In U.Hoppe, F. Verdejo, J,. Kay (eds.) Proceedings of the International conference in Artificial Intelligence in Education. Sydney July 20-24, 2003. (AIED 2003). IOS press. pp. 279-286. ISBN 0922-6389.

CourseVis: Externalising Student Information to Facilitate Instructors in Distance Learning

Riccardo MAZZA*, Vania DIMITROVA+
*Faculty of communication sciences, University of Lugano, Switzerland
+School of Computing, University of Leeds, UK
riccardo.mazza@lu.unisi.ch, vania@comp.leeds.ac.uk

Abstract. In this paper we present an approach for externalising student information to facilitate instructors involved in distance learning. Information visualization techniques have been used in CourseVis - a tool which obtains tracking data from course management systems, transforms the data into a form convenient for processing, and generates graphical representations that can be explored by course instructors to examine social, cognitive, and behavioural aspects of distance students. CourseVis is presented in the paper, and several examples of pictorial representations generated by the tool are discussed.

Keywords: web-based distance learning, external representations, instructor support, information visualization.

1 Introduction

Web-based distance education is rapidly becoming very popular. Computing technologies, such as Course Management Systems (CMS), that support the development and the delivery of distance learning courses over the Internet are widely exploited nowadays. Although these systems support many tasks related to teaching at distance, the instructors still face a number of problems with managing distance courses effectively, most of which are brought by the difficulty to gain sufficient understanding of social, cognitive, and behavioural aspects of distance students [2,4]. For example, instructors often need to obtain an overview of the performance of their classes, monitor discussions, cluster learner groups based on certain patterns of behaviour/performance, identify tendencies in different groups, discover common misconceptions, etc. CMS usually keep a thorough record of the students' activities in a distance course, but this complex information is rarely used by instructors since it is predominantly in a numerical format, often incomprehensible, with a poor logical organisation, and difficult to follow.

A challenge that developers of web-based learning environments face is to process complex multi-dimensional data provided by CMS and to transform the data into an appropriate form that can be used by instructors in their practice. In our work we examine one possible way to tackle this challenge by applying Information Visualisation techniques [1, 10, 12] to process the data from CMS and render it into an appropriate graphical manner. In this way, we aim to aid instructors in distance courses to gain understanding of their learners and become aware of what is happening in their classes [6].

Information Visualization is a field in Computer Science that examines techniques for processing and pictorially representing a vast amount of abstract data, so that the data can be comprehended and interpreted by people. Visualization is a powerful tool in three major tasks: *exploration* (searching for relationships, trends, and interesting phenomena); *confirmation* (validating or refuting hypothesises); and *presentation* (conveying information to others). By managing complex multi-dimensional data with appropriate visualization techniques people form *mental models* of the data and obtain a better understanding of specific features of the data [10]. Therefore, it can be expected that suitable pictorial representations of data from distance classes will help the instructors to form *mental models of individual students* as well as *mental models of groups of students*. By using these models, the instructors can provide more effective instruction.

We have exploited Information Visualization techniques in a tool called CourseVis which obtains tracking data from a CMS (WebCT is used in the current exemplification of CourseVis), transforms the data into a form convenient for processing, and generates graphical representations that can be explored and manipulated by course instructors to examine social, cognitive, and behavioural aspects of distance students.

Recent research has shown the need for externalising the information collected by CMS to analyse social aspects in distance learning groups. Reffay and Chanier [9] use graphs to monitor group communications in order to help instructors detect collaboration problems. Our approach is a step further in using graphical representations to facilitate instructors in distance learning. While Raffay and Chanier adopt only one form of externalisation, we employ several graphical representations and have followed a data visualization methodology to represent not only social, but also cognitive and behavioural aspects.

Some forms of visualizing cognitive aspects of students have been explored in open student modelling projects, e.g. ViSMod [13] uses concept maps to render a Bayesian student model; UM [5] exploits different types of geometric forms to represent known/unknown concepts; KERMIT [3] uses histograms to represent levels of a student's knowledge. However, the pictorial representations used in open student models have been chosen rather ad-hoc and no systematic analysis has been done to find out which graphical form is suitable in what case, while CourseVis is based on a systematic approach for selecting appropriate graphical representation which may be extended to suggest appropriate environments for externalising student models. Most open learner modelling systems deal with a representation of an individual student that hardly supports comparison with other students. Representations used in CourseVis could be employed in multi-user AIED systems to open models of groups to teachers and learners.

