

# Interactive Visualization Tools to Improve Learning and Teaching in Online Learning Environments

Kirsi Kuosa<sup>1</sup>, Damiano Distante<sup>2</sup>, Anne Tervakari<sup>1</sup>, Luigi Cerulo<sup>3</sup>, Alejandro Fernandez<sup>4</sup>, Juho Koro<sup>1</sup>, Meri Kailanto<sup>1</sup>

<sup>1</sup>Tampere University of Technology, Tampere (Finland)

<sup>2</sup>Unitelma Sapienza University, Rome (Italy)

<sup>3</sup>University of Sannio, Benevento (Italy)

<sup>4</sup>LIFIA, CIC/FI, National University of La Plata, La Plata (Argentina)

{kirsi.kuosa, anne.tervakari, juho.koro, meri.kailanto}@tut.fi  
damiano.distante@unitelma.it  
lcerulo@unisannio.it  
alejandro.fernandez@lifia.info.unlp.edu.ar

## ABSTRACT

This paper presents two interactive visualization tools for learning management systems (LMS) in order to improve learning and teaching in online courses. The first tool was developed at the Intelligent Information Systems Laboratory (IISLab) at the Tampere University of Technology (TUT). The tool is used to analyse students' activity from automatically recorded user log data and to build interactive visualizations. They provide valuable insights into the learning process and participation of students in a course offered to teachers and students. The second tool was developed at the Unitelma Sapienza University. It extends navigation and search functionalities in the discussion forum of an LMS with a topic-driven paradigm. The tool analyses forum content and automatically identifies discussion topics. It then enhances the original forum with a topic-driven navigation structure and an interactive search graph. Both tools have been developed as plug-ins for the Moodle LMS, but their analysis processes and techniques can be adopted into any LMS.

**Keywords:** e-learning, learning analytics, information retrieval, data mining, information visualization, visualization tools.

## INTRODUCTION

As online learning spreads and becomes pervasive, the need for tools that monitor students' participation in online courses and measure knowledge and skills development has greatly increased. Most learning management systems (LMS) and Web-based solutions that are adopted to offer online courses are able to record rich amounts of information about students' activity, such as communication, collaboration, and participation in online courses. Many LMSs also offer reports and even simple graphs about each student's (elementary) actions. However, these reports are poor indicators of students' activity and participation if they are considered singly (Soller,

2001). Exploiting the log data of LMSs in order to understand the activity in an e-learning environment, how to improve student learning processes, and how to support the process of teaching requires tools and strategies that help to select the right information in a timely fashion and present it at the right location in a correct, clear and effective way (Tufte, 2001; Tervakari et al., 2014).

To achieve this goal, at IISLab<sup>1</sup> was developed a visual analytics instrument called the TUT LA tool. The implementation of the tool was for TUT Circle<sup>2</sup>, an e-learning environment that is built using the Drupal<sup>3</sup> content management system (Silius et al., 2011; Tervakari et al., 2013). In this paper, we present the basic principles of an extended version of this tool, which was re-implemented as a plug-in for the widely adopted Moodle<sup>4</sup> LMS.

Discussion forums are one of the main means of asynchronous communication used in e-learning environments. They are available in all LMSs. Forums are used in online courses to facilitate the learning and teaching processes, such as supporting student-to-teacher interaction, student collaboration, group work and cooperative learning. Discussions held in a forum within a certain period are a potentially useful source of information for any student that subsequently accesses the forum online. However, the success of a forum as an information source and means of communication is determined by the richness of its content. Moreover, its content (discussions and messages) can be navigated and searched. While the first property depends on user participation and use of the medium, the second depends on the navigation structure and the search features provided by the forum.

To support the discovery and retrieval of relevant information in discussion forums in the e-learning environment of distance-learning through the Unitelma Sapienza University, we defined a data analysis process that exploits information retrieval techniques. It employs topic models (Blei, 2011) and formal concept analysis (Ganter and Wille, 1999) to identify the topics discussed in a given forum and to provide topic-driven searches and navigation (Cerulo and Distanto, 2013)(Distanto et al., 2014).

In this paper, we briefly describe these processes and describe the implementations of the plug-ins for the Moodle LMS. They provide interactive visualizations that help both students and teachers find the information they need and monitor students' learning and actions in the learning environment.

## RELATED WORK

Recently, on-line education systems have become widespread tools adopted by both historical and newly founded educational institutions. E-learning and e-teaching are new contexts for education through which large amounts of information are generated and available to users worldwide. A large portion of the available information takes the form of free text without the structure required for automated knowledge retrieval.

Learning analytics provides valuable support for teachers and students to understand learning habits and to obtain rich information about learning and teaching processes. Learning analytics is

---

<sup>1</sup> <http://www.tut.fi/en/about-tut/departments/mathematics/research/iislab/index.htm> (last accessed on 1 October 2014).

<sup>2</sup> TUT Circle is a social media-enhanced learning environment developed by the IISLab of the Tampere University of Technology. TUT Circle promotes students' networking, collaboration and communications in small groups by allowing them to send private messages, chat, create, contribute to and comment on content, exchange opinions, share resources and create communities for different needs. Students can also control the visibility of their information and allow access to the information either for the members of their small group or for all students. [www.tut.fi/piiri](http://www.tut.fi/piiri)

<sup>3</sup> [www.drupal.org](http://www.drupal.org)

<sup>4</sup> [www.moodle.org](http://www.moodle.org)

commonly used to measure, collect, analyse and report user data about learners in order to understand and optimize learning, learning environments and teaching (Siemens and Long, 2011). For example, Carr (2012) reported that a great number of open online courses monitor every user's action, such as pausing a video, increasing feedback speed, responding to quiz questions, revising assignments and participating in forums. The data are then used to analyse student behaviour and test how people learn. Hence, the teacher can tailor the learning environment to fit each student's learning style and needs.

The lack of quality in collected tracking data represents a major challenge to ensure its effective utilization. Moreover, large amounts of data can be difficult to analyze, interpret and understand because processing data requires complex cognitive processes. However, if the data are packed and presented in a correct, clear and effective manner, users can process an incredible amount. When the data presented are complex to perceive and process, adequate visual presentations are mandatory (Berg, 2012). Appropriate visualizations can help to process information, such as by increasing memory and the number of available processing resources, reducing the number of searches for information, enhancing the recognition of patterns, enabling perceptual inference operations and using perceptual attention mechanisms for monitoring (Ware, 2012). A good visualization helps the users understand the data, confirms what the users already know, and provides insights into the data. However, to be understood, the context of the data used should be familiar and interesting to the users.

Historically, data mining has been used to extract knowledge from free text (Baeza-Yates and Ribeiro-Neto, 1999). Knowledge extractions from e-learning systems, particularly from user generated data, was introduced in (Castro, Vellido, Nebot and Mugica, 2007b; Hanna, 2004). Tang and McCalla (2005) investigated patterns of system usage in teachers and students' learning behaviour. Data clustering was suggested to promote group-based collaborative learning and to diagnose students incrementally (Castro et al., 2007a).

