Integration of a Flexible Analytics Workbench with a Learning Platform for Medical Specialty Training

Tilman Göhnert
University of Duisburg-Essen
Lotharstr. 63/65
47048 Duisburg, Germany
goehnert@collide.info

Sabrina Ziebarth
University of Duisburg-Essen
Lotharstr. 63/65
47048 Duisburg, Germany
ziebarth@collide.info

Per Verheyen
University of Duisburg-Essen
Lotharstr. 63/65
47048 Duisburg, Germany
verheyen@collide.info

H. Ulrich Hoppe
University of Duisburg-Essen
Lotharstr. 63/65
47048 Duisburg, Germany
hoppe@collide.info

ABSTRACT

In this paper we present a generic and extensible analytics workbench and show how it can be connected to learning platforms in order to analyze the activities on the platform from different perspectives. We show that as the analytics workbench already supports a wide range of analyses like network analysis, statistical analysis, and analysis of activity logs, the main effort needed for connecting a learning platform to it lies in transporting the log data of the platform into the workbench. However the analytics workbench is also designed for extensibility so if desired more specific analysis capabilities can be added to it easily. We present an analysis of the online platform of the KOLEGEA research project for supporting medical doctors in their specialty training as a case example for how the integration can be done, for what kind of analyses can be conducted, and for how the analysis results can help both the developers and the operators of such a learning platform.

Categories and Subject Descriptors

D.2.13 [**Software Engineering**]: Reusable Software; D.2.11 [**Software Engineering**]: Software Architectures; K.3.1 [**Computers and Education**]: Computer Uses in Education

General Terms

Design, Measurement, Experimentation, Human Factors

1. INTRODUCTION

Software tools for supporting learning often do not provide support for learning analytics. If there is analytic support (e.g. in larger open source projects like Moodle[1] or commercial solutions), it is mostly integrated into the specific learning environment and cannot be easily reused for other environments. Thus, much effort is invested in implementing even basic analytical features again and again for different learning environments.

In the context of web analytics the need for simple and economic solutions for getting (statistical) information about the visitors' behavior on web pages led to generic tools like Piwik[2] or Google Analytics[3], which can be easily applied to arbitrary web pages and portals. Regarding learning analytics there are first approaches for generic analysis environments. For example, the LeMo application provides tools for monitoring of learning processes on arbitrary learning management systems (LMS) [1]. However, for using these tools the contents and logs of the LMS that should be analyzed have to be transformed into the comprehensive and very specific LeMo data base structures. There are only few connectors, e.g. for Moodle or Clix, thus for most LMS there would be much effort to transform the data. Furthermore, LeMo is focussed on analysis of LMS, which are only one aspect of learning environments. In general, the essential challenge is the flexible integration of a generic analysis module with different types of learning environments.

We have developed an analytics workbench[4] [5], which covers a wide spectrum of analyses and can be easily applied to a wide range of data. This includes data from arbitrary learning environments, since lots of standard data formats are supported and the internal data representations are simple and generic, thus most data can be mapped. The workbench can be used exploratively by analysts, but it can also be used for performing stored analysis workflows automatically and thus for providing analysis results for other user groups [7]. In this paper, we present the analytics workbench and show how it can be easily coupled with an existing learning platform and used for meaningful analyses regarding different target groups. As a case example we use the integration of the analytics workbench with the KOLEGEA occupational learning platform for doctors in specialty training [10].

2. THE ANALYTICS WORKBENCH

Major parts of the analytics workbench as it is now have been developed in the context of the SiSOB[5] project. This project, funded by the European Community under the Science in Society (SIS) theme, aimed at measuring impact of

---

[1] http://moodle.com/

[2] http://piwik.org/

[3] http://www.google.de/intl/en/analytics/

[4] http://workbench.collide.info/

[5] http://sisob.lcc.uma.es/

science and research on society. The overall goal of the workbench development in this project was building a generic and extensible analysis framework with an integrated user interface that would enable even non-computer experts to access the full analytical power behind the tool and that would also allow reusing and sharing the created analysis workflows.

