Revealing the evolution of semantic content through visual analysis

D. A. Gómez-Aguilar 1, M. Á. Conde-González 1, R. Therón 1, F. J. García-Peñalvo 2

1 Computer Science Department
2 Computer Science Department / Science Education Research Institute / GRIAL Research Group
University of Salamanca, Spain
{dialogoa, mconde, theron, fgarcia}@usal.es

Abstract— The literature in educational research has established that monitoring student learning is a crucial component of high quality education. When instructors and students are not in a face-to-face traditional environment, new forms of student monitoring must be explored. The effective use of CMS requires that instructors be provided with appropriate means of diagnosing problems so that they can take immediate actions to prevent or overcome those situations. The aim of this research is to support the comprehension of the evolution of the semantic content within eLearning environments through uncovering by means of visual representations. Therefore, we have carried out the mining of an eLearning database and represented the more relevant results by depicting them using a visualization based in the tag cloud visual representation. In this paper, we introduce a visual representation, which integrate a wave-graph and a bar-graph into a tag cloud representation to understand the evolution of words

Keywords-component; Visualization, e-learning, timeline, tagcloud, Moodle.

I. INTRODUCTION

Many universities and colleges, around the world, that deliver undergraduate programmes in science and engineering are currently incorporating virtual instruments as teaching measurement and analysis tools for student learning. The word Moodle is an acronym for Modular Object-Oriented Dynamic Learning Environment; Moodle is a license free open-source software platform. Those involved with eLearning also call it as Learning Management System (LMS), or Virtual Learning Environment (VLE)). Moodle is designed to help educators and content experts to create online courseware with opportunities for rich interaction. Its open source license and modular design allows content experts to develop additional functionality.

From the architectural point of view Moodle is based on a model-view-controller controller. The adoption of the Service Oriented Architectures (SOA) and its integration in Moodle require a deep knowledge of a system core library that is not particularly consistent. Extending Moodle to use Web Services (WS) gives more flexibility in controlling the connecting operation to obtain the optimal service, reduces the selection service by using accurate attributes of both consumer and service, is a friendly and easily accessible mechanism to discover and select of services and helps to build a high level of trust.

This API consists on a set of WS that encapsulate most of the services that an external (and even internal) application shall need from a Moodle server. In October 2008 this WS layer was integrated in the Moodle 1.9.3 and is going to be the standard interoperability subsystem for the future versions.

The extraction of information from these systems is not trivial and in most cases means hardcoding the LMS and/or visualization tool. This approach involves establishing a link with the platform and, worse still, with its version. To accomplish this, different approaches arise, several of them based on SOA [2]. According to [1], Visual Analytics supports analytical reasoning through interactive visual interfaces. This area has an important limitation towards their implementation within the LMS. This article will use the Moodle web services layer [3] as a proxy to retrieval and exchange information.

The philosophy of Moodle includes a social constructionist approach towards education, emphasizing that both students and teachers can contribute to the educational experience in many ways. Moodle is also useful in an outcomes-based interactive classroom environment that could be continuously improved by analysing the captured data of all activities such as ‘views’ and ‘posts’ for all the learning objects or components of a courseware hosted on a system. The value of a VLE is to fully enable "learning anywhere at any time" by providing an array of resources, opportunities for active participation, mastering content and self-learning.

Munoz-Organero, et al. analysis the effects of motivational states on student performance in an eLearning scenario and measures the correlation between the student interaction patterns with the VLE and his/her level of motivation. One of his results indicates that it is possible to predict deficits in autonomous and eLearning specific motivations by analysing the interactions of the students with the VLE. The number of hits to read eLearning content, the number of hits to participate in forums, and a student’s updating his or her profile have been positively correlated with autonomous and eLearning-specific motivations [4].

With this in mind, we are in a time when the subjects will change and will come up new learning and teaching initiatives. But in order to change and to apply the new available tools, it is necessary to analyse the existing information based on experience. With this information we could determine what method or tool to apply and how and when we are going to do it.

In Moodle “Reports” is one such object that could be very useful in analysing the level of interactivity between the instructor and students during the delivery of a courseware.
Nevertheless, have a brief interactivity and it becomes unusable when representing great quantities of data.