Information retrieval techniques consider that texts are represented by a vector of terms that are usually weighted by their occurrence in documents. The availability of different media (e.g., image, video, audio etc.) has recently demanded more sophisticated information retrieval techniques that are capable of treating new kinds of information (Goodrum, 2000; Foote, 1999; Beel et al., 2009). E-learning forums are usually represented as text. Hence, traditional information retrieval techniques are usually adopted to search for relevant information. Association-rules mining has been widely adopted in e-learning, specifically in recommendation systems (Za'iane, 2002; Yang et al., 2010), learning material organization (Tsai et al., 2001), student learning assessments (Romero et al., 2005), course adaptation to the student's behaviour (Hogo, 2010) and evaluation of educational Websites (dos Santos Machado and Becker, 2003). The development of cooperative learning and knowledge sharing inside student groups predominates in recent research trends (Jakobson et al., 2012). In particular, Web technologies take advantage of the opportunities raised by mixing the social and the semantic Web (Ghennane et al., 2012). Dicheva and Dichev (2006) investigated concept maps and their use of navigation in educational contexts. They propose a framework and a set of tools for the development of ontology-aware repositories of learning materials. The idea is similar to our topic-driven navigation structure. In our approach, topics are extracted from free text semi-automatically by leveraging information retrieval techniques which are then validated by the user, whereas concepts have to be manually defined by the authors of the learning materials (Dicheva and Dichev, 2006).

## **THE PROPOSED INTERACTIVE VISUALIZATION TOOLS**

This section presents our interactive visualization tools for improving learning and teaching process in online courses. First is a tool for learning analytics, which analyses log data collected by the learning management system and abstracts from them a set of interactive graphical reports on the activities of teachers and students. The second tool enhances an existing discussion forum in an e-learning environment using a navigation structure and an interactive visualization graph, which enables searching for and navigating forum messages and discussion threads according to their topics. Both tools were developed as plug-ins for the Moodle LMS, but they can be applied in principle to any other LMS.

The tools were designed by utilizing a method of user-centred design. This process involves designing Web-based services that focus on the end users. The design process is reiterative during which the service is repeatedly designed, modified and evaluated. The end users are actively involved in the design and development process to ensure that the product meets the end users' expectations (Preece, Rogers and Sharp, 2002). Information about end users' needs and requirements, and feedback from end users were collected through surveys, one-on-one interviews and group interviews.

First, we present the data analysis implemented by each tool, their architectures and the adopted implementation technology and finally the different visual user interfaces and functionalities that they offer.

### **Data Analysis Processes**

Both tools collect the data for the data analysis processes from Moodle LMS. However, because the processes in both tools have unique features, they are presented separately in detail in the following sub-sections.

#### *Data Analysis Process in the TUT LA Tool*

The TUT LA tool, which takes the approach of visual analysis in discovering students' activity and participation, follows a data visualization pipeline developed by the IISLab in which user interaction is applied in data filtering and visual transformations. Figure 1 shows this process. Because the tool consists of multiple views, the process mainly differs in the pre-processing part depending on the type of visualization the user has chosen.

1. *Data pre-processing.* First, the raw log data that represents the users' activity in the LMS are collected and analysed. They are then enriched with different data sources that describe users and content artefacts. Integrated in the data collecting process, the definition of context is added to define what kinds and which parts of the data are collected.
2. *Filtering and grouping.* The prepared data are set up and grouped for selection. Scales and thresholds are defined according to the data. This step is revisited when users choose filters within the visualization system.
3. *Visual mapping.* The set-up data are transformed to visual structures.

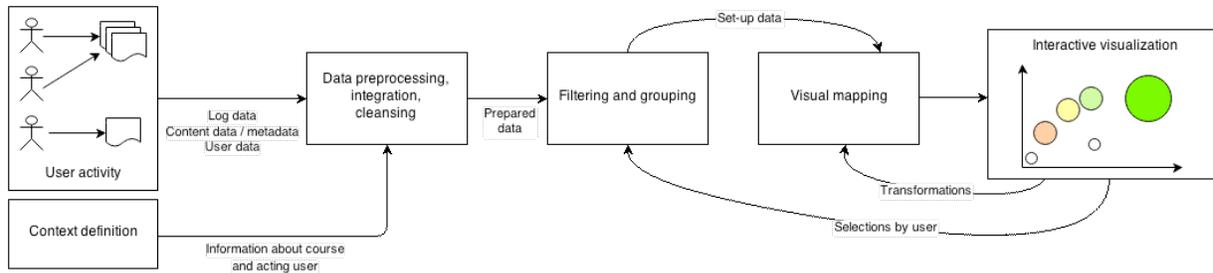


Figure 1. The analysis process for introducing LA functionalities in Moodle

### Data Analysis Process in the TDForum

The process of topic-driven forum enhancement was recently introduced by some authors in Cerulo and Distanto (2013) and is represented in Figure 2. Discussion topics are extracted from the forum's textual content using a semi-automatic information retrieval analysis process. It is based on a statistical technique that is able to extract from a corpus of documents, sets of terms, known as topics, which co-occur frequently. It can also produce a lattice of concepts (discussion topics) and derive a concept hierarchy from a collection of objects and their properties.

The enhancement process consists of four main steps: 1) *Data pre-processing*, 2) *Discussion topics identification*, 3) *Topics lattice generation*, 4) *Documents-to-topics assignment*.

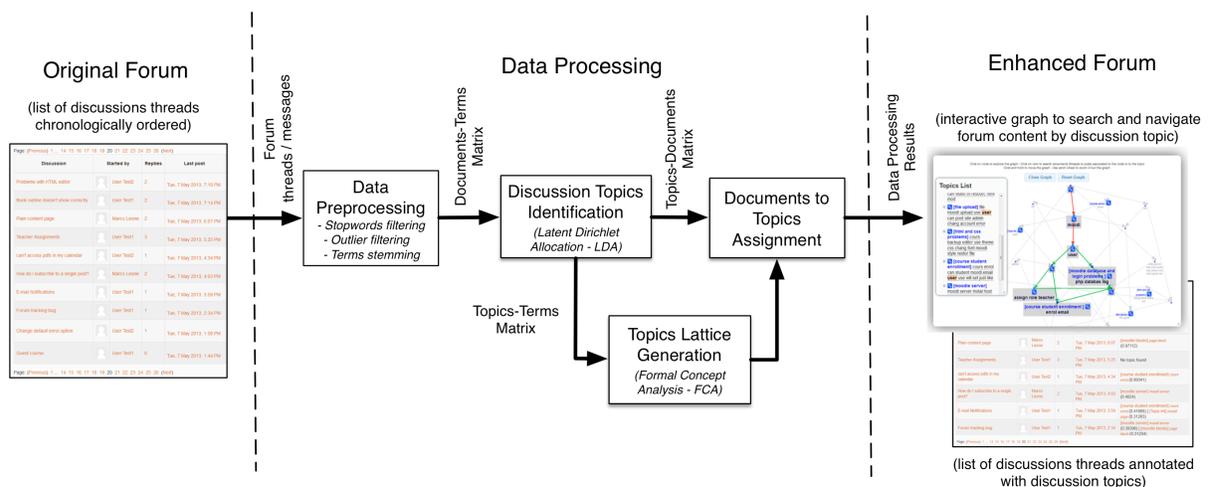


Figure 2. The data analysis process used to introduce a topic-driven search and navigation paradigm in a forum

1. *Data pre-processing*. Forum messages, represented as vectors of indexing terms, are filtered by a standard text analysis pipeline that is usually adopted in information retrieval, which includes: outlier filtering, stop words filtering, and stemming. A document-term matrix is generated.
2. *Discussion topic identification*. The document-term matrix obtained from the previous step is factorized into a topic-document matrix and a topic-term matrix by means of topic models. Intuitively, each topic is generated by a combination of semantically related terms, and each document is represented by a combination of topics.