As depicted in Figure 1 the workbench offers a web-based user interface for designing analysis processes. The workflows are represented in a visual language based on a pipes-and-filters metaphor, in which modules of the language represent analysis steps and links between these modules describe the data flow. In this representation workflows can be stored, loaded, and also shared with other users of the tool. This user interface is backed by a multi-agent system and each of the modules in the visual language corresponds to one agent in the backend.

A wide range of analysis modules are currently available. Apart from quite generic components for putting data into the workbench, retrieving data from the workbench, or duplicating data for splitting a workflow into parallel branches, there are also over ten different modules connected to processing and analyzing graphs (including several modules for community detection and a module offering a wide range of centrality measures), around ten modules connected to processing and analyzing activity logs (including modules for deriving statistical information, creating networks, and doing sequence analysis), and a wide range of different visualizations for both graphs and statistical information.

The evaluation studies of the analytics workbench conducted during the SiSOB project[6] showed that researchers from different fields working in the area of scientometric research experienced the workbench as a tool they could imagine to use in their day-to-day work. One aspect that was specifically highlighted were the advantages of the explicit workflow representation in the pipes-and-filters metaphor. The study participants saw re-usability of workflows, the possibility to exchange workflows with other researchers, and the possibility to communicate about the workflows based on this representation as something that would be very beneficial for themselves. First results of an evaluation study conducted in the context of teaching social network analysis using the workbench in a master level university course confirm these findings.

2.1 Data handling

In principle the formats used for data exchange between the individual analysis components may be chosen completely arbitrary as long as each agent in an analysis workflow understands the input it is given. However in order to facilitate an easy data exchange between the individual agents, some formats have been chosen as main exchange formats. The format used for data tables and the format used for graph data both stem from the SiSOB project[7]. Both are very flexible JSON based formats which allow converting to and from almost any other data format available for these kinds of data.

Currently the workbench offers transformations between the SiSOB Graph Format (SGF) and the edge list graph format used by the Pajek[8] network analysis tool, the adjacency matrix format of the UCINET[9] tool, and GML[10], a very widespread and expressive graph format, which can be used for example as input and output format of the igraph[11] network analysis library. The SiSOB Datatable Format (SDT) can be converted to and constructed from comma-separated value (CSV) files in order to allow data exchange with almost any tool that can handle tabular data.

The third main data exchange format used in the workbench is the Activity Streams format[12]. It is also a JSON based format and offers a quite flexible representation for activity log data.

2.2 Architecture and extensibility

As already stated the workbench combines a web-based user interface with a multi-agent system as analysis backend. The system uses an SQLSpaces server [9] as communication and data exchange platform. The SQLSpaces system is an open source implementation of the "tuple space" [4] concept focusing especially on ease of development using the server as communication platform for distributed multi-agent systems and on the support of language-heterogeneity in these systems. Supported programming languages include Java, Python, and JavaScript. Figure 2 shows an overview of the architecture.
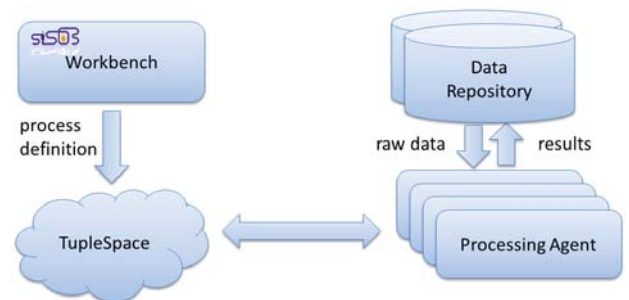

Figure 2: Architecture of the SiSOB analytics framework.

The communication between the elements of the analytics framework is based on simple tuple-based protocols. There are specific tuple formats for the different communication purposes with the most important being those for initiating and controlling analysis processes or those for data exchange between the agents.