This paper presents CourseVis and focuses on the visualization techniques used. Next in the paper, the gathering of design requirements for CourseVis will be outlined in Section 2, Section 3 will present the architecture of CourseVis, and the following section will discuss several visualization examples based on data from a distance learning course run at the university of Applied Sciences of Southern Switzerland. Finally, we will present some conclusions and will discuss plans for future work.

2 Gathering Design Requirements for CourseVis

The first step in developing usable systems is gathering design requirements. A survey was conducted to find out what information about distance students instructors need when they run courses with CMS and to identify possible ways to help instructors acquire this information. An online questionnaire was sent to a number of people involved in distance learning[1], 98 responses from 11 countries were received from a representative sample of

[1] The questionnaire was sent to several discussion forums and mailing lists, see [7].

potential users of a system like CourseVis. We also had several discussions with distance learning instructors in Lugano and Leeds to uncover aspects that had not been captured by the questionnaire. The study and the results have been presented in detail elsewhere [7], here we focus on informing the design of CourseVis and sketch out social, cognitive and behavioural aspects of students that tutors in distance courses may be interested in.

2.1 Social aspects of learners

There are various kinds of interactions that occur in a web based course run with a CMS, such as interactions between individual students, interactions between students and the instructor, interactions between the students and the system, and vicarious interactions[2]. The main tools in on-line courses that engage students and instructors in communicative activities are discussion forums, e-mail and chat. The study showed that discussions were considered as the most important in the evaluation of both the course and the students' activities. Although qualitative analysis of discussions provides deep insight into social aspects in distance classes, it is usually laborious and time consuming. Instead, by using suitable visualization techniques, quantitative analysis can be performed to discover general tendencies and phenomena about social aspects of students as well as to highlight parts of the interaction for further qualitative analysis. The students' participation in group work is also crucial for the instructors and appropriate visualization of the data about this kind of activity should be considered. E-mail and chat, while often used in on-line courses, have been rated in the study as not very relevant for teaching purposes.

2.2 Cognitive aspects of learners

Many instructors stressed the importance of understanding cognitive aspects of their students and pointed out that appropriate support was needed. Cognitive aspects are usually monitored in distance learning by considering the students' overall course performance, their results on a selected evaluation proof (e.g. quiz or assignment), the students' familiarisation with a specific topic of the course, etc. It was highly recommended to give a suitable external representation of the overall performance in the course and the level of knowledge achieved by each student for every domain concept of the course. A visualization system should provide instructors with a clear and immediate external representation of students having difficulties with domain concepts. The representation should facilitate comparison of individual students with the whole class.

2.3 Behavioural aspects of learners

Instructors use behavioural indicators to judge factors such as active learning, motivation, engagement, and, in general, to assess the success or failure of a distance course. The participants in the survey pointed out various behavioural indicators required in distance teaching: the students' attendance in the course, students performing very well or very badly, students that are progressing too fast or too slowly, etc. It was stressed that the monitoring of the students' access to the course was vital and suitable representations of the attendance data should be provided. It was also identified as crucial to supply appropriate visualisation for monitoring the students' readings of course material (access, time spent),

---

[2] Vicarious interaction takes place when a student actively observes and processes the interaction between other students or between other students and the instructor, without taking an active part in interaction [11].

the performing of evaluation proofs, and the participation in discussions. An indication about how the students' are progressing with the schedule of the course should be provided.

These requirements have been taken into account in the design of CourseVis to decide how to use data from CMS and to identify suitable visualization techniques. The design of CourseVis is illustrated in the next section where the CourseVis architecture is presented.

3 The CourseVis architecture

CourseVis is developed as a tool that extends existing CMS by processing and pictorially representing data collected by the CMS to help instructors understand the needs of their learners and manage distance classes effectively. CourseVis considers social, cognitive, and behavioural aspects discussed in section 2. Figure 1 presents the architecture of CourseVis.

