Enhancing Online Discussion Forums with a Topic-Driven Navigational Paradigm: An Implementation for the Moodle Learning Management System

Damiano Distante[1], Luigi Cerulo[2], Aaron Visaggio[2] and Marco Leone[2]
[1]*Unitelma Sapienza University, Rome - Italy*
[2]*University of Sannio, Benevento - Italy*
damiano.distante@unitelma.it,{lcerulo,visaggio}@unisannio.it,mleone85@gmail.com

Keywords: Discussion forums, navigability, searchability, information search, information extraction, text mining, topic modeling, e-learning, learning management systems.

Abstract: One of the most popular means of asynchronous communication and most rich repository of user generated information over the Internet is represented by online discussion forums. The capability of a forum to satisfy users' needs as an information source is mainly determined by its richness in information, but also by the way its content (messages and message threads) is organized and made navigable and searchable.
To ease content navigation and information search in online discussion forums we propose an approach that introduces in them a complementary navigation structure which enables searching and navigating forum contents by topic of discussion, thus enabling a topic-driven navigational paradigm. Discussion topics and hierarchical relations between them are extracted from the forum textual content with a semi-automatic process, by applying Information Retrieval techniques, specifically Topic Models and Formal Concept Analysis. Then, forum messages and discussion threads are associated to discussion topics on a similarity score basis.
In this paper we present an implementation of our approach for the Moodle learning management system, opening to the application of the approach to several real e-learning contexts. We also show with a case study that the new topic-driven navigation structure improves information search tasks with respect to using Moodle standard full-text search.

1 INTRODUCTION

Online discussion forums represent one of the main sources of user generated information (i.e., social media) over the Internet and enable asynchronous communication between Internet users in the form of message posts. Most visited websites, including blogs and social networks, use forums to support user interaction and knowledge sharing. In several domains ranging from e-commerce (Otterbacher, 2008)(Gruen et al., 2006), to news (Li et al., 2010), and healthcare (Sudau et al., 2014) discussion forums constitute rich and widely accessed repositories of information for Internet users.

As an example, software developers forums are an effective source of information where programmers search for and describe solutions to specific problems [1]. In e-learning contexts, discussion forums enable asynchronous communication student-to-student, and teacher-to-student, e.g., to support collaborative learning and group work (Stefan, 2009)(Hrastinski, 2008). Whatever the forum domain, discussions held in a certain period of time become a source of information for any user accessing the forum afterwards.

In general online forums organize messages into into a chronological order. A user starts a new discussion by posting an initial message, other users post their replies or comments to it, and the list of messages form a *discussion thread*. If users are allowed to reply to other users' replies in additional to the original message, discussions take the form of trees, with discussion branches.

The effectiveness of a discussion forum as information source depends on its richness in information, but especially on the searching paradigm users can adopt to find contents of their interest.

[1]An example of such forum is the Microsoft MSDN Developer Network forum. http://social.msdn.microsoft.com/Forums/en/categories/

Search features usually provided with online discussion forums are limited to full-text search which returns a list of forum messages that include (and/or do not include) one or more of the query keywords in their body and/or their title. Such a search feature may return too many or too few results (depending on the forum size and the query keywords) and may miss messages which are semantically related to the query keywords but do not actually include them (Baeza-Yates and Ribeiro-Neto, 1999).

Hierarchical graphs constitute an effective paradigm to represent users' knowledge (Zhang and Peck, 2003). In a previous work (Cerulo and Distante, 2013) we have introduced an approach to improve information retrieval and content navigation in online discussion forums by introducing in them a complementary hierarchical topic-driven navigation structure. Information Retrieval (IR) techniques, specifically Topic Models (Blei, 2011) and formal concept analysis (FCA) (Ganter and Wille, 1999), are used to discover discussion topics and hierarchical relations between them in the forum content. Then, forum messages and discussion threads are associated to discussion topics based on a similarity score, thus to enable searching and navigating them on a topic-driven basis, additional to conventional chronological order and full-text search approaches.

In this paper we present an implementation of this approach as a plugin for the Moodle learning management system which makes the topic-driven navigation approach accessible and evaluable in several e-learning contexts. We also present a case study which provides a first qualitative assessment of the benefits of topic-driven navigation and access to forum content, with respect to traditional full-text search.

The rest of the paper is organized as follows. Section 2 describes our forum navigation enhancement approach introduced earlier in (Cerulo and Distante, 2013). Section 3 presents the implementation of the approach for the Moodle[2] learning management system. Section 4 reports on a case study we have conducted to qualitatively assess the benefits of the topic-driven forum enhancement approach in searching forums for information of interest for the user. Section 5 overviews related work and Section 6 draws conclusions and introduces future works.