In typical usage, Tagclouds are created by mapping a dimension associated with a term in an underlying data to a dimension parameter determining how that term should be displayed. For example, the prevalence of a term in the set could be represented by its size. However, in Moodle the principal task of tagclouds is Recognition/Matching.

Tagclouds can evolve as the associated data source changes over time. However, while tagclouds seem to invite exposure of their evolution over time, they do not explicitly represent them. This becomes a significant cognitive demand on people who want to understand how a tagcloud evolved.

We organize this paper as follows. In the next section we outline the related work to provide context for our description of Temporal Words Clouds, which follows in Section 3. Section 4 describes the temporal tagclouds. We then conclude the paper with a principal contribution.

II. RELATED WORK

Currently, the increasing use of new technologies to support learning has fostered the creation of tools that help extract information that is not available at first sight. This is essential for the improvement of the learning process from the point of view of institutional decision makers, educational content providers, teachers, and students, all of whom will benefit from the use of effective analytical tools for current e-learning platforms. For a better comprehension of the related works, were organized in two categories: tagcloud and supporting eLearning.

A. Tag Clouds

A tagcloud usually has a particular purpose: to present a visual overview of a collection of text. By this criterion, the first example may have been the outcome of an experiment carried out by social psychologist Stanley Milgram in 1976 [1]. Milgram asked people to name landmarks in Paris, and then created a collective “mental map” of the city using font size to show how often each place was mentioned.

Tagclouds support navigation to the underlying items, serving as automatically created tables-of-contents or indices into a block/batch/set of content. And, much as a table of contents or index can do for a book and a menu of categories can do for a website, they provide a means for users to form a general impression of the underlying set of content and a “gist” of what the book or site is about.

Now at days, among high-profile websites exist a considerable research to improve tag cloud layouts: Flickr1; Del.icio.us2; TagCrowd3; Tagline Generator4; Tag Cloud Generator5; Wordle6; Manyeyes7. Tagline Generator allows people to generate a sequence of tag clouds that are associated with time, from a collection of documents; a dynamic slider control is used to navigate the time points, but only one tag cloud is shown at a time.

Additional enhancements include the use of spatial algorithms to pack the words in a tag cloud into a smaller area, and clustering algorithms so tags which are used together or which have similar meanings are placed near each other. Kaser and Lemire organized tags in nested tables for HTML based sites by using an Electronic Design Automation (EDA) packing algorithm [2]. In the work presented by Torniai et al. the tag cloud employs the size and colour of tags to convey to teachers information describing the tags popularity and relevancy, respectively [3]. Seifert et al. proposed a new algorithm to address several issues found in the traditional layouts [4]. It creates compact and clear layouts by reducing whitespace and featuring arbitrary convex polygons to bound the terms. Tree Cloud arranges words on a tree to reflect their semantic proximity according to the text [5].

Research efforts that attempt to understand the effectiveness and utility of tag clouds generally fall into one of two categories; those which investigate the visual features of tag clouds and those which compare tag clouds with different layouts. Bateman et al. compared nine visual properties of tag clouds for their effects on visual search for tags [6]. Their results show that font size and font weight have stronger effects than others such as colour intensity, number of characters, or tag area.

Rivadeneira et al. conducted two experiments [7]. In the first study, they examined the effect of font size, location, and proximity to the largest tag, asking participants to recall terms (for 60 seconds) that were previously presented in tag clouds (for 20 seconds). In the second study, they investigated the effect of both font size and word layout on users’ abilities to form an impression (gist). From both studies, in accordance with previous research, they observed a strong effect of font size. Halvey and Keane compared tag clouds with traditional lists (horizontal and vertical), each with regular vs. alphabetical order by asking participant to find a specific tag [8]. They found that lists perform better than tag clouds and that alphabetical order further accelerates the search speed.

In use, tag clouds can evolve as the associated data source changes over time. Interesting discussions around tag clouds often include a series of tag clouds and consider trends of their tags over time. This desire to study trends and understand how text content or topics evolve over time has been the purpose of other visualizations such as the commonly used line graphs and bar charts. However, despite the significant amount of research on tag clouds, there has not been much research on how to visualize trends in tag clouds.

Parallel Tag Clouds (PTCs) is designed to provide an overview of a document collection by incorporating graphical elements of parallel coordinates with the text size encoding of traditional tag clouds [9]. While PTCs do show multiple clouds simultaneously, they do not explicitly represent trends, and thus comparing multiple tag clouds to ascertain trends places the cognitive demands on the person.