3. *Topics lattice generation.* The topic-term matrix is used to prune a lattice of topics by using Formal Concept Analysis (FCA), which provides a computational way to derive a concept hierarchy from a collection of objects and their properties.
4. *Documents-to-topics assignment.* Forum messages and discussion threads are associated with discussion topics based on a similarity probabilistic score and a threshold.

## **Architectures and Technology**

The architecture in the tools can be divided into client-side (i.e., front-end) and server-side (i.e., back-end) components. The server-side has the main responsibility in the data processing phase, which is described in detail in the previous section. The client-side focuses on visualizing the data for the user using different JavaScript libraries, for example. This section describes the technical architectures of the tools in detail.

### *Architecture of the TUT LA tool*

The TUT LA tool consists of two separate components: a server-side module, which retrieves and processes the data, and a set of client-side visualization scripts that map the data, render the visual representations in the browser and provide interactivity. Figure 3 depicts the architecture of the TUT LA Moodle plug-in.

Because the client-side visualization components used in the first implementation of the TUT Circle are reusable, they can be applied to different LMSs, such as Moodle. The data models of student activity used in the TUT Circle can also be built in Moodle. This facilitated the re-use of the visualization components used in the TUT Circle in developing the first version of the TUT LA Moodle plug-in. In both environments, the server-side component performs the following: 1) defines where and how the visualization view is shown (e.g., some elements may be disabled depending on the user's role and visualization settings); 2) analyses, processes and cleans log data collected by LMS; and 3) prepares it for use by JavaScript libraries in the Document Object Model (DOM). In Drupal and Moodle, the log data and other data needed for representing the user activity are stored in a relational database. Different visualizations use distinct data models because they represent user activity from different points of view. To build a data model that provides information about student activity, the data must often be heavily processed. The cleansing of data is needed to remove unnecessary records, such as irrelevant activity and, in some cases, to anonymize users. In Drupal, laborious queries, such as acquiring users' content view counts, are carried out hourly by utilizing a Unix cron job. The results are saved to the database for later use. This feature is currently being developed in the TUT Moodle plug-in to decrease loading times when a visualization utilizing this type of data is accessed.

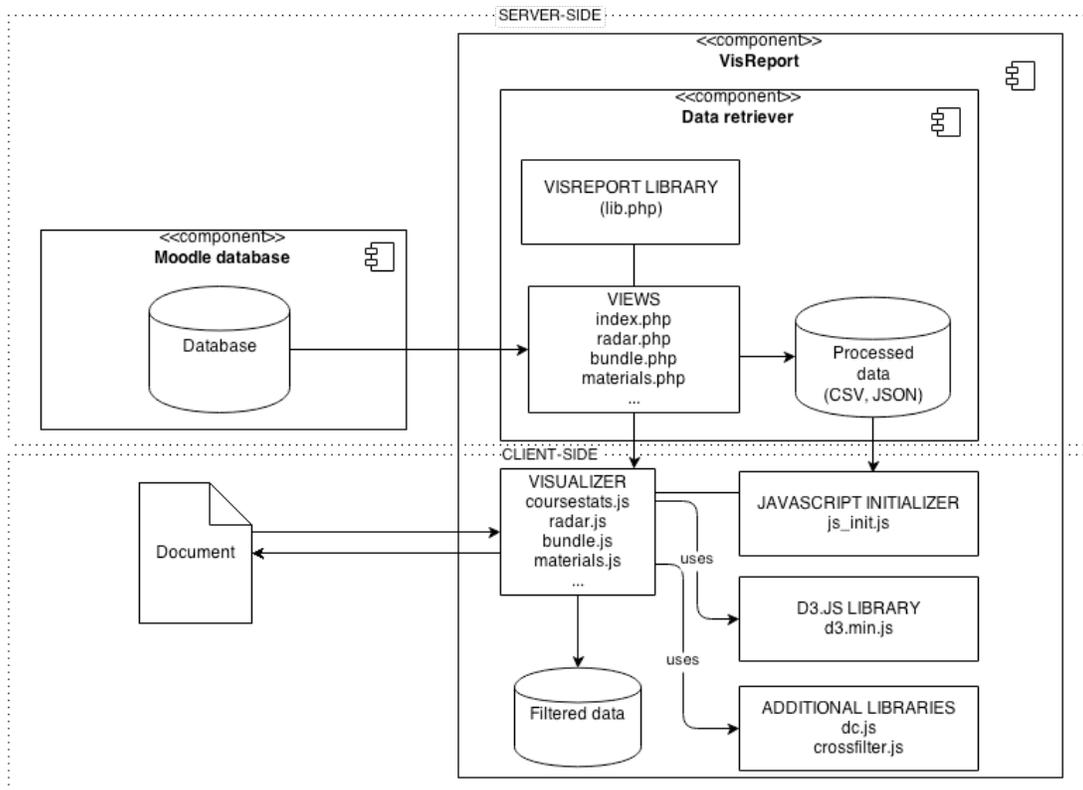


Figure 3. The architecture of the TUT LA Moodle plug-in

A report-type Moodle plug-in, the TUT LA tool is built so that different visualization views are available in the navigation system within the Moodle course. When a view is loading, depending on type of visualization, a PHP script is run to retrieve and pre-process the course data and to initialize it in a DOM available for JavaScript libraries. In basic form, the activity data are in CSV or JSON format in which each row represents a data point, such as the properties of a user linked to an activity about content creation or a page view. Some additional data, such as course creation date, are used to scale the visualization.

The main interactive visualization views of the tools are built with D3.js (<http://d3js.org/>), a JavaScript library used to manipulate documents with data (Bostock, 2011). In a narrative visualization view, Raphaël (<http://raphaeljs.com/>) is also used for its animation capabilities (see Figure 9). Among the different Web-based visualization libraries and tools available, D3 provides the flexibility and scalability needed to build different types of interactive visualizations and to use existing solutions to be applied in the context of learning analytics. With the library, input data initialized from the LMS can be bound to document elements by applying dynamic transforms to generate and modify DOM content. In this case, visualizations are displayed in scalable vector graphics (SVG) format. The ability to interact with visualizations is built in visualization scripts that which filter the data and transform data-bound DOM elements according to the selections and filters the user has chosen.

The TUT Moodle plug-in enables the visualization of learning log data in the context of a single course with functionalities similar to the visualizations in the TUT Circle, but it is strongly focused on courses which are heavy on forum activity, i.e. where students report their assignments in forums and comment on them. So far the plug-in uses mainly log, forum, forum posts,

forum discussions and user tables in the Moodle database to retrieve the activity data. However, an administrator interface that will facilitate the selection of the variables used in the visualization is in under development. This interface will make it possible to modify these visualizations to represent different kinds of activity data in addition to the data related to discussion forums.

### *Architecture of the TDForum*

The architecture of the TDForum tool, which provides content navigation and search enhancement in a forum, is represented in Figure 4, which shows the distinction between the front-end and back-end components. The front-end represents the graphical interfaces with which users interact, while the back-end includes components that are in charge of implementing the data analysis process represented in Figure 2, as well as and interacting with the Moodle database.