[2]www.moodle.org

2 THE TOPIC-DRIVEN FORUM NAVIGATION ENHANCEMENT PROCESS

The topic-driven forum navigation enhancement process, introduced recently by some of the authors of this paper (Cerulo and Distante, 2013), is shown in Figure 1. It consists of four main steps represented in the figure as rectangles and described briefly in the following.

2.1 Terms extraction

We represent a forum message as a vector of indexing terms, $\{t_1, \ldots, t_m\}$, extracted, from the corpus of n messages, $\{d_1, \ldots, d_n\}$, through a standard text analysis pipeline usually adopted in Information Retrieval that comprises: outlier filtering, stopwords filtering, and stemming (Baeza-Yates and Ribeiro-Neto, 1999).

The outcome of this step is a document-term matrix $\mathbb{DT}$, where each element $\{\mathbb{DT}\}_{jp}$ is the *tf-idf* of the term t_p in the forum message d_j (Baeza-Yates and Ribeiro-Neto, 1999).

2.2 Topic modeling

Topic modeling, in particular Latent Dirichlet Allocation (LDA), is a statistical technique that is able to extract frequently co-occurring terms, known as *topics*, from a corpus of documents (Blei, 2011). The input is the document-term matrix, $\mathbb{DT}$, obtained from the previous task, while the output is a topic-document matrix, $\mathbb{TD}$, and a topic-term matrix, $\mathbb{TT}$. The number of topic k is a parameter that controls the granularity of the topics and must be fixed a priori.

Intuitively, the top terms of a topic are semantically related and represent some real-world concepts. For example the concept related to problems e-mails setup is represented by the terms "mail", "problem", "setup". The topic membership of a document describes which concepts are present in that document. Table 1 shows and example of topic-document and topic-term matrices.

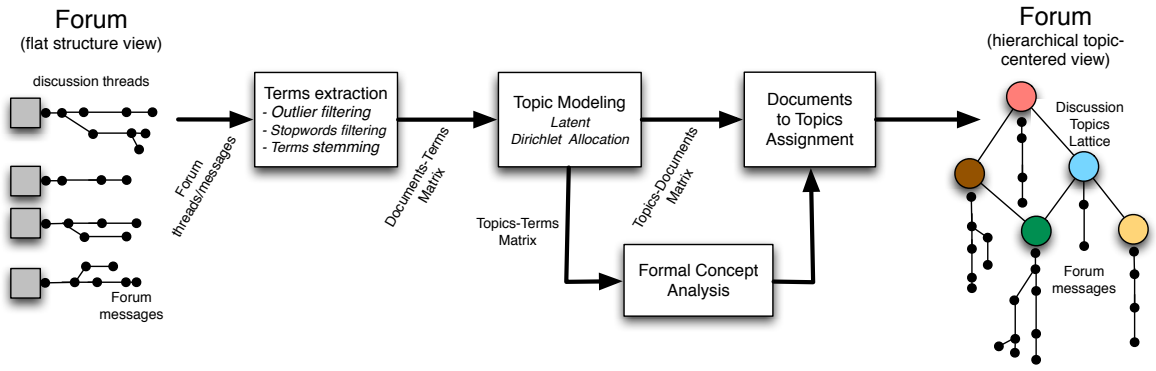

Figure 1: The topic-driven forum navigation enhancement process (Cerulo and Distante, 2013).

	topic-term	topic-document			
	Top terms	d_1	d_2	d_3	d_4
z_1	*problem, email, setup*	0.6	0.7	0.1	
z_2	*problem, email, connection, setup*	0.3	0.1	0.1	0.5
z_3	*problem*	0.1			
z_4	*problem, video, decoder, setup*	0.1	0.2	0.8	0.5
z_5	*problem, video*	0.1	0.2		

Table 1: Examples of topic-term and topic-document matrices

	problem	*email*	*connection*	*video*	*decoder*	*setup*
z_1	×	×				×
z_2	×	×	×			×
z_3	×					
z_4	×			×	×	×
z_5	×		×	×		

Table 2: The formal context obtained from the topic-term matrix shown in Table 1

2.3 Formal concept analysis

Using the topic membership of a term, we prune a topic lattice by means of Formal Concept Analysis (FCA). FCA is a computational way to derive a concept hierarchy or formal ontology from a collection of objects and their properties (Ganter and Wille, 1999)(Birkhoff, 1967).

We model the topics as the objects of a formal context and the terms as their attributes. The relation R of the formal context is computed from the topic-term matrix $\mathbb{TT}$ by means of a decision threshold h_T, *i.e.*, a term (attribute) t_p belongs to a topic (object) z_i, $(z_i, t_p) \in R$ iif $\{\mathbb{TT}\}_{ip} \geq h_T$.

