find patterns within large data sets [2, 6, 7] and two clustering algorithms were chosen to study the initial data set. First, the *K*-means clustering algorithm was used. *K*-means is an unsupervised learning, iterative descent algorithm that partitions n data observations into K clusters. Each cluster is assigned a centroid and cluster membership is determined by minimizing the distance from each cluster member and the centroid. The second algorithm in the initial experiment was the Expectation-Maximization (EM) algorithm [4]. EM is a model-based iterative algorithm that examines data observations and represents each cluster as a probability distribution. Given n data observations, EM maximizes the likelihood of the observed distributions by estimating the means and standard deviations of each cluster.

Neither algorithm is without flaws and our experimental results in section 6 show this. K-means primary weakness is that the number of clusters must be determined *a priori*. This weakness, however, is inherent in many partitional clustering algorithms and may require an experimentally selected n. Another weakness is that K-means is sensitive to outliers – data that are distant from the centroid may pull the centroid away from the real centroid in a given data set. Finally, it is difficult to understanding which feature contributes more to cluster membership, since every feature is assumed to have the same weights. EM's core weaknesses are the relative speed with which the clusters converge, and the possibility of convergence at the boundary of a cluster.

5 Experiments

Two experiments were performed on the initial feature set with a few variations for comparison. The first experiment was designed to run the K-means algorithm with $n = 12$ and the Euclidean distance function, referred to as K12. The n for this initial experiment was derived from the total number of sessions analyzed over the period (~1,400) divided by the number of users invited to participate (~120). This provides a baseline for comparison with the other algorithms. For comparison, the expectation maximization algorithm was chosen for the remainder of the experiments. EM was first chosen to automatically select n clusters using cross validation, referred to as EM^*. This provided a baseline to compare the algorithm's performance against the K-means algorithm (K12). The last experiment was run using EM again with a fixed cluster size of 12, referred to as EM12.

6 Evaluation and Results

Table 2 shows each of the algorithms and the sizes of the largest clusters they produced. EM^* produced 10 clusters, the top 4 of which represent 84% of all the data. Similarly, for EM12, the top 5 largest clusters represent 84% of the data. Finally, K12 shows a similar trend, with the top 5 clusters representing 87% of the data. For the purposes of evaluation we consider the top 4 clusters in EM^*, top 5 in EM12, and the top 6 clusters in K12, since the K12 distribution of clusters was more sparse.

Table 3 shows the features with the greatest means of the top clusters for each algorithm. The cluster labels represent UI features for example, *A1* represents clicks on the *Interactive Resources* tab, A2 the *Shared Stuff for This Concept* tab, *A6* for the *Embedded Assessments* toggle element, *A12* for the *Images/Visuals* tab, and so on. The top features of the largest cluster in EM^* (A1, A2 and A4) correspond to CCS UI tab clicks on *Interactive Resources*, *Shared Stuff for This Concept* and *Shared Stuff for This Activity*. This top cluster suggests a pattern of activity that is focused on both CCS-suggested interactive resources *and* shared resources that others have saved, which may indicate the importance of what *others* have saved as well the automatically generated interactive resource list.

The EM12 and K12 algorithms indicate very similar patterns. For example, EM12's largest cluster shows the exact same pattern as EM*. Similarly, K12 shows A1, A3 and A4 as its top features. Examining other clusters show cluster 4 of K12 and cluster 11 of EM12 share similar patterns over features A3, A6, A8 and A11. This pattern corresponds to clicks on *Instructional Support Materials*, *Embedded Assessments*, *Answers* and *Teaching Tips* system areas. Similarly, cluster 2 of K12 and cluster 1 of EM12 share similar features along A3, A7 and A14, which correspond to the *Instructional Support Materials* and *Activities* tabs, suggesting time being spent on preparing or reviewing student activities and corresponding materials.

Table 3 : Top cluster features and their cluster membership

	Cluster #	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A11**	**A12**	**A13**	**A14**	**A16**
K12	1 (8%)	■									■			
	2 (5%)			■				■					■	
	3 (14%)	■	■				■							
	4 (4%)			■			■		■	■				
	8 (4%)	■												■
	9 (58%)	■		■	■		■							
EM*	0 (26%)	■	■			■						■		
	1 (11%)				■			■					■	
	6 (10%)	■				■		■					■	
	7 (37%)	■	■		■									
EM12	1 (5%)			■	■			■					■	
	3 (5%)	■	■			■								■
	6 (17%)	■				■		■					■	
	7 (47%)	■	■		■									
	11 (11%)			■			■		■	■				

Each cluster algorithm revealed data that was consistent with overall system use seen in the server logs, though the smaller cluster sizes show greater differentiation of features. That there was not complete agreement in cluster features or sizes, however, may indicate more experiments are required.