While agents offering analysis techniques are the majority in the system, agents also control input and output of data to and from the analysis process. Therefore connecting the system to new data sources, feeding output of workflows to other systems, and adding new analysis techniques can

---

[6]http://sisob.lcc.uma.es/repositorio/deliverables/
D63-Final.pdf

[7]http://sisob.lcc.uma.es/repositorio/deliverables/
SISOB-D52.pdf

[8]http://pajek.imfm.si/doku.php?id=pajek

[9]https://sites.google.com/site/ucinetsoftware/

[10]http://www.fim.uni-passau.de/fileadmin/
files/lehrstuhl/brandenburg/projekte/gml/
gml-technical-report.pdf

[11]http://igraph.sourceforge.net/
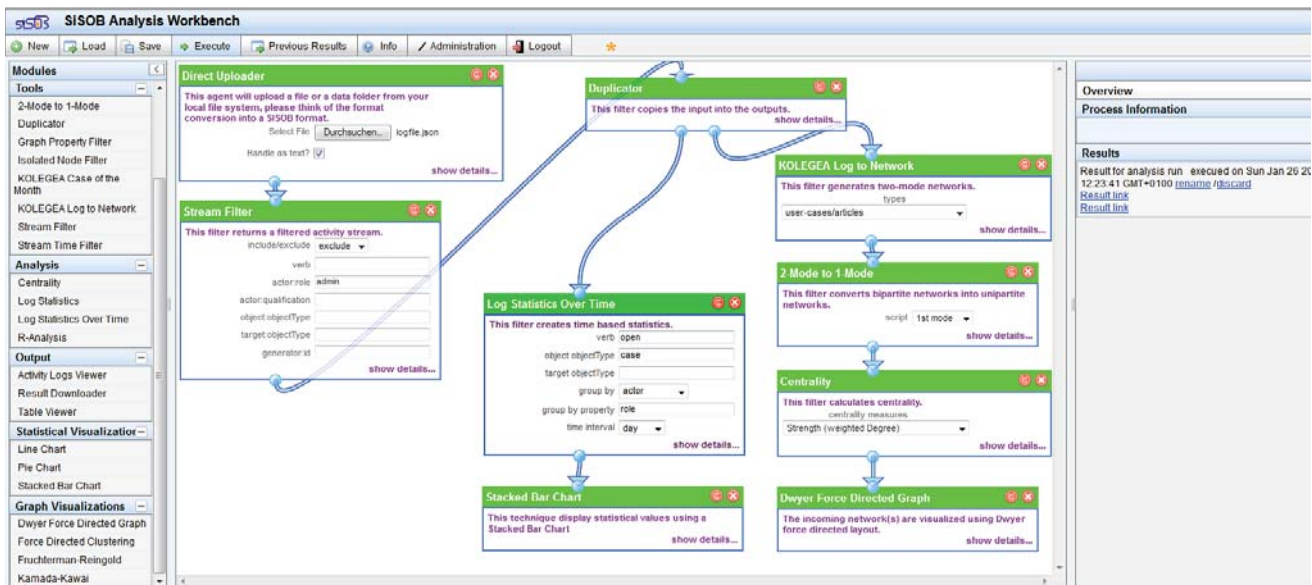
[12]http://activitystrea.ms/

Figure 1: SiSOB analytics workbench user interface.

be realized by adding new agents to the analytics workbench. Apart from the technical basis, the connection to the SQLSpaces, a newly developed agent only needs to comply with the tuple protocols used in the system. There is no further restriction regarding the structure or the functionality of the agents.

Another possible entry point for adding new functionality to the analytics workbench is the *R-Analysis* agent, which acts as a wrapper for scripts to be executed in R[13], a language for statistical computing. As there is a wide range of R libraries for different analysis purposes including network analysis available, this allowed and still allows easy integration of analysis features based on these libraries. One example for the integration of experimental or newly developed analysis techniques based on R is the integration of a novel approach for multi-relational blockmodeling [5].

3. THE KOLEGEA LEARNING PLATFORM

The project KOLEGEA[14], which is facilitated by the German Federal Ministry of Education and Research, aims to support doctors specializing in family medicine (general practice (GP)) by providing a platform for collaborative learning in occupational, social communities. It thus addresses the problem of missing opportunities for networking, which highly restricts occupational knowledge exchange as well as collaborative learning in peer communities among young doctors. This problem results amongst others from the many job changes which are required by the broad curriculum of family medicine. Since GP specialty training is based on learning by solving real problems/cases in every day working life, learning is problem-based, self-directed and based on intrinsic motivation. Thus, KOLEGEA's pedagogical approach (see also [10]) is focused on collaboratively working with user-generated cases in the spirit of problem-based

learning (PBL) (see e.g. [2]). Users can share and discuss cases in self-regulated or mentor-supported small groups and share the results with the community. Cases can be enriched with media like pictures or videos, tags and links to medical guidelines or external online information like recent publications. Apart from the work in small groups, there are tools for community support like forums, which can be used for occupational, but also social exchange.