Regarding the front-end plug-in, TDForum offers users two graphical user interfaces (GUIs): Main View and Discussion Topics View. Main View presents the forum content in as a chronologically ordered list of discussions, each of which is augmented by the set of discussion topics identified in it and the associated similarity scores (blue-dashed rectangle in the right side of Figure 11). The Discussion Topics View presents an interactive visual navigation graph that shows the set of discussion topics and associated terms discovered in the information retrieval analysis process (Figure 10). The latter view is built using the JIT JavaScript InfoVis Toolkit (<http://philogb.github.io/jit/>), which is an open-source library of JavaScript's classes and objects for the interactive visualization of data within Websites. JIT is based on MooTools (<http://mootools.net/>), which extends the JavaScript language by a set of constructs for the management and inheritance of classes and objects, making it similar to an object-oriented language. This allows developer to have a set of default classes that model different modes of viewing data: graphs (Area Chart, Bar Chart, Pie Chart), trees (TreeMap, SpaceTree, HyperTree) or further graphs (Radial Graph, Force-Directed Graph). Each class inherits the basic core functionality of the library and extends it by the behaviour of the adopted visualization mode. The user can then re-implement some features that are typically associated with events on the graph (e.g., on BeforePlotNode and on CreateLabel) or customize parameters without having to modify the source code. The fundamental characteristic of each view is the non-use of external images: all the elements needed to display the data objects that are implemented as a canvas for rendering a dynamic bitmap. In the case of the TDForum, we adopted the visualization model Radial Graph (RGraph). In this visualization mode, it is possible to centre the graph on one of the nodes that comprise it, which is selected by the user via a single click, and then access the remaining nodes in successive layers that consist of concentric circles (which can be displayed or not) depending on the distance (defined as the number of arcs) from the selected node. The graph is populated with an Asynchronous JavaScript and XML (AJAX) call by a php script that takes care of drawing the necessary data from the database and converting them to the JavaScript Object Notation (JSON) format required by the JIT library. The obtained information is then passed to a JavaScript function that initializes the graph and renders the arcs and nodes that comprise it. The adopted RGraph visualization model was customized in order to adapt it to the context of the usage. First, we re-implemented the functions that render nodes in order to distinguish two different types of nodes contained in the generated topic lattice: the root node and the intermediate nodes. Two functions were implemented for the recursive search of arcs that connect a selected node to the root and the leaves of the graph. These functions are used in the function that renders edges in order to highlight the inbound and outbound paths of a node in different colours.

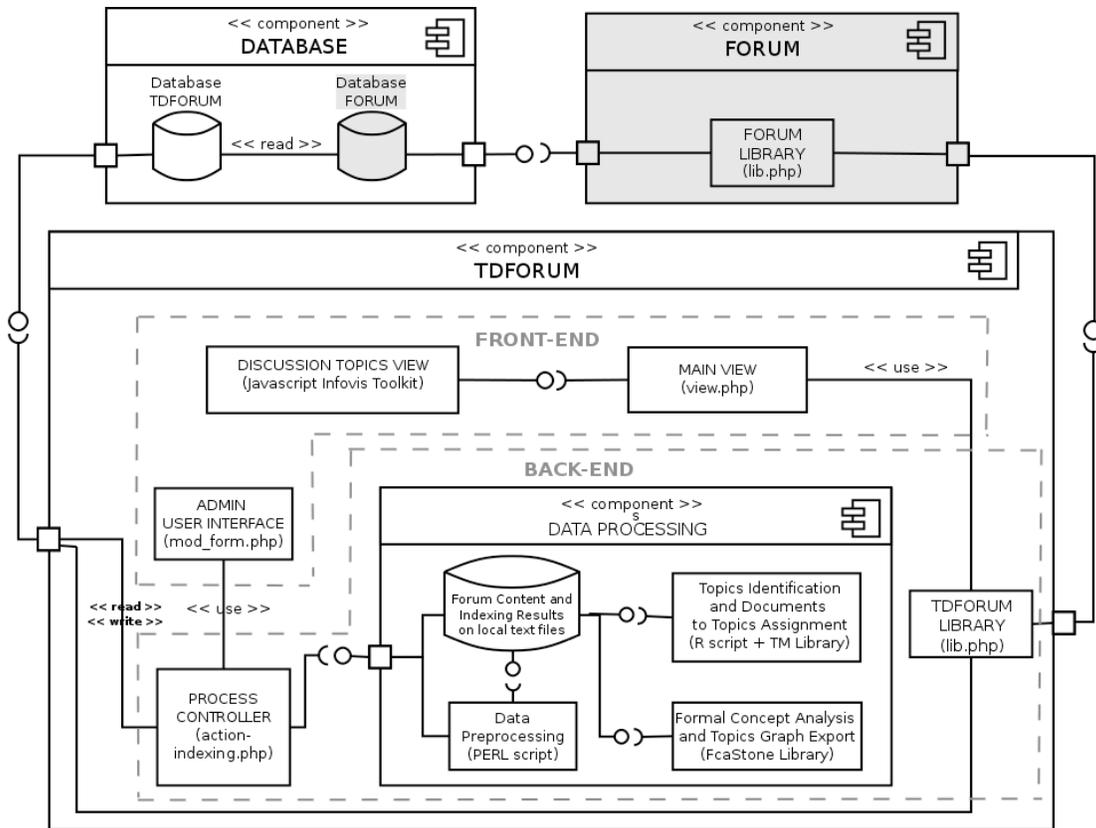


Figure 4. The architecture of the TDForum Moodle plug-in

### Visual User Interfaces and Functionalities

TUT LA tool and TDForum plug-ins enhance Moodle with visual user interfaces. They provide users and easy way to access the information provided by the tools.

#### Visualizations in TUT LA tool

All the visualizations of the learning log data available in the TUT LA Moodle plug-in are interactive, which allows the users to investigate data and help them better understand it. Four main perspectives are used in the visualizations: represent students' activity (for example Figure 5), help navigate or find relevant contents (e.g., Figure 6), show the effectiveness of the activity according to scores (Figure 8) and the visualizations that motivate students to participate and be active (Figure 9). The interactive visualization dashboard represents the activity and participation of each member of the student group (Figure 5). The dashboard consists of three types of interactive visualizations that are connected by means of linking and brushing (i.e., filtering made in one visualization automatically reflects in the other visualizations) (Keim, 2002), as follows: 1) set of interactive live histograms that users click and drag in order to filter information by dimensions, such as time, unique viewers per content, total views per content, and date; 2) a pie chart representing types of user content; and 3) a bubble chart in which each bubble represents a different student.

In the bubble chart, each bubble indicates a different student. The properties of each bubble are determined according to the values that indicate the students' activity. The position of the bubble depends on the value of the x-axis (content read) and the value of the y-axis (amount of content produced). The size of the bubble represents the number of words in the content that were produced by the student, and the colour of the bubble separates passive students from active students. The content created by the students is listed below the visualization, and the list is updated in real time depending on the selected filters.

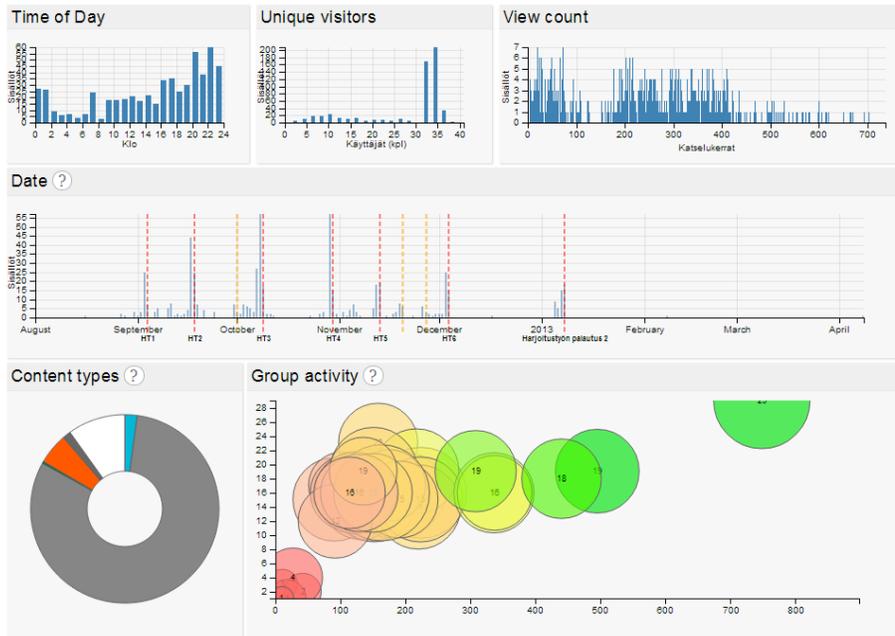


Figure 5. The interactive visualization dashboard consists of three types of visualizations representing the activity and participation levels of each member of student group

The visualization dashboard (Figure 5) provides the students with the possibility of comparing their own activity with those of other students. It also helps teachers to monitor the evolution of the students' activity and participation over time and identify passive students who may be at risk of failing or dropping out (Silius et al., 2013; Tervakari et al., 2014).

