# Discovering occurrences of user-defined patterns in historical data representing collaborative activities in virtual user environment

Jozef Wagner[2], Ján Paralič[1], František Babič[1]

[1] Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 042 00 Košice
frantisek.babic@tuke.sk, jan.paralic@tuke.sk
[2] Centre for Information Technologies, Faculty of Electrical Engineering and Informatics, Technical University of Košice, B. Němcovej 3, 042 00 Košice
Jozef.wagner@tuke.sk

**Abstract.** The paper deals with analyses of performed collaborative activities in virtual user environment, focused on pattern discovering. All activities are monitored and recorded into separate database within defined log format. This log format provides sufficient historical data for various analytical purposes as visualization through timeline or extraction of different statistics based on user expectations or requirements. Historical overview of performed activities through timeline displays series of various events in chronological order and some subsets of them may be interesting or important for users. These subsets can be represented in form of patterns and can be used to discover subsequent pattern occurrences in historical data.

**Keywords:** event log, pattern, discovering

## 1 Introduction

Collaborative activities realized within virtual user environment consist of various elements that are in relations and represents possible inputs for analytical purposes. It is important to evaluate past instances of performed activities in order to improve their future realization through more effectively usage of resources, better group composition, identification of critical points, discovering of best practices, etc.

Simple example of typical collaborative activity is creation of conference article in collective of several persons. They need collaborative space to share their inputs and contributions, to display their progress and to communicate and comment published versions of the paper. The whole creation process consists of several phases (initial draft, etc.) and each of these phases requires interaction between participants based on their roles, knowledge and theoretical background. The virtual environment provides functionalities to accomplish these expectations and requirements. The important function is the ability to trace performed steps and visualize them through timeline in order to provide historical overview in chronological order. The created timeline

displays the whole progress and relations between related events, objects and users and can be used for evaluation. Patterns are important part of evaluation process, because they can identify critical points or best practices for future realization.

### 1.1 Related Work

Several similar analytical approaches can be identified in relevant domain, as utilization of suitable data mining algorithm or various visualization forms, etc.

Semantic Spiral Timelines represents an interactive visual tool aimed at the exploration and analysis of information stored in investigated collaborative learning systems [21]. It provides an interesting way of presenting the events in form of spiral timeline that contains sequences of color-coded events.

Interesting and similar approach to the proposed solution is described in [33]. This approach covers constrain-based analytical approach for pattern discovering, i.e. defining filters during the pre-processing phase that reduces the search space; constraints during the mining phase (association rule mining, sequential pattern analysis, clustering and classification) accelerate and control pattern discovering.

Hardless and Nulden proposed the Activity Visualization (AV) as technical supporting functionality for understanding of learning processes in virtual environment. AV uses information from the environment to visualize aspects of the whole learning process, to give users the opportunity to view activities, progress, and usage patterns from various perspectives [4], e.g. analysis of message lengths to reveal usage patterns and relationships, message counts in relation to time, etc.

Process mining tool, called ProM[1], developed in the Eindhoven University of Technology, represents state of the art in the process analysis. ProM tools offers vast number of analyses of the event logs themselves, or analyses where also the process model is present. For the purposes of analyses, the ProM tool expects that it is possible to divide activities from the event log into the separate sets, where each set represents one instance of a process, called ‘case’. Given activity presents in the event log must belong to exactly one process instance.

The OCAF framework [1] shares similar motivation to our approach. It also places objects at the center of the collaborative activity. Those objects are studied as entities with their own history and evolution procedures. Synergo, which builds on the OCAF framework already supports the automatic association of some kind of events, e.g. insertion, or modification event. The authors also list serious shortcomings of log analysis (especially quantitative indicators may be misleading) so they recommend mainly qualitative analysis of collaborative activities.

## 2 Timeline-based visualization

Timeline-based visualization provides features and methods to display the whole evolution process (sequence of actions/activities) or its particular part with relevant interactions and relations [6]. Input data can be filtered based on group selection,

[1] http://prom.win.tue.nl/research/wiki/

particular time-frame or relevant part of virtual user environment. This approach produces a complex view of performed activities and gives the possibility to focus on potentially interesting facts. The proposed solution offers creation of basic timeline or several different timelines for comparison; possibility to add external events performed outside virtual system but relevant to examined activities; commenting; patterns definition, etc.

### 2.1 Data

Historical data recorded for analytical purposes represents action/activities performed in the virtual user environment. Each event is monitored and logged within designed and implemented logging services on the middleware layer of the system. These services provide two-side communication within presentation layer and log repository. This repository is implemented within MySQL relational database in order to provide scalable and responsive solution [5].

## 3 Patterns

Activities performed within the virtual user environment cannot be fully described by some well defined, rigid process. It is in the very nature of the collaborative user environment, that the complex activities performed in it are ill defined and not easily formulated. Model of such activity must cover such cases and be able to dynamically build, modify and customize the description of it. Moreover, it is not possible to explicitly define an activity, where its instances are almost always unique in some of its parts. To cope with this problem, we suggest generalizing activities into so called patterns. These patterns will be able to formally and explicitly define some parts of the examined activities. Having formal description of the pattern, pattern discovery service could search the subsequent event logs for the next occurrences of this pattern, greatly helping user in the identification and comprehension of what is happening in the virtual environment.

These patterns usually (see Fig.2 – current pattern elements are highlighted) lead to some critical moments in time, which can mean, for example, a significant progress, discovery of new knowledge/approach, or on the other hand they may indicate non-success of a particular process or its early finish. Such kind of patterns may also conceptually represent interesting learning paths emerged within particular user activities – either being positive (something like best practices), or negative (worst practices).