As a clarification example consider the formal context shown in Table 3 and the topic lattice obtained from such a formal context shown in Figure 2. Topics are mapped on circles and hierarchical relationships are represented by arcs. Large circles are mapped on topics extracted with the topic modeling approach, while small circles are intermediate topics extracted with the formal concept analysis. The lattice shows the hierarchical relationships between topics. In the lattice the top most topic (z_3) is the most general topic. A path starting from the top most topic is a more specific topic. For example z_2 is reachable by the path from z_3 (problem), setup, z_1 (email), and z_2 (connection), and represent the more specific topic of problems related to the email connection setup.

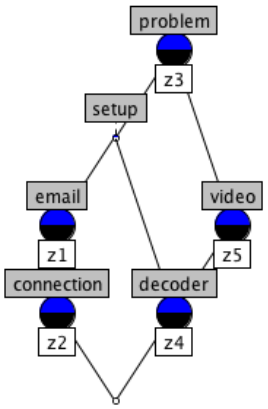

Figure 2: The topic-lattice pruned from the formal context shown in Table 3.

2.4 Documents to topics assignment

During this step each document (*i.e.* forum message/thread) is mapped onto the topics to which it is more likely to belong by estimating the probabilities of each topic for that message (topic-document matrix). For this purpose we adopted the topic-document matrix $\mathbb{TD}$ and a decision threshold h_D, *i.e.*, a document d_j belongs to a topic z_i iif $\{\mathbb{TD}\}_{ij} \geq h_D$.

2.5 Parameter setting and accuracy evaluation

Selecting the number of topic k is one of the most problematic modeling choice in topic models (Wallach et al., 2009). We adopt a metric, introduced by Meilă (Meila, 2003) for clustering comparison, that measures the *Variation of Information* as the entropies the mutual information associated with cluster assignments. Intuitively, the entropy measures the uncertainty of allotting an item to a cluster, while the mutual information measures the reduction of such uncertainty when the allocation in the other cluster is known. Following the approach adopted by Wallach *et al.* (Wallach et al., 2009) the assignment of documents (forum messages or threads in our context) to topics can be assimilated to a sort of cluster assignment. In our previous work (Cerulo and Distante, 2013) we showed that above a certain value of k no significant increment of Variation of Information can be observed in a specific context. We consider such a value the optimal number of topic in that context.

We evaluated the document assignment task to check whether the documents assigned to topics by the Latent Dirichlet Allocation were congruent with the semantics of their content (Bakalov et al., 2012). In our previous work (Cerulo and Distante, 2013) we addressed this question with a controlled experiment obtaining in average a precision ranging between 52% and 74%.

3 TDFORUM: A PLUGIN FOR THE MOODLE LEARNING MANAGEMENT SYSTEM

Topic-Driven Forum (TDForum) is a Moodle plugin (particularly, an *activity module*) that implements the topic-driven forum navigation enhancement approach described in Section 2 for the Moodle open-source learning management system.

In Moodle, *activity* is a general name for a group of features in a course. Usually an activity is something that a student will do that interacts with other students and/or the teacher. Assignments, quizzes, surveys, workshops, chats, and forums are examples of activities that can be created in a course and that are provided in Moodle by default. Each activity is implemented by a software module (plugin) located in the *mod* sub-folder of the Moodle instance. Additional activities can be included by installing the corresponding Moodle plugin[3].

From a source code point of view, each Moodle activity module consists of a series of mandatory files (e.g., install.xml, lib.php, and view.php) used to install the module and integrate it within the Moodle system, and other files specific to the plugin.

Figure 3 shows the architecture of the TDForum Moodle plugin that we developed. In the figure, we can distinguish the components representing the plugin *front-end* (the graphical interfaces that Moodle users interact with), and those that are part of the plugin *back-end*.

The plugin front-end comprises the components *Main View* and *Discussion Topics View* corresponding to the two possible views on the forum content: (*i*) standard chronological list of discussions augmented with discussion topics and scores, and (*ii*) navigable hierarchical discussion topics graph). The last view is built using the *JavaScript InfoVis Toolkit*[4]). It also includes the *Admin User Interface* component which lets administrators (*i*) manage the forum data processing activity and customize the visualization plugin parameters.

The plugin back-end contains the components implementing the forum analysis and indexing process described in Section 2 to build the additional topic-driven navigation structure. In particular, the *Process Controller* controls the process by executing the commands provided through the plugin admin user interface. It also exports forum content from the Moodle database into a local temporary csv text file and imports the data on the new navigation structure from the local filesystem into the Moodle database.