The bundle visualization represents the relationships between the students, the topics of the learning material, the Web resources referenced, and the discussions in which the students have participated (see Figure 6). Forum messages are crawled in order to automatically collect meaningful keywords and find URLs, which are then associated with the corresponding users and discussions. The visualization helps the teacher to monitor how the students use various parts of the learning material and the external resources as references. This can also help teachers evaluate and improve the quality of their online learning material. Being able to see the references that other students have used can help the students to find new references, and motivate them to search for new references for use in their own works (Tervakari et al., 2014).





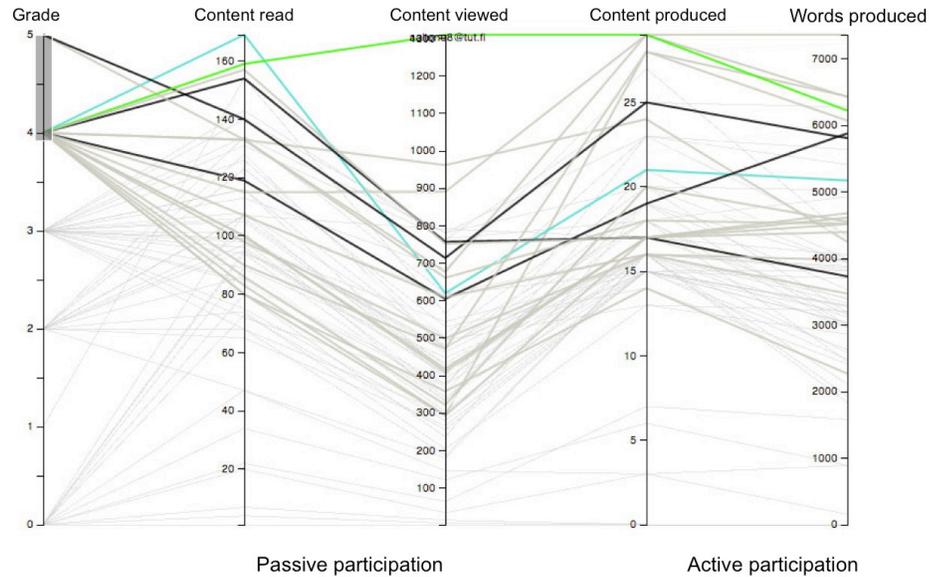


Figure 8. This interactive parallel coordinates visualization represents relationship between the students' academic performance (grade) and participation. The grades 4 and 5 are selected when visualization shows the levels of the students' active and passive participation related to those grades.

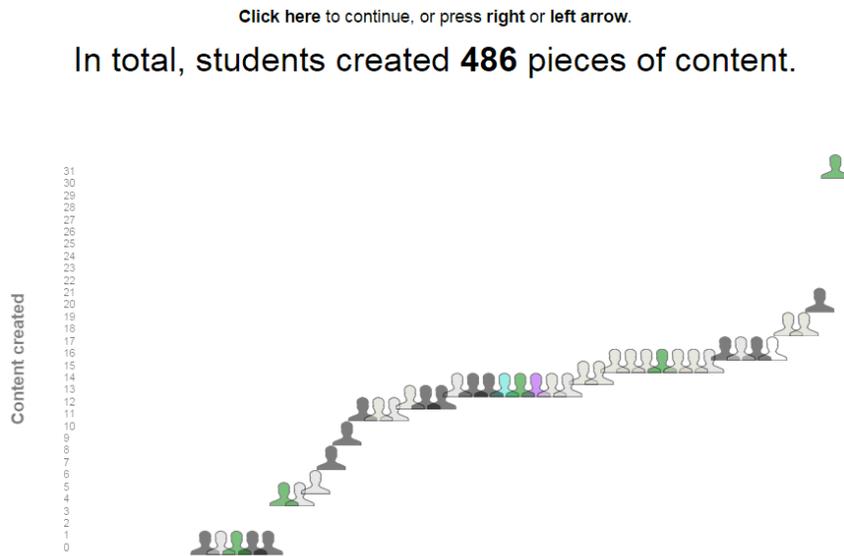


Figure 9. Narrative visualization tells a story using data about the students' activities during the course.

### User interfaces in TDForum

The interactive visual navigation graph offered by the TDForum Moodle plug-in enables a new topic-driven navigational paradigm with the content of a forum. The user can navigate through

the identified forum discussion topics by panning and zooming the graph, and mining the hierarchical relations between a selected topic and its ancestors (macro-topics) and descendants (sub-topics). When the user finds a topic of interest, by simply clicking once, she or he can retrieve the list of discussions or messages associated with it. The list of returned discussions or messages is ordered by a similarity score with respect to the selected topic. Figure 10 shows an example of the interactive discussion topics visualization graph. Figure 11 shows the enhanced version of a Moodle forum featuring the list of discussion topics associated with each forum discussion thread, and provides hyperlinks to similar or related discussions.

The TDForum navigational paradigm is similar to semantic networks, that is, networks that represent semantic relations between concepts (Sowa, 1987). The main difference is that in our case topics are extracted from data by using a statistical model automatically. Instead, semantic networks are usually built from scratch and are maintained manually. Moreover, we adopt information retrieval techniques to enhance the information retrieval capability of semantic networks. This is evident when a user reaches a concept in the graph and inspects the messages ordered by using an information retrieval scoring scheme similar to the classical vector model (Baeza-Yates, 1999).