CourseVis imports records of students’ actions supplied by the CMS employed to run an on-line distance learning course. Despite differences in the format of the tracking information provided by various CMS, there are commonalties in the content and the structure, e.g. history of the pages visited, marks the students receive for each quiz, messages posted to discussion forums, etc. This has motivated our work on designing a general course visualization tool that can be applied to a wide range of distance learning environments. Although CourseVis is currently based on data imported from WebCT, it can be adapted to supplement other CMS. The only module in CourseVis that is dependent on the CMS is the *Student Data Exporter* which transforms the data provided from the CMS into an XML format. This enables the other modules in CourseVis to be based on a unified data syntax and semantics, and to be independent from the particular CMS used. We have currently experimented with importing data from two different courses both managed with WebCT. In order to run CourseVis with data from another course-management platform, e.g. Blackboard, an appropriate Student Data Exporter has to be developed. The current Student Data Exporter for WebCT is written in Perl and includes simple text parsing algorithms. The transformed data is stored in a *Raw Data Repository*.

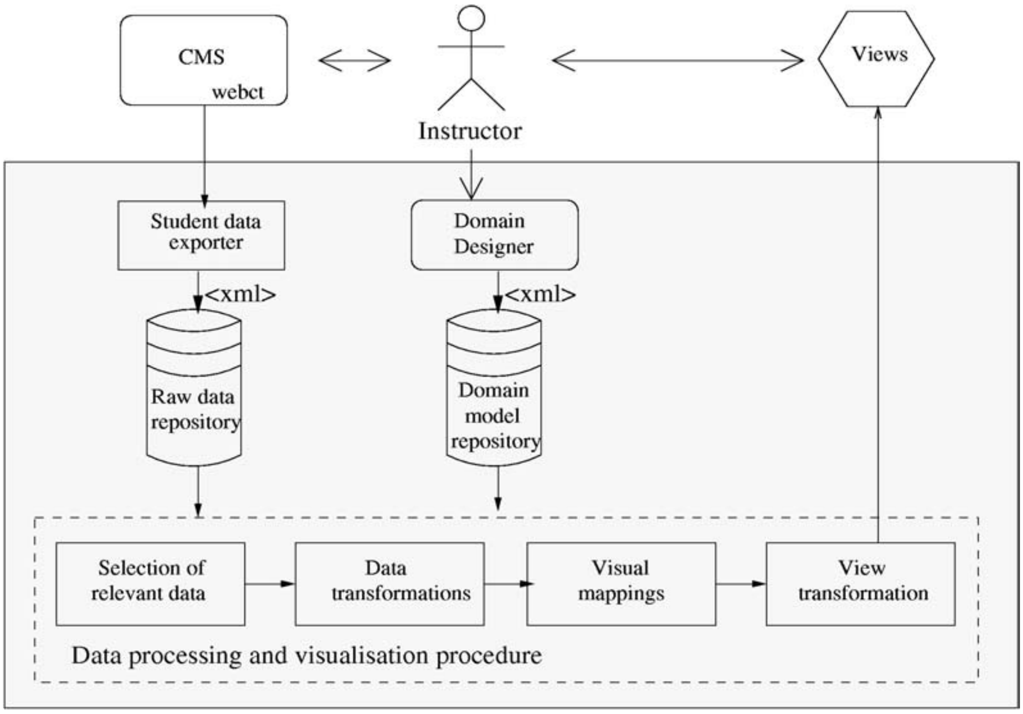


Figure 1: Structure of CourseVis.

The *Domain Designer* is a module that is provided for instructors to define the domain repository of the course. The *Domain Model* is encoded in XML and contains a list of concepts where each concept is associated with a set of pages and evaluation proofs (i.e. quizzes, group exercises) from the *Raw Data Repository*. To enable flexible use of CourseVis, we have not included dependencies among concepts in the current version. Possible inclusion of hierarchical concept relations is being considered.

The *Data Processing Procedure* follows the "visualization pipeline" model proposed by Stuart Card [1]. Data stored in the repositories is passed through a pipeline of four stages:

(a) *Selection of relevant input data*: extracts from the repositories data relevant to the particular visualization.

(b) *Data transformation*: generates intermediate form of the analytical abstraction of data together with metadata of the abstract representation.

(c) *Visualisation mapping*: transforms the pre-processed filtered data into geometrical primitives with appropriate attributes, such as colour or opacity. Visualization mapping is the core of the visualization process.

(d) *View transformation*: generates the image by using the geometric primitives from the mapping process to generate the output *Views*.