The *Data Processing* component includes the following sub-components:

- *Data Preprocessing*: a Perl script which extracts threads and messages from the csv file into separated text files and performs terms extraction and text filtering such as stopwords and stemming (cf. Section 2.1).
- *Topics Identification and Documents to Topics As-*

[3] A rich and up-to-date list of Moodle plugins can be found in the Moodle Plugins Directory at http://www.moodle.org/plugins

[4] http://philogb.github.io/jit/

signment: a R[5] script which uses the Topic Model library[6] to perform discussion topics identification and documents to topics assignment. The matrices Topics-Terms and Topics-Documents of the detected forum discussion topics and scores associated to them are generated in this step (cf. Sections 2.2 and 2.4).

- *Formal Concept Analysis and Topics Graph Export*: this component uses the FcaStone[7] Formal Concept Analysis command-line utility to generate the lattice representative of the hierarchy of topics and to export the topics graph used in the graph view of the plugin (cf. Section 2.3).

The TDForum activity implemented by our plugin offers the same features provided by a standard Moodle forum (particularly, a *main view* which lists forum discussions and messages organized in a chronological order, the functionality of posting new messages or replying to existing ones, full-text search of messages, etc.) and adds to them a *discussion topics view* which acts as a topic-driven navigation index to the forum content.

The *main view* (Fig. 4) presents the list of discussion threads of the forum in a chronological order and adds to each of them the list of discussion topics in it identified, and the calculated similarity score (column 'Discussion topics' in the figure). Score values range between 0 and 1 (with 1 representing the maximum similarity value) and the list of topics associated to a discussion is ordered by score. By clicking one of the topics of the list, the user can search for discussions or messages which are related to the selected topic. The results of this search is presented sorted by decreasing values of score.

The *discussion topics view* (Fig.5) shows the list of discussion topics found by the analysis process for the considered forum (scrollable list on the left side of the figure) and a graph that the user can pan and zoom which highlights the hierarchical relations between the identified topics. The user can navigate the discussion topics graph or the topics list and once she finds a topic of her interest she can retrieve the list of discussions/messages associated to it with a click.

The plugin has been designed to extend a standard Moodle forum and, at the same time, to be independent from it. As such, if it is installed, applied on a forum, and then deactivated, none of the content of the original forum are lost, nor the additional messages/discussions that will have been added in it after the plugin instantiation.

[5]http://cran.r-project.org/

[6]http://cran.r-project.org/web/packages/topicmodels/index.html

[7]http://fcastone.sourceforge.net/

4 CASE STUDY

We evaluated qualitatively that, with the topic-driven approach, searching and browsing tasks of forum contents can be significantly improved with respect to traditional full-text search. The case study has been conducted on forums inside an instance of the Moodle learning management system. The context is composed by a reduced version of 2 forums extracted from the Moodle user and development communities (Table 3). The *Installation Help Forum* includes all discussions about user difficulties with first Moodle installations or errors happening during the installation process, or with migration to different OSs, or to newer Moodle versions. The *General Help Forum* includes discussions about problems not included into other Moodle community forums, such as problems with database access, file upload, block modules and student enrollment.

We evaluated effectiveness in 11 searching tasks in terms of (*i*) the number of items (forum posts) the user had to inspect in order to satisfy the information need, and (*ii*) the time spent to accomplish the task (Table 4). The nature of the 11 searching tasks has been defined by the first two authors of this paper. For each task the search goal, *i.e.,* the expected posts that should be retrieved, is known beforehand. The other two authors performed the searching tasks with two complementary approaches: (*i*) full-text search, and (*ii*) topic driven navigation. The first approach is accomplished with the default full-text search engine implemented in the Moodle platform which performs a full-text search from a set of user defined keywords. By default, the keywords are linked by an AND operator and the system retrieves the list of all posts containing all searched terms. The second approach is executed with the TDForum Moodle plugin introduced in Section 3. While performing the tasks, we collected the number of items inspected and the time needed to achieve the search goal. With the Moodle full-text search the number of inspected items is computed by counting the number of posts examined before finding the correct expected posts. With the topic-driven navigation approach the number of inspected items is the sum of two quantities: the number of links followed to reach the closest topic in the Discussion Topics View of the TDForum plugin and the number of posts examined before finding the correct expected posts.

Table 5 reports the results obtained by executing the evaluation protocol on the 11 tasks. The table reports for each search goal the number of inspected items and the time spent to find the correct posts. For Moodle full-text search we also reported the number