The current version of the KOLEGEA system consists of a web portal[15] and mobile applications (tablet[16] and smart pen) for multimedia note taking and editing. Smart phone apps for (simple) note taking and accessing the system are currently in development.

The portal provides access to the virtual community and its knowledge artifacts (user-generated content as well as medical guidelines and links to artifacts of current research). Figure 3 shows an example of a case description, its discussion and metadata (e.g. tags). The case description is structured based on the phases of the medical consultation. Following Steinkuehler et al. [8], we consider group discussions to be asynchronous and threaded, since this can produce discussions and results of higher quality than synchronous communication and is more flexible regarding time. Discussions can take place directly connected to cases, but also in the community's forums.

"An important issue in groupware and CSCW is awareness – generally having some feeling for what other people are doing or having been doing" [3], since "the success of a collaborative learning experience depends on the informed involvement of curriculum designer, teacher, evaluator, and students" [6]. Each of these roles needs different awareness information. In KOLEGEA target groups for aware-

---

[13]http://www.r-project.org/
[14]http://www.kolegea.de/

[15]https://beta.kolegea.org/kolegea/
[16]https://play.google.com/store/apps/details?id=
info.collide.kolegea.npad.views&hl=en
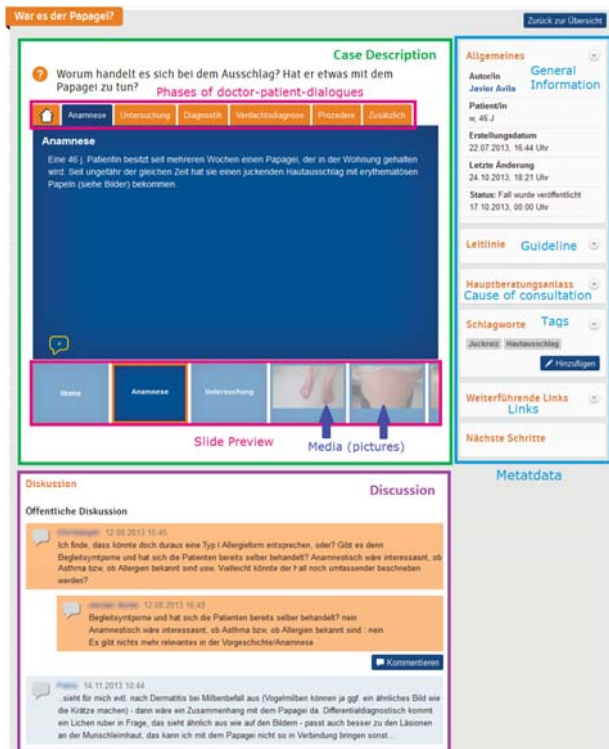
Figure 3: KOLEGEA Web Portal – case representation with discussion.

Figure 4: Planned coupling of KOLEGEA system (orange) and SISOB workbench (blue) – the backchannel from the workbench to the platform is not yet available.

ness information are the researchers designing and evaluating the platform, the platform operators, the mentors and of course the trainees. Researchers and platform operators are interested in the success of the designed tools, processes and stimuli to further improve the system. Furthermore, platform operators need input for operating the platform like which case to select as "case of the month" or which trainees have the potential for supporting (or "tutoring") self-regulated groups or even become mentors after becoming medical specialists. Mentors e.g. need information on how the members of their groups work together to support collaboration. Trainees might be interested in the activity of groups and their topics before asking for membership. These are just a few examples for awareness needs in the KOLEGEA project.