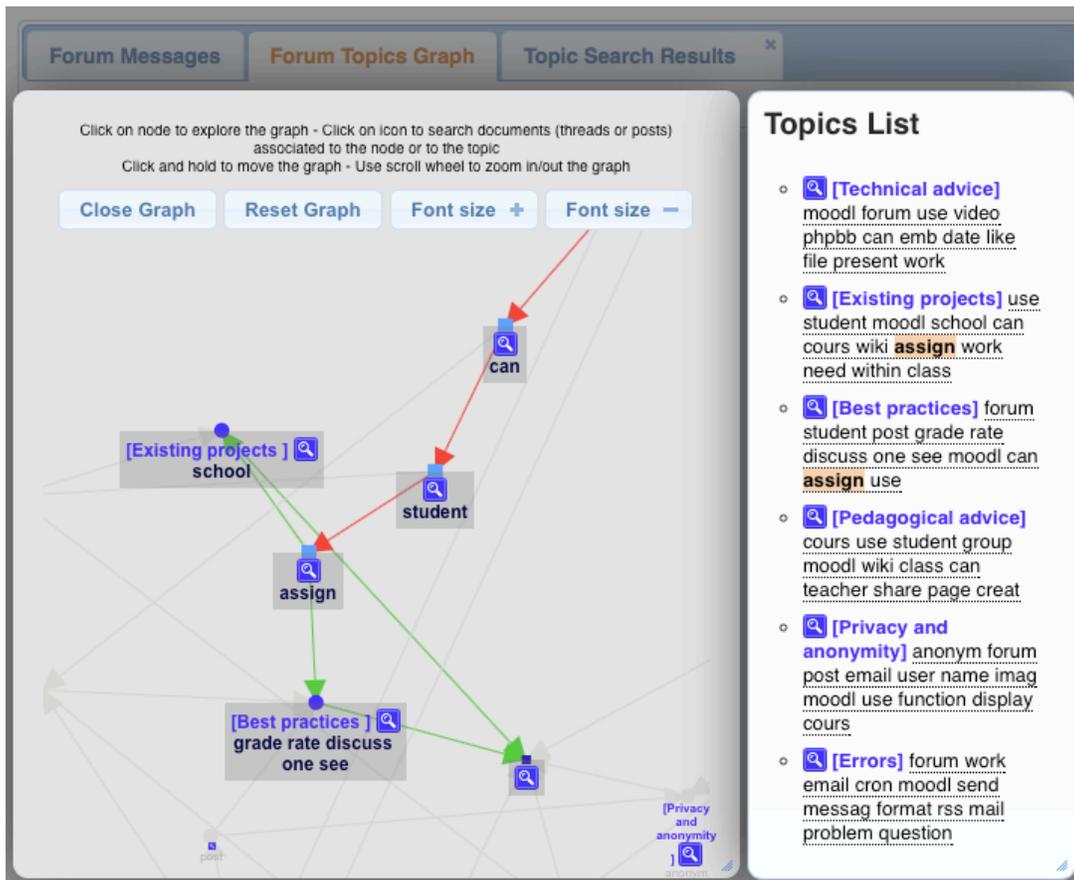


Figure 10. The interactive search and navigation discussion topics graph recovered from a forum in Moodle

Discussion	Forum	Started by	Replies	Last post	Discussion topics
how to use Moodle to organize and manage a large class?	Pedagogical issues	anonfirstname4 anonlastname4	3	anonfirstname4 anonlastname4 Wed, 26 Mar 2014, 2:05 PM	[Pedagogical advice] <i>cours use student group moodl wiki</i> (0.99974)
Anyone using Moodle in large (150+) classes?	Pedagogical issues	anonfirstname2 anonlastname2	1	anonfirstname2 anonlastname2 Wed, 26 Mar 2014, 2:04 PM	[Existing projects] <i>use student moodl school can cours</i> (0.99967)
Messages work but forum notifications do not work	Technical issues	anonfirstname2 anonlastname2	15	anonfirstname3 anonlastname3 Wed, 26 Mar 2014, 2:03 PM	[Errors] <i>forum work email cron moodl send</i> (0.99988)
How to integrate phpbb or Vbulletin in moodle	Technical issues	anonfirstname3 anonlastname3	3	anonfirstname4 anonlastname4 Mon, 24 Mar 2014, 12:46 PM	[Technical advice] <i>moodl forum use video phpbb can</i> (0.98594)
Integrate phpbb in	Technical	anonfirstname1	7	anonfirstname3 anonlastname3 Mon, 24 Mar	[Technical advice] <i>moodl</i>

Figure 11. A Moodle forum enhanced with the topic-driven navigation structure

## QUALITATIVE ASSESSMENT OF THE PROPOSED TOOLS

From a practical viewpoint, the quality of the tool or any other product can be determined by evaluating how useful the tool is in practice. To be useful, the tool must be usable and have high utility. If the tool is usable, it is easy and efficient, easy to remember, has few errors and it is subjectively pleasing. The tool has high utility if it supports users in achieving their goals. (Nielsen, 1993.) To ensure the quality and effectiveness of the tools different quality assessments have been made.

### *Qualitative assessment for TUT LA tool*

To study the perceived usefulness of the interactive visualizations provided by the TUT LA tool from the students' point of view, we conducted a survey among hypermedia students regarded as early adopters in different hypermedia courses that were offered in autumn 2012 at TUT. We used a Web questionnaire (Silius et al., 2013). Another survey was conducted among hypermedia students in spring 2013. In addition, during the course, a Web analytics platform called Mixpanel was used for collecting information about how the end users use the interactive visualization tools. (Tervakari et al., 2014).

In both cases, the students were asked to rate the usefulness of the visualizations provided using a seven-level semantic with contrasting adjectives at each end (0 = totally useless, 6 = very useful). The survey instrument included open-ended questions that were developed to gather information about the reasons they considered the particular visualization useful or useless. The statistical data from the surveys were analysed, and the frequency distributions were counted. The qualitative data from the open-ended questions were analysed using content analysis, a

method used to examine the meanings, themes and patterns that were manifest or latent in the students' answers, and identifying similarities and differences among the answers. researcher triangulation was used to ensure the validity of the analysis (Silius et al., 2013; Tervakari et al., 2014.)

According to the results of the first survey (42 respondents), one of the most useful visualizations was the narrative visualization (Figure 9). Two thirds of the students (66%) considered it useful. According to the respondents' opinions, narrative visualization presented information in an enjoyable, interesting format that was easy to understand. The narrative visualization was also said to offer a good overview of the evolution of students' activity during the course. The students were also asked to describe the kind of story the visualization presented. One student summarized the feeling of the visualization: 'Once upon a time there was a course in which there were students and a few lecturers. In the course were a few hyperactive students one of whom was a superman. If that was not you, shame on you!' Most students (93%) said that the narrative visualization told a story that used data about the activity and participation of individual students and their evolution during the course. The purpose was to motivate students to evaluate their activity and participation and to make strategic decisions to improve their performance, if needed. (Silius et al., 2013). The story represents the information in a way that is easy to understand, remember, and share, which then supports collaboration and effective decision-making. Because narrative visualizations focus on presentation rather than exploration and analysis, they can be considered communication media (Kosara and Mackinley, 2013).

According to the results of the second survey, 64 respondents said that the most useful visualization was the interactive bundle visualization (see Figure 6) which represented the relationships between students, the topics of the learning material, the Web resources referenced and the discussions in which the students participated. Forty-six per cent of the students considered the visualization useful. According to the students, the visualization helped them find new and interesting references to use in their own work. Almost half the students (47%) thought that the circle package visualization was useful (Figure 7). The visualization represents the hierarchically structured content of the online learning material. This visualization helped the students obtain an overall view of the topics discussed in the learning material, navigate the material, and search for interesting information in the learning material, thus providing true practical value for the students. According to the results of the survey, most students seemed to prefer the visualizations that offered concrete and practical support in their studying, such as by helping them to find new information to add to their own works, instead of the visualizations that helped them evaluate their own activity and participation. However, the results of monitoring the actual use of the visualization by Mixpanel showed that the interactive visualization dashboard (Figure 5) representing activity and participation of the students was used the most often. In most cases, the students chose a single student from the bubble chart (Tervakari, Silius, Koro, Paukkeri and Pirttilä, 2014.)