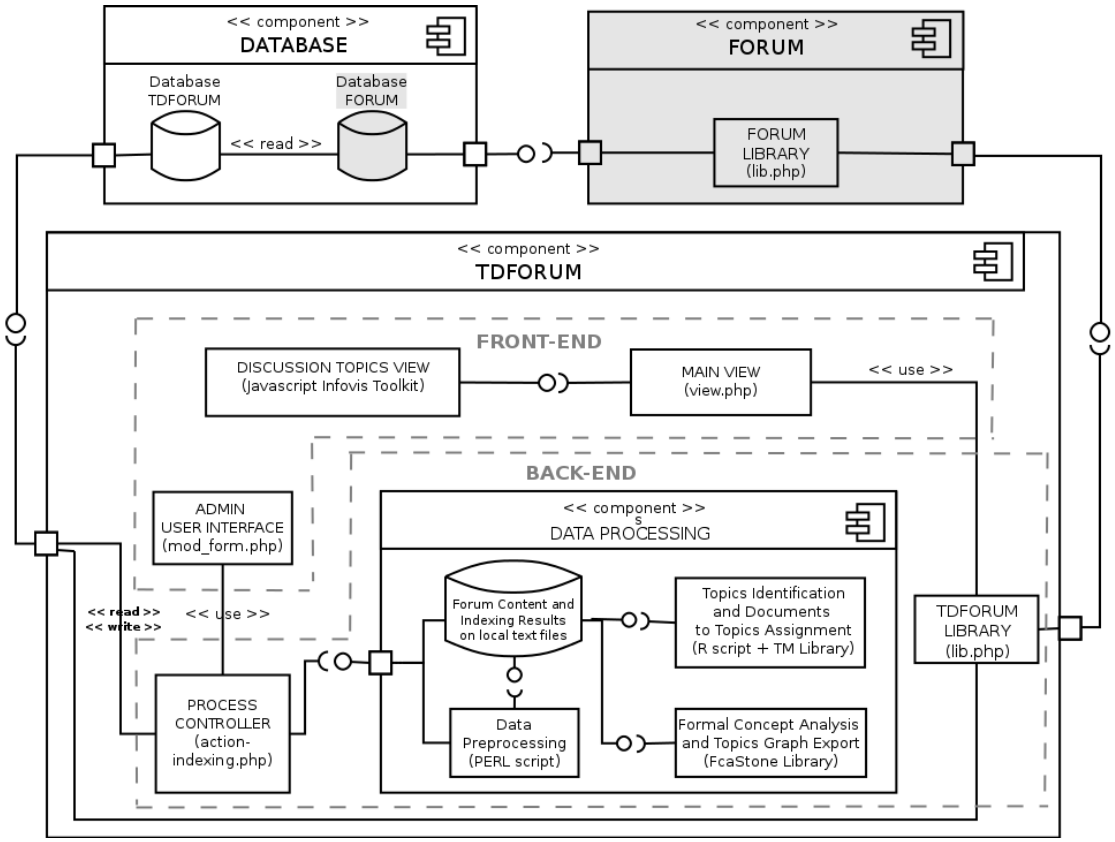

Figure 3: Architecture of the TDForum Moodle plugin (with a gray background color, standard Moodle components).

Forum	# threads	# posts	# users	time period
Installation Help	253	777	78	May 1, 2013 – May 24, 2013
General Help	115	714	107	Jul 15, 2013 – Aug 8, 2013

Table 3: Case study context

of search attempts (queries) performed with different search keywords necessary to reach the goal. In general the number of items inspected with full-text search is in average higher than the number of items inspected with TDForum (14 vs 9). The time necessary to obtain the correct answer is in average less in TDForum (137 sec. vs 170 sec.) because with full-text search more time is spent to choose the correct search keywords. The difference is not statistically significant due to the limited number of samples, thus further experiment are necessary to draw more general conclusions.

Task ID	Moodle full-text search # queries	# items	time	TDForum search # items	time
1	2	15	201	20	254
2	1	5	109	5	92
3	1	17	131	8	168
4	3	12	230	7	135
5	3	13	225	11	187
6	5	40	275	9	113
7	1	5	134	4	62
8	2	42	287	10	131
9	1	4	122	6	90
10	1	1	86	16	192
11	1	0	70	6	85
average	2	14	170	9	137

Table 5: Case study results (time in seconds)

5 RELATED WORK

Recently, on-line education systems are becoming widespread tools adopted by both historical and newly founded educational institutions. E-learning and e-teaching are new contexts for education where large amounts of information are generated and ubiq-