4. INTEGRATION DESIGN

The analytics workbench provides many standard analysis modules for SNA and statistical analysis of log files. Furthermore, it can be easily extended for specific analyses needed for the actors of the KOLEGEA system. Since it is more efficient to reuse existing software (modules) instead of developing it from scratch (e.g. reduced implementation time, less bugs due to prior testing and usage) the analytics workbench was coupled with the KOLEGEA system for performing the needed data analysis.

In KOLEGEA all textual contents as well as the logging information are stored in a PostgreSQL data base, media files like images or videos are stored in the file system of the web server. While the data is directly accessed by the web
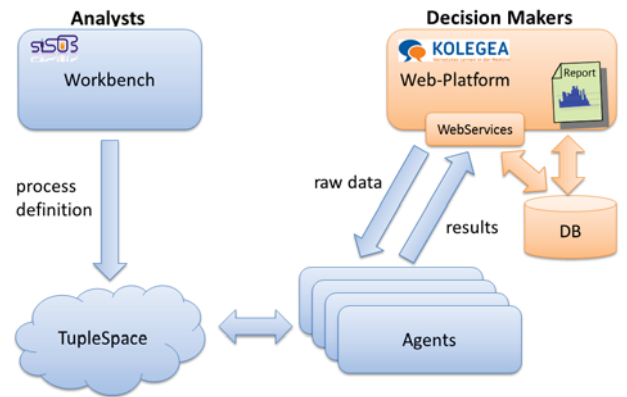
platform, the mobile applications use a web service interface (REST) to write and access data. The mobile devices also send their log files to the central data base using the web service interface. The KOLEGEA logging format is proprietary.

To integrate the analytics workbench with the KOLEGEA system the KOLEGEA WebServices were extended to provide raw data for the analysis (log files as well as information on objects and users stored in the data base). This data is preprocessed by a specialized module which converts the KOLEGEA log file format into the Activity Streams format that is supported by the workbench. Furthermore, there is a specialized module for creating actor-artifact networks. After this preprocessing, mainly generic modules for log analysis and social network analysis provided by the workbench are used for analyzing the data.

At the moment the analysis results are manually compiled into reports for the project members. We plan to extend the KOLEGEA WebServices to receive analysis results that are stored in the internal data base and provided by the platform. The complete analysis process (see Figure 4) will then contain the following steps: Analysts define analysis workflows in the workbench and save them in the tuple space. A scheduling agent takes the workflow definition and initiates its execution by writing the appropriate command tuples in defined intervals. In each workflow run specific input agents access the KOLEGEA WebServices to retrieve the data needed for the analysis. The specified analysis agents prepare and analyze this data and visualization agents create appropriate visual representations of the results (e.g. sociograms). The textual and visual results are transferred to the platform by an output agent, which accesses the KOLEGEA WebServices. These save the results in the platform's database to be retrieved and visualized on demand.

5. DATA ANALYSIS

The following sections describe exemplary analyses on different levels conducted with the analysis workbench to support different types of actors regarding the KOLEGEA system.

5.1 Effects of stimuli

Log statistics over time can be used to identify effects of stimuli. Figure 5 shows the number of distinct users per day in the "closed beta" phase of the project. Since the platform was only used by a closed group of users, there was no registration via the platform, but the login information was send via mail to the preselected test users. On this day, there was the highest amount of distinct users. After a week an activation mail was send to the trainees informing about new cases. This resulted in a higher number of distinct users. Direct contacts with the test users on the DEGAM day of family medicine as well as the later activation mail do not seem to have had much effect.
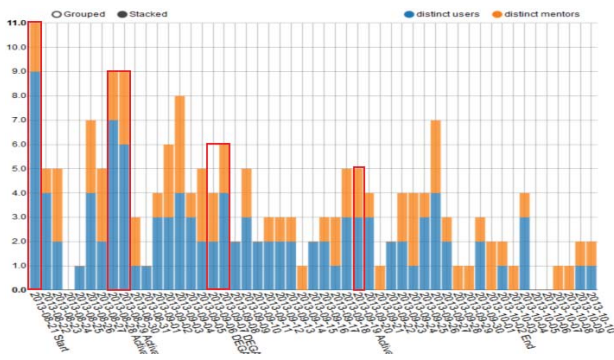

Figure 5: Distinct users per day in "closed beta".