CourseVis is implemented in Perl and Java except the last two stages of the Data Processing Procedure (Visualisation mapping and View transformation) which are implemented by using the OpenDX visualization tool [8]. OpenDX is a powerful open source package developed at IBM for visualizing analytical (as well as scientific and engineering) data. It allows the construction of complex visualizations by combining components in an extended data flow architecture. OpenDX provides a simple GUI builder that enables user customisation, and includes interfaces that allow OpenDX functionality to be incorporated into stand-alone applications.

4 Examples of Graphical Representations Produced in CourseVis

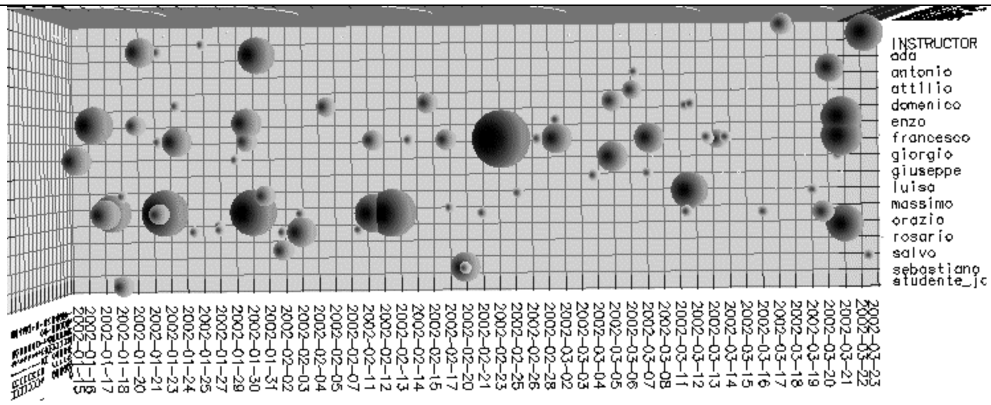
This section illustrates pictorial representations generated in CourseVis. The data[3] used is from an on-line course in Java Programming run by the first author at the Department of Informatics and Electronics of the University of Applied Sciences of Southern Switzerland. We present one example for each of the student aspects discussed in section 2.

4.1 Example of visualizing social aspects of learners

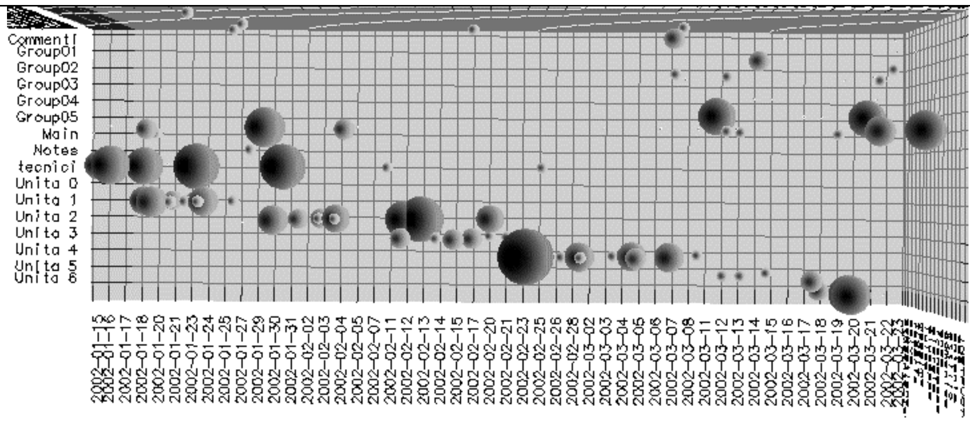
In this example a 3D scatterplot is used to represent data about student discussions throughout the course, see Figure 2. There are three dimensions in the data set: time, discussion topic, and student. An additional dimension – the size of the discussions – is represented by using different size: the size of the spheres represent the number of follow-ups in a discussion. The instructor may use operations, such as brushing, zooming and rotating, to manipulate the visualization. Figure 2 shows two rotations that enable the analysis of different relations by examining the same data set. The top scatterplot presents the discussion treads open by each student during the course. Certain social characteristics of students can be discovered. For instance, it can be seen that the discussions were predominantly initiated by students, and two individuals (Francesco and Massimo) dominated in opening treads followed up by others. The instructor might delegate communication tasks to such students, e.g. to moderate discussions. Domenico has also

[3] To keep the anonymity of the data, all real student names have been replaced with aliases.

been active in opening discussion treads but has not been followed by others. The instructor may decide to look at some of the treads initiated by this student and encourage longer discussions, if appropriate. The figure identifies students who have not been active in opening discussions. The instructor may need to pay more attention at these students, indeed, one of them (Ada) dropped off the course.