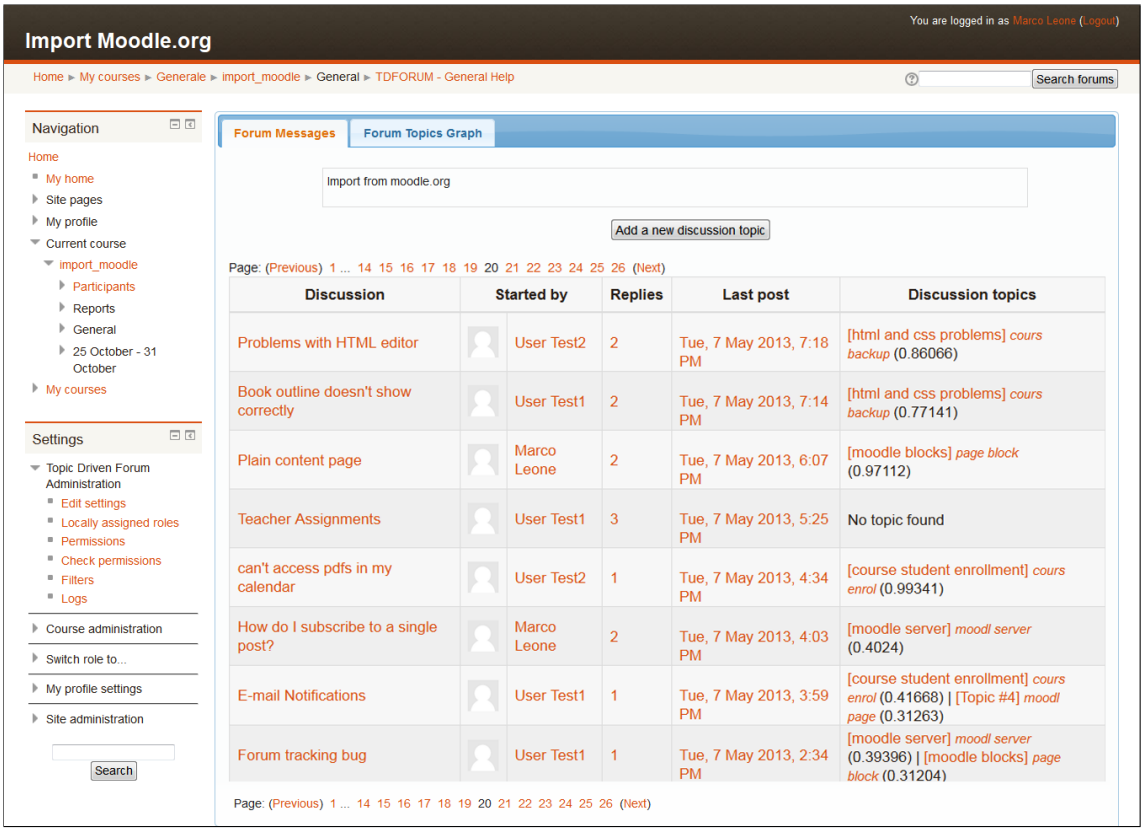

Figure 4: The main view of TDForum showing the list of forum discussion threads enhanced with discussion topics and scores associated to them.

uitously available. Most of generated information has the form of free text without a structure crucial for automating knowledge retrieval.

Data Mining has been historically used to extract knowledge from free text (Baeza-Yates and Ribeiro-Neto, 1999). Knowledge extraction from e-learning systems, in particular from user generate data, has been introduced in (Castro et al., 2007b; Hanna, 2004). Patterns of system usage by teachers and learning behavior by students has been investigated in (Tang and McCalla, 2005). Data clustering was suggested to promote group-based collaborative learning and to diagnose students incrementally (Castro et al., 2007a).

Web Mining techniques to meet some of the current challenges in distance education was presented in (Sung Ho Ha, 2000) where a clustering of forum messages are in fact grouped into similar discussion topic classes. Association Rules mining has been widely adopted in e-learning, in particular recommendation systems (Zaíane, 2002; Yang et al., 2010), learning material organization (Tsai et al., 2001), student learning assessments (Romero et al., 2005), course adaptation to the students behavior (Hogo, 2010), and evaluation of educational websites (dos Santos Machado and Becker, 2003). In educational research the development of cooperative learning and knowledge sharing inside student groups constitute recent research trends(Jakobsone et al., 2012). To this aim, Web technologies should grasp the opportunities raised by mixing the Social and the Semantic Web (Ghenname et al., 2012) and on adopting Semantic and Artificial Intelligence techniques for discovering information objects and restructure large digital collections (Martin and Leon, 2012). Concept maps and their use for navigation in educational contexts has been investigated in the recent past by different authors. As a representative of this research effort we cite the work of Dicheva and Aroyo (Dicheva and Dichev, 2006). In this work the authors propose a framework and a set of tools for the development of ontology-aware repositories of learning materials. While the idea and use of *concept maps* is similar to our topic-driven navigation structure, in our approach topics are extracted from free text in a semi-automatic way, by leveraging information retrieval techniques and then validated by the user, while concepts have to be manually defined by the authors of the learning