7 Related Work

Much of the work here has been influenced by the body of work in web use analytics, which break down into two categories : (1) content analytics and (2) usage analytics [12]. This work is focused on usage analytics. Broadly, use analytics aims at understanding the aggregate activity and use patterns of a website primarily using advanced server log analysis. Such analytics often aim at understanding aspects of the site that are popular, content that seems to be frequently accessed, times of frequent/infrequent use, etc. with the goal of developing a sense of where the site could be improved or enhanced for optimal performance, increased advertisement penetration or site content enhancement through recommender techniques [13]. Such use analytics are invaluable for developing site content, but also useful in developing models of user behavior. Website session characteristics are commonly studied to determine how users are accessing the site and statistical techniques are used to determine tasks being performed within a website, revealing cluster usage patterns in the ways we have discussed here. Markov models have been used to derive the meaning of certain behaviors within a session by observing page transitions and their probabilities to develop behavioral models of use [5]. Work has also been done to connect page semantics to web usage, for example [9] use Probabilistic Latent Semantic Analysis to determine if the content and subsequent usage of a page implies an underlying task. Finally, user interface event mining [8] aims at developing techniques to exploit detailed user experience and interaction data.

8 Discussion

The initial experiments presented in this paper offer some insights into the planning behavior of teachers online. However, two areas of improvement can be immediately discussed : improved feature selection and expanded algorithm experiments and comparison. The initial experimental feature set provides interesting insights into the behavior of teachers for the visual components selected in this initial observation. However, the flattened visual hierarchy of the CCS interface only provides a convenient way to discretize each visual element of the system without advancing the notion of the *semantic* structure of this hierarchy. For example, while it is clear that the *Interactive Resources* tab of the interface was widely used, there are substructures under that tab which also contain widely used features. The current feature set is not capable of capturing this hierarchy or its implied semantic structure, though considering it might yield new insights into the semantics of the features commonly accessed by users. Further extensions to the feature set might also include adding link-to-link features, for example, exploring high frequency transitions might reveal unique relationships between UI features and functionality.

The EM and *K*-means algorithms are commonly used in data mining, and while some clusters in K12 and EM12 had similar characteristics, all of the top clusters were not similar enough to say both algorithms were converging on *exactly* the same feature sets. This may underscore the differences in each algorithm or in the way they each treat the features. It may also reinforce the effects parameter sensitivity (e.g. *n* clusters) and feature selections have on the results. The focus of the next round of experiments will be to experiment further with EM and *K*-means parameters, and also to expand algorithm coverage to hierarchical-based algorithms. Such experimentation may also fit well with the semantic features already suggested and allow a comparison of the hierarchies that are produced from a semantic-based structure with the clusters already observed.

As with all learning algorithms, it is challenging to determining if the experimental data would be predicted by and hold up to some gold standard or human expert evaluation. Determining if the discovered behaviors match the observed data in practice is difficult and further research is underway to study actual and reported system use through on-site observation and survey instruments, which should lead to a higher fidelity confirmation of the patterns discovered thus far.

References

[1] Brophy, J. and Good, T.L., Teacher behavior and student achievement. In M.C. Wittrock (Ed.) *Handbook of research on teaching*. McMillan : New York, 1986.
[2] Cios, K.J., Pedrycz, W., Swiniarski, R.W. and Kurgan, L.A. *Data mining: A knowledge discovery approach*. Springer Verlag, 2007.
[3] Collins, A. and Halverson, R. *Rethinking education in the age of technology: The digital revolution and the schools*, 2009.
[4] Dempster, A.P., Laird, N.M., Rubin, D.B., et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
[5] Deshpande, M. and Karypis, G. Selective Markov models for predicting web page accesses. *ACM Transactions on Internet Technology (TOIT)*, 4(2):163–184, 2004.
[6] Frank, E. and Witten, I.H. *Data Mining: Practical machine learning tools and techniques with java implementations*. Morgan Kaufmann, 2005.
[7] Han, J. and Kamber, M. *Data mining: Concepts and techniques*. Morgan Kaufmann, 2006.
[8] Hilbert, D. M. and Redmiles, D. F. Extracting usability information from user interface events. *ACM Computing Survey*, 32(4):384–421, 2000.
[9] Jin, X., Zhou, Y., and Mobasher, B. Web usage mining based on probabilistic latent semantic analysis. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 197–205. ACM New York, NY, USA, 2004.
[10] Leu Jr, D.J., Leu, D.D., and Len, K.R. *Teaching with the internet: Lessons from the classroom*. Christopher-Gordon Publishers, Inc., 480 Washington St., Norwood, MA 02062, 1997.
[11] Liu, B. *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer, 2007.
[12] Liu, H. and Motoda, H. *Feature selection for knowledge discovery and data mining*. Springer, 1998.
[13] Mobasher, B. Data mining for web personalization. *Lecture Notes in Computer Science*, 4321:90, 2007.
[14] Smith, R., Clark, T. and Blomeyer, R.L. A synthesis of new research on K-12 online learning. Naperville, IL: North Central Regional Educational Laboratory. 2005.
[15] Sumner, T., Devaul, H., Davis, L. and Weatherley, J. A curriculum customization service. In *Proceedings of the 2009 Joint International Conference on Digital libraries*, pages 479–480. ACM New York, NY, USA, 2009.
[16] Sumner, T., Khoo, M., Recker, M. and Marlino, M. Understanding educator perceptions of "quality" in digital libraries. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 269–279. IEEE Computer Society Washington, DC, USA, 2003.
[17] Tomlinson, C.A. and McTighe, J. Integrating differentiated instruction and understanding by design: Connecting content and kids. Alexandria, VA: Association for Supervision and Curriculum Development, 2006.