Visualization of discussion treads focusing on the students who have initiated the treads



Visualization of the discussions related to the course units and group activities

Figure 2: Rotating a 3D scatterplot representation of discussions to analyse different relationships.

Using the second representation, the instructor can gain insights into the discussed topics. Some treads can be identified, e.g. there are lengthy discussions about technical problems at the beginning of the course, most units have had fairly active discussions. There has been a group work assigned, the top lines of the representation show the communication in each group. While group 5 has communicated intensively, groups 3 and 4 have not communicated at all (at least not using the tools in the course). The instructor may intervene the group work by encouraging the students to communicate and collaborate more effectively.

4.2 Example of visualizing cognitive aspects of learners

Figure 3 shows an example of representing the student performance on the course quizzes. The data set has two dimensions – student name and concept. A third dimension – the level of knowledge of a student about a concept - is indicated by using colour (darker colour means a higher level of knowledge). The matrix in Figure 3 enables global analysis of the overall performance of students on the course topics and comparison between topics. This can promote the instructor's reflection on his/her practice. For instance, surprisingly, array was found a difficult concept by most of the students, and looking at the course material,

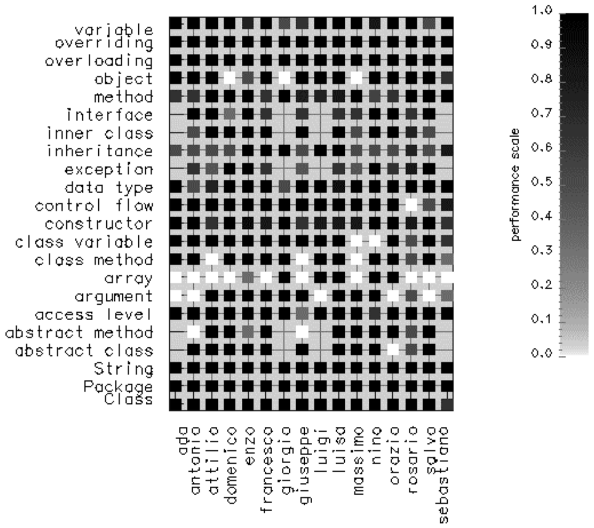


Figure 3: A matrix for visualizing the student's performance on quizzes related to domain

the instructor found vague presentations and quizzes that had to be improved. The matrix also allows local analysis of the performance of a particular student on a specific topic and comparison between students. For example, it can be identified that even though Massimo was successful in opening long discussion treads (see Section 4.2), his performance at quizzes was rather poor. This may point out that this student could be encouraged to pay more attention at reading course material (indeed, the reading time can be analysed with another representation).

4.3 Example of visualizing behavioural aspects of learners

The last example shows the students' access to the course. This is one of the behavioural features required by the instructors. A simple matrix formed by student's names and dates of the course is used to represent the course access, where a corresponding bullet represents at least one access to the course made by the student on the selected date (Figure 4). The instructor has an overview at a glance of the global students' access to the course with a clear identification of patterns and trends. For instance, there is a high concentration of access between the 4th and 23rd of March, corresponding to the group activities, while little access is made at the end of the course. The same image gives attendance information about a specific student, e.g. Michele made little visits to the course material, while Francesco was quite consistent (indeed, as can be seen from the three representations, Francesco was one of the students who were doing very well in the course).

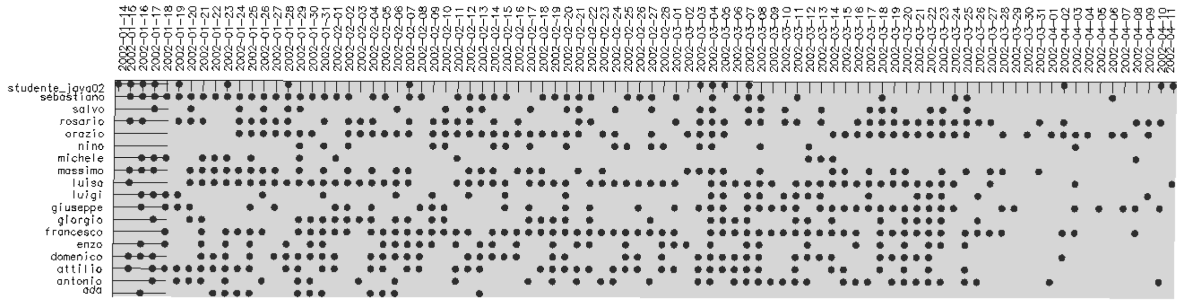


Figure 4: The course access represented in a matrix.