Bongshin, et al. integrate sparklines into a tag cloud to convey trends between multiple tag clouds. Also they makes controlled study to explore the efficacy of temporal representation on tag cloud see SparkClouds [10].

The first attempt to use the tag cloud view as a part of the coordinated multiple views (CMV) system was presented by Matkovic, et al. [11], there the tag could was used to identify the months when the immigrants are most successful (most landings). Weiwei, et al. introduce a visualization method that couples a trend chart with word clouds to illustrate temporal content evolutions in a set of documents [12]. To better convey the evidence of change across multiple tag clouds, we developed a new breed of tag cloud that integrates wave graph and bar graph into a tag cloud.

B. Supporting eLearning

Due to limited space, some experiences in the literature that deal with the analysis of information generated by the CMS through visual representations are listed and briefly explained below. The reviewed works focus on different aspects of online learning and some basic concepts will be explained. Each message has a sender, date, and topic. A set of posts on the discussion topic, comprised of an initial post and all its responses is called a thread. The person who sent the initial message in a thread is called the originator. Mazza and Milani [13] showed the instant in which users enter the platform and a representation of the frequency of reading and writing in the fora, as well as the thread originator. In [14] the visits and posts over time for each person in a CMS were shown, while in [15] the authors presented the mapping of temporal relations of discussions on software, aimed at helping analyse temporal aspects of online educational course discussions. Finally, Mazza and Dimitrova [16] suggested a scatter-plot-based representation of the online discussions and a matrix to visualize the students’ performance on quizzes related to domain concepts.

Another group of works deals with the use of visualization, rather than information analysis, as part of the learning process or as a supportive resource for coursework [17] [18]. Dichev et al. [19] make use of ontologies and propose the display of thematic maps with the support of semantic information, in addition to their interactive administration. In our previous work [20], we have proposed interactive visualizations of the social networks that are formed among the participants around an activity on the VLE. For a review of the search patterns in the interaction of the learning networks refer to [21].

III. DESCRIPTION OF TEMPORAL WORDS CLOUDS

Visual representations help users to quickly perceive salient aspects of their data. Augmenting the cognitive reasoning process with perceptual reasoning through visual representations permits the analytical process to become faster and more focused. The main goal of the visualization is to provide a compact representation of the overall use of the forums’ interaction on VLE, thus providing an overview of the eLearning platform interest, activities and its evolution over the time.

![Fig. 1. Representation of a single tag on Temporal tagcloud](image1)

Usually, a tagcloud presents a certain number of most often used tags in a defined area of the user interface. A tag’s popularity is expressed by its font size (relative to the other tags) and is therefore easily recognized. Sometimes, further visual properties, such as the font colour, intensity, or weight, are manipulated (for an overview see [6]). Next to their visualization function, tagclouds are also navigation interfaces as the tags are usually hyperlinks leading to a collection of items they are associated with. However, tagclouds are not only used to display taglinks but are also increasingly applied in other contexts and for various data sets, for instance, in the areas of information visualization or text summarization (cp. [22]). Furthermore, several layout variations emerged on the basic design principles of tag clouds. Most popular is the ‘classic’ rectangular tag arrangement with alphabetical sorting in a sequential line-by-line. The goal of the representation for each tag is to represent not only how much it is highlighted, but also the evolution of this representativeness over the time.

![Fig. 2. Global view of temporal tagcloud of forum posts](image2)
The interactive visual tool can be used to analyze the usage of a CMS over time. Later, this view can be adapted to the user’s requirements, so she can explore all the available discussions, forums, courses and users data, going from the overview to the detail of a given person, course or discussion within a period of time.

The interactive visual tool can be used to analyze the usage of a CMS over time. Also, shows the most relevant words from the forums’ activity. Here the user can chose, through the contextual menu interaction (see Fig. 4), to view among the keywords: the users, the courses, the discussions and the subjects of forum posts and the obtained from the analysis of all the semantic content of all forum posts exchanged in the platform.

One of the interested visual analytic technic used on this work is the semantic zoom, which consists to balance the detail and context in data visualization, known as semantic zooming or multi-scale interfaces. A physical zoom, on the one hand, changes the size and visible detail of objects. A semantic zoom, on the other hand, changes the type and meaning of information displayed by the object [23]. The Temporal tagcloud allows both types of zooming, but the emphasis is on the different shapes that is formed depending on the chosen degree of detail, i.e., semantic zoom in the mouse double click interaction over the word.