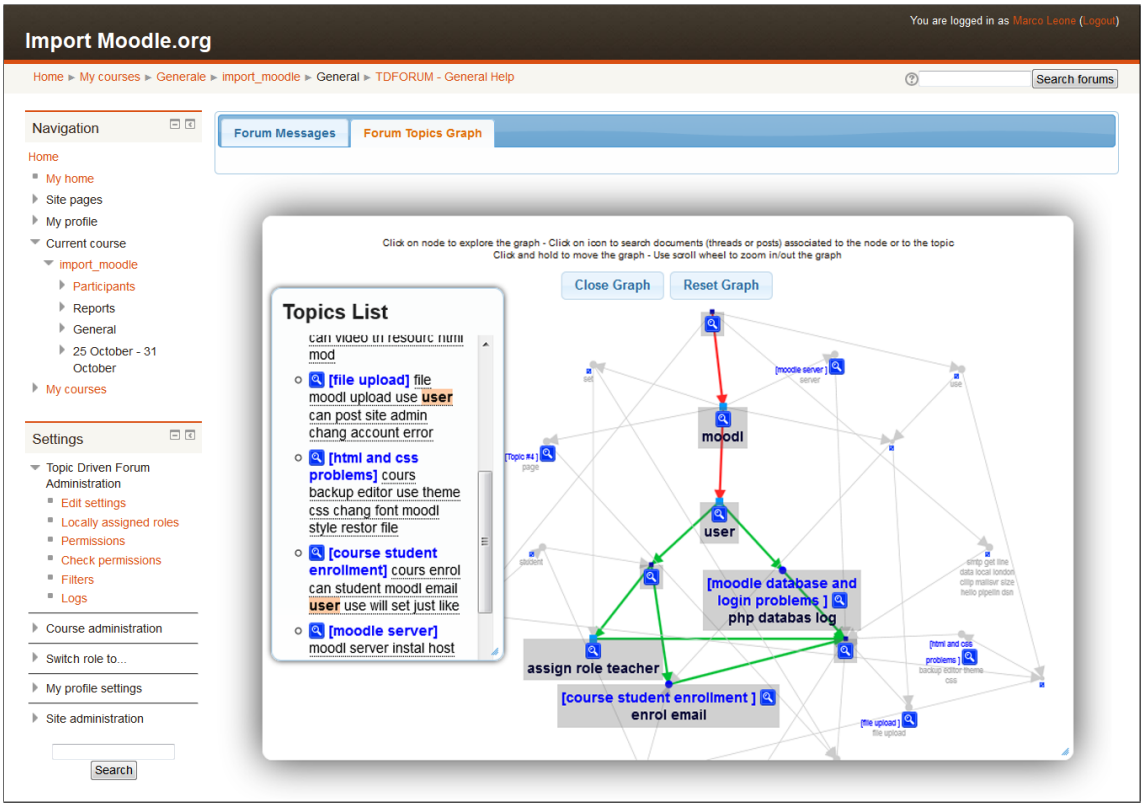

Figure 5: The Discussion Topics view of TDForum showing the topics list and the hierarchical topics graph.

materials in the work of Dicheva and Aroyo.

6 CONCLUSIONS AND FUTURE WORK

Online discussion forums are one of the main asynchronous communication means and repositories of user generated content over the Internet. Learning management systems (LMSs), such as Moodle, use forums to support interaction and collaboration between students and students-to-teachers. Discussions taken place in a forum at some time represent a source of information for users accessing the forum afterwards. However, the effectiveness of a forum as a source of information for its users, additionally to be closely related to its richness in content, is also influenced by the way its contents are organized made searchable.

In this paper we presented an approach and a plugin for the Moodle LMS that enhances content navigation and information search in online discussion forums with a topic-driven navigational paradigm. The approach enables the automatic recovery of a lattice of discussion topics from the forum content, and the introduction of an additional navigation structure and graphical user interface which enable navigating and searching forum contents by topics of discussion.

While the approach has proven correctness for both the identified topics and the document-to-topics assignment (Cerulo and Distante, 2013), in this paper we have also shown with a case study that the additional navigation structure significantly improves the search of information stored in forum discussions.

In the future we aim to apply our approach in the context of social networks, in order to explore how it could improve social organization and user interaction. As a matter of fact, social networks are increasingly used in e-learning as side means for connecting students and teachers.

ID	Search goal	Adopted Keywords
1	Retrieve the 10 posts related to css problems in the General help forum	*css*, *problem*
2	Retrieve the 5 posts related to login issues in the General help forum	*login*, *problem*
3	Retrieve the 3 posts related to uploading files problems in the General help forum	*file*, *not*, *upload*
4	Retrieve the 4 posts related to changing Moodle fonts in the General help forum	*change*, *font*
5	Retrieve the 5 posts related to not sent enrollment email in the General help forum	*enrollment*
6	Retrieve the 3 posts related to editing Moodle theme in the General help forum	*change*, *theme*
7	Retrieve the 5 posts related to web hosting in the General help forum	*moodle*, *web*, *hosting*
8	Retrieve the 10 posts related to Moodle upgrading problems in Installation help forum	*problem*, *moodle*, *upgrade*
9	Retrieve the 5 posts related to editing admin password in the Installation help forum	*admin*, *password*
10	Retrieve the unique post related to missing files after Moodle migration in the Installation help forum	*missing*, *files*, *after*, *migration*
11	Retrieve the 2 post related to slower system after upgrade in the Installation help forum	*css*, *problem*

Table 4: Search tasks definition

REFERENCES

Baeza-Yates, R. A. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Bakalov, A., McCallum, A., Wallach, H. M., and Mimno, D. M. (2012). Topic models for taxonomies. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '12, Washington, DC, USA, June 10-14, 2012*, pages 237–240.