The examples illustrate *micro/macro design* – a method used in Information Visualization for presenting large quantities of data at high dimensions in a way that a broad overview of the data is given. The method enforces local and global comparison [12]. The examples also show the *use of colour* for layering and separation. Example 1 demonstrates also *perspective projection* to present multiple dimensions of data on a 2D flatland [12]. The use of popular visualizations, such as *scatterplots* and *matrixes* [10], is illustrated. Example 1 shows also *rotation* of the visualization to analyse different relationships between the data.

5 Conclusions and Future Work

We have presented here an approach for externalising student information to facilitate instructors involved in distance learning. Applying Information Visualization techniques, we have developed CourseVis – a tool that processes student data collected by web-based course management systems and creates pictorial representations (2D and 3D) to help instructors gain understanding of social, cognitive and behavioural aspects of distance students. The advantages of CourseVis are: tutors and students continue to use the CMS in the usual manner; the information visualised complies with the instructors' requirements gathered from a survey; and the tool can be used independently from a CMS.

Our plans for future work are to carry out several studies to examine the effectiveness of CourseVis. We have conducted a focus group study with several distance learning instructors at Leeds University to examine the appropriateness of the pictorial representations in CourseVis. The study enabled us to gather feedback on the usability and usefulness of CourseVis. Participants pointed out possible uses of CourseVis in distance learning to help teachers identify individuals that need attention, discover patterns and trends in distance groups, and reflect on their teaching practice. The next stage is to distribute CourseVis to distance learning instructors and to gather their feedback whether the tool helped them manage distance courses more effectively.

Acknowledgement. This work has been partly supported by a fellowship of the first author awarded by the Swiss National Science Foundation and undertaken at the Informatics Research Institute, School of Computing, Leeds University.

References

[1] K. Stuart Card, Jock D. Mackinlay, Ben Shneiderman. *Readings in Information Visualization, using vision to think*. Morgan Kaufmann, Cal. USA. 1999.

[2] K. Gratto. *Distance Education Best Practices and Problems*, Report to the Distance Education Evaluation Task Force, North Carolina A & T State University. 1999.

[3] Danita Hartley, Antonija Mitrovic. *Supporting Learning by Opening the Student Model*. Proceedings of ITS 2002, Biarritz.

[4] D. Helic, H. Maurer, N. Scherbakov. *Web Based Training: What do we Expect from the System*. Proceedings of ICCE 2000, Taiwan.

[5] J. Kay. *The UM toolkit for cooperative user modelling*. User modeling and user adapted interaction journal. Vol. 4, 1995.

[6] R. Mazza. *Diagnosing the State of the Student's Knowledge in a Web-Based Learning Environment*. International Conference on Intelligent Tutoring Systems, Young Researchers Track, 2002.

[7] R. Mazza, V. Dimitrova. *Informing the design of a course data visualisator: an empirical study*. To appear to the 5th International Conference on New Educational Environments, May 26-28 2003 Lucerne, Switzerland.

[8] http://www.opendx.org

[9] Christophe Reffay, Thierry Chanier. *Social Network Analysis Used for Modelling Collaboration in Distance Learning Groups*. Proceedings of ITS 2002, Biarritz.

[10] Robert Spence. *Information Visualisation*. Addison-Wesley, 2001.

[11] Leah A. Sutton. *The Principle of Vicarious Interaction in Computer-Mediated Communications.* International Journal of Educational Telecommunications, Vol. 7 N.3 2001.

[12] Edward R. Tufte. *Envisioning Information*. Graphics Press, Cheshire, Connecticut, USA 1990.

[13] Diego Zapata-Rivera, Jim E. Greer. *Exploring Various Guidance Mechanism to Support Interaction with Inspectable Learner Models*. Proceedings of ITS 2002, Biarritz.