Patterns are manually selected and customized subsets of actions from the awareness repository. They are formalized as a sequence of activities, a list of pattern elements. Each pattern element represents one generalized event which is essentially a list of key-value pairs. User can specify the element based on any of the events properties, including custom ones. In these key-value pairs, user specifies which parts of the generalized event are important and which should be generalized. User can specify the element based on any of the events properties.

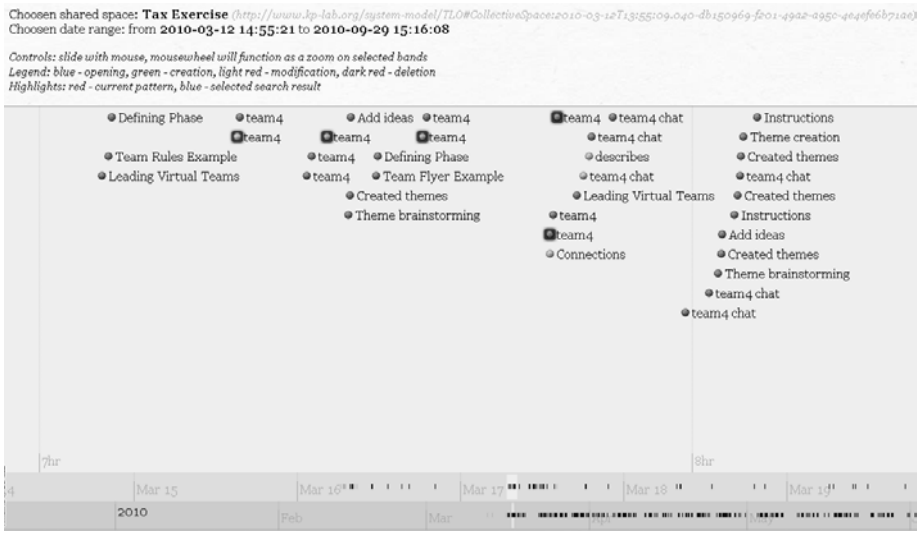


**Fig. 2.** Example of environment for pattern definition

### 3.1 Implementation

Pattern discovery service will find the matches of the given pattern within the specified time range, and within the activities by specified group of users, if given. The service returns list of matches, each represented as an array of events, comprising the given pattern.

Current version of pattern discovery service is implemented in the emerging lisp language called Clojure[2]. This dynamic programming language for JVM emphasizes functional approach to the programming and usage of immutable data structures. In Clojure, collections are generalized into the sequences, for which most of the operations provides lazy evaluation. For each pattern element, pattern discovery service construct an SQL query in order to find matching events in the log.

As the pattern is a sequence of such pattern elements, the resulting matches produce a tree of matches, in which the results are in the leaf nodes with given depth. A non-informed depth search is performed, collecting results on the way. The lazy evaluation, easily achieved in Clojure, realizes only those parts of the result tree, which are actually used. This subsequently lessens the number of queries sent to the MySQL database, dramatically speeding up the whole matching process, if the user is interested only in small number of results. Full source code can be found at [77].

[2] http://clojure.blip.tv/file/982823/

## 4 Conclusion

Patterns represent important way how to identify interesting sequences of user actions during collaborative activities in virtual environment. These sequences may lead to some critical moments or represent best practices in achievement of specified objectives. In our case patterns are manually identified by users based on created visualization through timeline. This visualization represents historical overview of performed activities/actions based on recorded data. Current version for basic patterns definition is implemented within KP-Lab System[3] and is being tested within organized pilot courses as supporting feature for complex evaluation of user practices. Current results from first testing phase are mainly oriented on usability of time-line visualization. Patterns implementation is still ongoing. List of user experiences and recommendations for this feature will be available in the fall 2010.

Future implementations of the pattern service plans to incorporate more flexible definition of a pattern, including the support for specifying relationships between individual elements and the ability to specify relevance for the individual elements, which helps to find better partial matches for a given pattern.

**Acknowledgments.** This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) (50%) and Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) (50%), both supported by the Research & Development Operational Program funded by the ERDF.

## References

1. Avouris, N.M., Dimitracopoulou A., Komis, V., Fidas, C.: OCAF: An Object-oriented Model of Analysis of Collaborative Problem Solving, Proceedings of CSCL 2002, Colorado, January, 2002, Hillsdale: Erlbaum, pp. 92-101.
2. Gomez-Aguilar, D. A., & Theron, R., & Garcia-Penalvo, F. J.: Semantic spiral timeline as a support for e-learning," Journal of universal Computer Science, vol.15, no.7, pp.1526-1545, 2009
3. Zaïane, O. R., Luo, J.: Towards evaluating learners' behaviour in a Web-based distance learning environment. In: 2nd IEEE international conference on advanced learning technologies (ICALT'01) 2001.
4. Hardless, C., Nulden, U.: Visualizing Learning Activities to Support Tutors, In Proc. CHI '99 Extended Abstracts on Human Factors in Computing Systems, CHI '99, 1999, pp.312-313.
5. Babič, F., Wagner, J., Jadlovská, S., Leško, P.: A logging mechanism for acquisition of real data from different collaborative systems for analytical purposes. In: SAMI 2010 : 8th International Symposium on Applied Machine Intelligence and Informatics, Herľany, Slovensko. IEEE, 2010. pp.109-112. ISBN 978-1-4244-6423-4.
6. Paralič, J., Richter, Ch., Babič, F., Raček, M.: Timeline-based analysis of collaborative knowledge practices within a virtual environment. I-KNOW 2010, Graz, Austria, in press.
7. Pattern discovery service: <http://kplab.tuke.sk/trac/browser/hpa/trunk/src/clj/hpa>

[3] http://2d.mobile.evtek.fi/help/index.php/Tutorial_Videos