Birkhoff, G. (1967). Lattice theory. In *Colloquium Publications*, volume 25. Amer. Math. Soc., 3. edition.

Blei, D. M. (2011). Introduction to probabilistic topic models. *Communications of the ACM*.

Castro, F., Nebot, A., and Mugica, F. (2007a). Extraction of logical rules to describe students' learning behavior. In *Proceedings of the sixth conference on IASTED International Conference Web-Based Education - Volume 2*, WBED'07, pages 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*, pages 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 *Global Engineering Education Conference (EDUCON), 2013 IEEE*, pages 303–310.

Dicheva, D. and Dichev, C. (2006). Tm4l: Creating and browsing educational topic maps. *British Journal of Educational Technology*, 37(3):391–404.

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*, pages 360–361. IEEE Computer Society.

Ganter, B. and Wille, R. (1999). *Formal concept analysis: mathematical foundations*. Springer.

Ghenname, M., Ajhoun, R., Gravier, C., and Subercaze, J. (2012). Combining the semantic and the social web for intelligent learning systems. In *Global Engineering Education Conference (EDUCON), 2012 IEEE*, pages 1 –6.

Gruen, T. W., Osmonbekov, T., and Czaplewski, A. J. (2006). eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business Research*, 59:449456.

Hanna, M. (2004). Data Mining in the e-Learning Domain. *Campus-Wide Information Systems*, 21(1):29–34.

Hogo, M. A. (2010). Evaluation of e-learning systems based on fuzzy clustering models and statistical tools. *Expert Syst. Appl.*, 37(10):6891–6903.

Hrastinski, S. (2008). What is online learner participation? a literature review. *Computers & Education*, 51(4):1755 – 1765.

Jakobsone, A., Kulmane, V., and Cakula, S. (2012). Structurization of information for group work in an online environment. In *Global Engineering Education Conference (EDUCON), 2012 IEEE*, pages 1 –7.

Li, Q., Wang, J., Chen, Y. P., and Lin, Z. (2010). User comments for news recommendation in forum-based social media. *Information Sciences*, 180:49294939.

Martin, A. and Leon, C. (2012). An intelligent e-learning scenario for knowledge retrieval. In *Global Engineering Education Conference (EDUCON), 2012 IEEE*, pages 1 –6.

Meila, M. (2003). Comparing clusterings by the variation of information. In *Computational Learning Theory and Kernel Machines, 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003*, pages 173–187.

Otterbacher, J. (2008). Searching for product experience attributes in online information sources. In *Proceedings of the International Conference on Information Systems (ICIS 2008)*. Association for Information Systems.

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):425–464.

Stefan, H. (2009). A theory of online learning as online participation. *Computers & Education*, 52(1):78–82.

Sudau, F., Friede, T., Grabowski, J., Koschack, J., Makedonski, P., and Himmel, W. (2014). Sources of information and behavioral patterns in online health forums: qualitative study. *Journal of medical Internet research*, 16:e10.

Sung Ho Ha, Sung Min Bae, S. C. P. (2000). Web mining for distance education.

Tang, T. and McCalla, G. (2005). Smart Recommendation for an Evolving e-Learning System: Architecture and Experiment. *International Journal on e-Learning*, 4(1):105–129.

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, pages 429–438, London, UK, UK. Springer-Verlag.

Wallach, H. M., Mimno, D. M., and McCallum, A. (2009). Rethinking lda: Why priors matter. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009*, pages 1973–1981.

Yang, Q., Sun, J., Wang, J., and Jin, Z. (2010). Semantic web-based personalized recommendation system of courses knowledge research. In *Proceedings of the 2010 International Conference on Intelligent Computing and Cognitive Informatics*, ICICCI '10, pages 214–217, Washington, DC, USA. IEEE Computer Society.

Zaíane, O. R. (2002). Building a recommender agent for e-learning systems. In *Proceedings of the International Conference on Computers in Education*, ICCE '02, pages 55–, Washington, DC, USA. IEEE Computer Society.

Zhang, K. and Peck, K. (2003). The effects of peer-controlled or moderated online collaboration on group problem solving and related attitudes. *Canadian Journal of Learning and Technology / La revue canadienne de lapprentissage et de la technologie*, 29(3).