To study the usefulness of the interactive visualization based on the learning analytics from the teachers' point of view, a research study was conducted, in which the data collection was performed in two stages. First, in the autumn of 2013, twenty-seven teachers and academic coordinators participating in the Peda Forum Days workshop in Finland were asked about the advantages and challenges related to the educational use of visualizations based on user data. At the beginning of the workshop, the participants were asked to discuss the advantages and challenges of visualizations in small groups. Next, the information about the participants' opinions about the advantages and challenges of visualization were collected by using the focus group method (Ter-

vakari, Silius, Koro, Paukkeri and Pirttilä, 2014). The results indicated that the visualizations (Figure 5) representing information about students' activity and participation would provide valuable information to teachers about how to improve the implementation, structure and schedule of the course. It would also help teachers identify students who might be at risk of failing or dropping out and provide them with personal guidance at the time it is needed. The results showed that the students liked the visualization (Figure 6) presenting information about the references used by other students. According to the results, the visualization could help the teacher monitor whether the students were using the learning materials and the external resources or not. The visualization could also help teachers to identify the parts of the learning material that the students did not use because it was too challenging (Tervakari, Silius, Koro, Paukkeri and Pirttilä, 2014.)

Additional data were collected in a diary study, which is a method for collecting information about users' experiences over time in the actual context of interaction with a system, which can reveal real user needs (Preece et al., 2002). Two researchers who have been involved in designing learning analytics visualization for TUT Circle used the TUT LA tool implemented in TUT Circle, which they used as a learning environment on their online courses from 2011 to 2014. The researchers were asked to record activities and events related to the use of interactive visualizations in a diary. They were also asked to compare the use of the interactive visualization to the use of default Moodle reports and to record their daily experiences in using the visualizations during their online courses (Tervakari, Kuosa, Koro, Paukkeri and Kailanto, 2014). Content analysis was used to analyse the diaries. According to the results, a main benefit of the interactive visualizations provided by the TUT LA tool was that they offered the possibility to monitor students' participation easily and effectively in real time during the online course, unlike the reports provided by Moodle. For example, Moodle offers user activity reports and course statistics based on users, times and activities. However, the utilization of the reports is difficult because the reports and statistics are unrelated and full of detailed data. In addition, Moodle does not provide any support in interactions with the data, which makes further investigation and analysis of the data difficult or even impossible. However, the interactive visualizations provided by TUT LA tool can offer teachers the possibility of investigating data from different viewpoints, thus helping them to gain better understanding of students' behaviour during the online course. For example, Moodle does not support teachers in comparing students' participation and activity levels or monitoring the development of the students' participation during the course. According to the teachers, a main benefit of using TUT LA tool is that it helps them to establish whether the students' learning is proceeding as intended, and to adjust the instructional design when required. For example, the information about students' last-minute behaviour in returning their weekly assignments just before the deadlines helped teachers to adjust the schedule of the course to support peer learning during the course (Tervakari, Kuosa, Koro, Paukkeri and Kailanto, 2014.)

### *Usefulness of the TDForum tool*

The usefulness of the TDForum tool and the topic-driven content search and navigation paradigm were also verified by the results of a controlled experiment. We compared (i) the number of items (forum posts) the user had to inspect, and (ii) the time spent in accomplishing a set of information search tasks with and without using the TDForum tool in a Moodle forum. The search tasks were performed using two different (but complementary) approaches: (a) full-text search, and (b) topic-driven navigation. The full-text search was accomplished using the default full-text search engine implemented in the Moodle platform. In this system, input query key-

words are by default linked with the AND operator, and the system retrieves a list of all posts containing the keywords used in the search.

The second approach was used the TDForum Moodle plug-in. While performing the tasks, we collected the number of items inspected and the time needed to achieve the information search goal. In the Moodle full-text search, the number of inspected items was computed by counting the number of posts examined before finding the expected posts. In the topic-driven navigation approach, the number of inspected items was the sum of two quantities: the number of links followed to reach the closest topic in the Discussion Topics View of the TDForum plug-in and the number of posts examined before finding the correct posts.

The findings showed that in general the number of items inspected using the full-text search was on average higher than the number of items inspected using TDForum. Moreover, the time necessary to obtain the correct answer was on average less in the TDForum. The reason is that in the full-text search more time was spent in choosing the correct search keywords. These results confirmed that the TDForum tool facilitates and improves information searches in the discussion forums on a Moodle platform.

## **CONCLUSIONS AND FURTHER RESEARCH**

The interactive learning data visualizations provided by the TUT LA tool can offer concrete and practical support for students in their learning activities in online courses. They can help students to monitor and evaluate their performance processes and learning outcomes in order to make strategic adjustments to improve, which could bring real benefits to students. Especially on massive open online courses (MOOCs) that require students to be responsible for their learning, self-regulation and other learning skills.

Teachers of online courses can also benefit from visualizations in many ways. For example, the visual analytic tools, such as the TUT LA tool, could help teachers to make decisions regarding pedagogical strategies, instructional guidance, actions and interventions that may be used to support student participation and activity. In order to support teachers in obtaining increased practical value by using the visual analytics tool, the controls used to specify data and manipulate views need further development. Visual analytics could also be used to retrieve valuable information about how students use the course platform, how they study subtopics and how they use learning materials and other resources, such as how much time students spend in watching a video or reading a chapter of the learning material, in which order students study the learning materials offered, and which parts of the assignments students have completed. This kind of information can help teachers to evaluate students' learning performance and monitor their progress during courses. In MOOCs with thousands of students, the TUT LA tool could offer valuable practical support for teachers, such as identifying students at risk of failing or dropping out. The information can also be used to identify topics that students skip and learning materials that they do not use, which may help the teacher to improve the instructional design of the course

Effective search and navigation capabilities associated with discussion forums in online courses are decisive in making forums a useful information source for students, as well as an additional means of communication. Moreover, according to earlier research, students prefer tools and information that can provide them with practical support in analysing and examining textual information, such as messages on discussion forums, in order to discover new information (Silius et al., 2013; Tervakari et al. 2014).

In this paper, we presented two tools for use with Moodle LMS: a tool for learning analytics and a tool for enhancing content searches and navigation in a forum. Both tools were designed to support students and teachers in e-learning environments.

IISLab at TUT in Finland, Unitelma Sapienza University in Italy, and the University of La Plata in Argentina have started a collaborative research that aims to extend and evaluate the usefulness of their visual analytics tools and forum enhancement approach by means of a set of controlled experiments replicated in different classrooms. We expect the measured advantages to be twofold: i) information is retrieved in less time, and ii) contents stored in on-line forums are managed and tracked in a more effective manner.