When the users use the semantic zoom, depending on the context, the tool selects for analyse the forum-posts related with the word zoomed. For example, if the user makes double click over a course, forum or discussion, the tool uses for his analysis all the forum posts from that course or discussion or forum. In the case of the user makes semantic zoom on a word, the tool takes all forum posts that contends the specific word selected to do the reconstruction of the temporal tagcloud.

Furthermore the user can choose what he wants to see. He can draw or hide every element of the representation on the visualization (notice the difference between Fig. 2, where all elements are showed, and Fig. 4 which have only represented the tags).

IV. TASKS THAT TEMPORAL TAGCLOUDS CAN SUPPORT

The propose of our visual tool is the study of “view, update and post” statistics obtained through the real time “Reports” from the logs of some temporal period. With that information which is intended to conduct a study to determine the moments of greatest activity in VLE and how often and for what purpose they are using the forums. On a VLE the professors create, edit and fill different forums, discussions and posts. The following are few common scenarios that the professors have in a VLE:

Any person, as a normally conversation or discussion, tries to maintain the track of all discussions, forums and posts that they make. In the Figure 2, we could see the most used words in a bigger size.

Also, the users need to review, update and to monitor the most frequent discussions, posts, specific problem or student. We can notice the increase focused on the months of May, June, July, August and part of September through the time representation of specific period of time on wave-graph inside of a tag (see Fig. 3 show the “problema” tagcloud).

Moreover, with the same behaviour interaction, the bar-graph permits to the professor see the specific date of the activities. Also we must to consider that thanks to the tool we can check when it has begun to repeat a word over time and what is the point at which most often occurs, and also we could consider the different types of posts depending of the colour.

In Fig. 1, we could observe this phenomenon with the word practice. Taking into consideration adding posts (in blue) we could see that the information of the Practice is added at the beginning of the course. Considering update posts (in green) we can see that there are very few updates in Practice information. Regarding to read-post, given its rightward shift on the word symbolizes that first appears midway through the course and stands towards the end, having a greater frequency near the end of teaching period.

As a part of monitoring a specific theme, course, forum or discussion, the users can find and go directly to the most significant discussion. This can be done drought the contextual menu (see Fig. 4) and select the item “kind tag”, and then “discussion”. The result of this is showed on Fig. 4, where we can recognize like the most relevant discussion is “tutorial”. Then the user can make shift and click over the tag having as an outcome the webpage of the discussion.

Other common case on this scenario is when a professor tries to understand the semantic content of a single forum, course, user or discussion. For this the user needs to goes, through the contextual menu, to the representation of a kind tag that he wants, on the case of this example “tutorias” discussion. Then use the semantic zoom on the specific tag (double click on it), and finally, change the representation to forum post kind tag. Following the same idea, if the user wants to know what is the outline of a semantic content in a specific period of time. Then she/he can select a specific period of the time, through Ctrl+ double click over the period that she/he wants (showed in Fig. 3). Yielding as a result a semantic zoom of the period selected.

Finally, we can mentioned that besides these words originally appeared other, but were excluded because they did not report relevant information, among them were articles, prepositions, student identification numbers and names. Should be added that the visualization tool is continually evolving and throughout the writing of this article, there are new features. Including synonyms and plurals binding and the representation of the space of words at a forum or course, but for this case study were not considered.
V. CONTRIBUTIONS AND CONCLUSIONS

Based on our research, this work is one the first to generate a visualization that uses tag clouds to depict evolving text content over time. As a result, our work offers unique contributions:

- The possibility to use the three tasks of the tagcloud: search, browsing and recognition through to the use of semantic zoom, the visual representation design and the user interactions design.
- Two-level visualization, on the same visual representation, that fuses a histogram chart and dynamic word clouds to illustrate temporal content evolution at multiple levels of detail.
- Time-based tag cloud layout that balances semantic coherence of content and spatial stability of the visualization to help users easily perceive content updates.

ACKNOWLEDGMENT

This work was supported by Spanish Government projects TIN2010-21695-C02-01 and FFI2010-16234, and by the Castilla y Leon Regional Government through GR47.

REFERENCES