## REFERENCES

- Baeza-Yates, R. A. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Beel, J., Gipp, B., Stiller, J. (2009). Information Retrieval On Mind Maps - What Could It Be Good For?. *5th International Conference on Collaborative Computing: Networking, Applications and Worksharing* (CollaborateCom'09). Washington, DC: IEEE.
- Berg, M. (2012). *Human abilities to perceive understand and manage multidimensional information with visualizations*. Doctoral dissertations 30/2012. Helsinki: Unigrafia Oy.
- Blei, D. M. (2011). Introduction to probabilistic topic models. *Communications of the ACM*, 55(4), pp.77-84, ACM Press.
- Bostock, M., Ogievetsky, V., and Heer, J.. (2011). *D<sup>3</sup> Data-Driven Documents*, IEEE Transactions on Visualization and computer graphics.
- Carr, N. (2012, September 27). *The crisis in higher education*. MIT Technology Review. <http://www.technologyreview.com/featuredstory/429376/the-crisis-in-higher-education/>. Last accessed on August 15<sup>th</sup>, 2014.
- Castro, F., Nebot, A., and Mugica, F. (2007a). Extraction of logical rules to describe students' learning behavior. *6th International Conference Web-Based Education - Volume 2, WBED'07*, pp. 164–169, Anaheim, CA, USA. ACTA Press.
- Castro, F., Vellido, A., Nebot, A., and Mugica, F. (2007b). *Applying data mining techniques to e-learning problems*. In Jain, L., Tedman, R., and Tedman, D., editors, *Evolution of Teaching and Learning Paradigms in Intelligent Environment*, volume 62 of *Studies in Computational Intelligence*, pp. 183–221. Springer Berlin Heidelberg.
- Cerulo, L. and Distante, D., (2013) Topic-driven semi-automatic reorganization of online discussion forums: A case study in an e-learning context. In Proc. of the 2013 *Global Engineering Education Conference (EDUCON)*, pp. 303-310, IEEE.
- Dicheva, D. and Dichev, C. (2006). Tm4l: Creating and browsing educational topic maps. *British Journal of Educational Technology*, 37(3), pp. 391–404.
- Distante, D., Cerulo, L., Visaggio, C.A., Leone, M. (2014), Enhancing Online Discussion Forums with a Topic-Driven Navigational Paradigm: A Plugin for the Moodle Learning Management System". In Proc. of the *6th International Conference on Knowledge Discovery and Information Retrieval (KDIR) 2014*, pp. 97-106. Scitipress Co.

- dos Santos Machado, L. and Becker, K. (2003). Distance education: A web usage mining case study for the evaluation of learning sites. In *2003 IEEE International Conference on Advanced Learning Technologies (ICALT 2003)*, 9-11 July 2003, Athens, Greece, pp. 360–361. IEEE Computer Society.
- Foote, J. (1999). *An overview of audio information retrieval*. *Multimedia Systems*, 7(1), pp. 2-10, Springer.
- Ganter, B., and Wille, R. (1999). *Formal concept analysis: mathematical foundations*. ISBN 978-3-642-59830-2, Springer.
- Ghenname, M., Ajhoun, R., Gravier, C., and Subercaze, J. (2012). *Combining the semantic and the social web for intelligent learning systems*. In Proc. of the 2012 *Global Engineering Education Conference (EDUCON)*, 2012, pp. 1–6, IEEE.
- Goodrum, A. A. (2000). Image Information Retrieval: An Overview of Current Research. *Informing Science* 3(2).
- Hanna, M. (2004). Data Mining in the e-Learning Domain. *Campus-Wide Information Systems*, 21(1), pp. 29–34. ISSN: 1065-0741, Emerald Group Publishing Ltd.
- Hogo, M. A. (2010). Evaluation of e-learning systems based on fuzzy clustering models and statistical tools. *Expert Systems with Applications*, 37(10), pp. 6891–6903, Elsevier.
- Jakobsone, A., Kulmane, V., and Cakula, S. (2012). Structurization of information for group work in an online environment. In Proc. of the 2012 *Global Engineering Education Conference (EDUCON)*, pp. 1 –7, IEEE.
- Keim, D.A. (2002). Information Visualization and Visual Data Mining, *IEEE Transactions on Visualization and computer graphics*. 8(1), Jan/Mar 2002, IEEE Computer Society.
- Kosara, R. and Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. *IEEE Computer (Special Issue on Cutting-Edge Research in Visualization)*, 46(5), pp. 44–50.
- Nielsen, J. (1993). *Usability Engineering*. Academic Press.
- Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. New York: John Wiley & Sons.
- Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning*. Retrieved from <http://www.eric.ed.gov/PDFS/EJ982976.pdf>
- Romero, C., Ventura, S., and Bra, P. D. (2005). Knowledge discovery with genetic programming for providing feedback to courseware authors. *User Modeling and User-Adapted Interaction*, 14(5), pp. 425–464.
- Siemens, G. and Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, vol. 46, no. 5. Retrieved from <http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education>
- Silius, K., Tervakari, A-M., and Kailanto, M. (2013). Visualizations of User Data in a Social Media Enhanced Web-based Environment in Higher Education. *International Journal of Emerging Technologies in Learning* 8(2), pp. 13-19.

- Silius, K., Tervakari, A.-M., Huhtamäki, J., Tebest, T., Marttila, J., Kailanto, M. & Miilumäki, T. (2011). *Programming of Hypermedia – Course Implementation in Social Media*. 2nd International Congress on Computer Applications and Computational Science. Advances in Intelligent and Soft Computing, 2012, Volume 144/2012, 369-376. Springer Berlin / Heidelberg
- Soller, A. (2001). Supporting Social Interaction in an Intelligent Collaborative Learning System. *International Journal of Artificial Intelligence in Education*, 12(1), pp. 40 - 62.
- Sowa, J. F. (1987). *Semantic Networks*. In Stuart C Shapiro. Encyclopedia of Artificial Intelligence. Retrieved 2008-04-29.
- Tang, T. and McCalla, G. (2005). Smart Recommendation for an Evolving e-Learning System: Architecture and Experiment. *International Journal on e-Learning*, 4(1), pp. 105–129, AACE, Chesapeake, VA.
- Tervakari, A.-M., Marttila, J., Kailanto, M., Huhtamäki, J., Koro, J., and Silius, K. (2013). *Developing learning analytics for TUT Circle*. In: Ley, Tobias; Ruohonen, Mikko; Laanpere, Mart; Tatnall, Arthur (ed.). Open and Social Technologies for Networked Learning. IFIP WG 3.4 International Conference, OST 2012, Tallinn, Estonia, July 30-August 3, 2012, Revised Selected Papers. IFIP Advances in Information and Communication Technology vol. 395, pp. 101-110. Springer Berlin Heidelberg.
- Tervakari, A.-M., Kuosa, K., Koro, J., Paukkeri, J., and Kailanto, M. (2014). Teacher's learning analytics tools in social media enhanced learning environment. In Proc. of the *International Conference of Interactive Collaborative Learning, ICL 2014*, 3-6 December 2014, Dubai, United Arab Emirates. In press.
- Tervakari, A.-M., Silius, K., Koro, J., Paukkeri, J., and Pirttilä, O. (2014). *Usefulness of information visualizations based on educational data*. In Proc. of the *2014 IEEE Global Engineering Education Conference, EDUCON*, 3-5 April 2014, Istanbul, Turkey.
- Tsai, C.-J., Tseng, S.-S., and Lin, C.-Y. (2001). A two-phase fuzzy mining and learning algorithm for adaptive learning environment. In Proceedings of the International Conference on Computational Science- Part II, ICCS '01, pp. 429–438, London, UK, UK. Springer-Verlag.
- Tufte, E. 2001. *The Visual Display of Quantitative Information*. Second edition. Graphics Pr. ISBN 978-0961392147.
- Ware, C. (2012). *Information visualization: perception for design*. Third edition. Boston, Morgan Kaufmann.
- Yee, K. P., Fisher, D., Dhamija, R., and Hearst, M., (2001). *Animated exploration of dynamic graphs with radial layout*. In Proc. of the 2001 IEEE Symposium on Information Visualization (INFOVIS), pp. 43-50, IEEE.
- Yang, Q., Sun, J., Wang, J., and Jin, Z. (2010). Semantic web-based personalized recommendation system of courses knowledge research. In Proc. of the *2010 International Conference on Intelligent Computing and Cognitive Informatics, ICICCI '10*, pp. 214–217, Washington, DC, USA. IEEE Computer Society.
- Za'iane, O. R. (2002). Building a recommender agent for e-learning systems. In Proc of the *International Conference on Computers in Education, ICCE '02*, pp. 55–, Washington, DC, USA. IEEE Computer Society.