This indicator was similarly used to evaluate the effects of events advertising KOLEGEA regarding registration numbers in the "open beta" phase in January 2014. Furthermore, we used login statistics for identifying a good weekly date for sending activation emails with information on recent platform events.

5.2 Recommendations for "case of the month"

In KOLEGEA each month one case is selected as "case of the month" and depicted on the top of the welcome page. While the case is ultimately selected by a group of doctors of the institute for family medicine of the Charité Berlin to guarantee the quality of its content, not all available cases should be considered to reduce their effort.

Figure 6: Indicators for "case of the month".

Indicators for interesting cases are a high number of comments and read events as well as distinct users and distinct returning users accessing the case. While most of this information could have been gained by using the existing generic components for log statistics, to have all information in one

table a specific "KOLEGEA Case of the Month" component was implemented. This component is applied after loading the logfile and filtering (excluding) actors of the role "admin" by existing components. The results (see Figure 6) are displayed by the table viewer, which allows sorting of the table rows by selecting a column caption.

The statistics were provided to the doctors as a basis for selecting the "case of the month" at the end of January 2014 for the first time. They chose a case with many comments and medium accesses, which was created by a doctor in training.

5.3 Identification of user roles

While the previous examples are mainly based on descriptive statistics generated from logfiles, in this analysis the log information is aggregated into a network for network analysis.
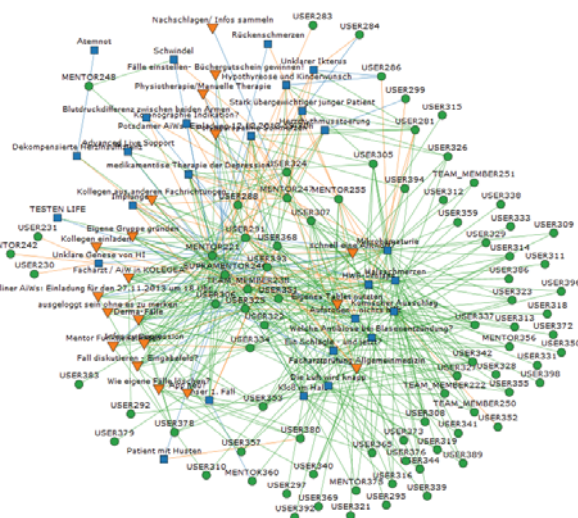

Figure 7: KOLEGEA community (three-mode network).

Figure 7 shows a three-mode network of the complete KOLEGEA community containing actors (green dots), cases (blue squares) and forum threads (orange triangles). There are three different types of edges: green edges indicate that an actor only read an artifact, orange edges show that the actor at least added one comment or tag to a case/forum thread and blue edges identify the creators of the artifacts. Thus, comments and tags are already aggregated in the edges.

Most actors only interact with artifacts by reading them, which is typical for online communities. The community has a core of actors being connected with many artifacts and a periphery of actors being only connected to few artifacts. The core contains two mentors, a KOLEGEA team member and ten users. Users with a high degree centrality regarding reading edges and no or little writing edges seem to be very interested in the artifacts and might need just a little nudge to participate actively.

Regarding the identification of potential tutors and mentors highly active users are interesting. Considering only

create (case/forum tread/comment) edges, the degree centrality respectively strength (weighted degree centrality) is an indicator for the level of activity in the platform. The the degree centrality of the cases in the multi-mode network corresponds to the number of distinct users in the "case of the month" statistics in section 5.2 .

Figure 8 shows an one-mode actor network, which is created by "folding" the actor-artifact network ignoring "read"-edges. From the point of view of data aggregation this folding operation transforms the previous multi-mode network. The nodes are sized regarding their betweenness centrality. A high betweenness centrality is an indicator that the respective actors have a mediator role in the community, thus these might especially qualify for becoming tutors or mentors. The four most central actors are three mentors and one user. This supports that betweenness centrality is a good indicator for identifying mentors/tutors. The high centrality of the mentors is a consequence of building a new community ("cold start problem"). Furthermore, user325 appears to be a good candidate for becoming a tutor.
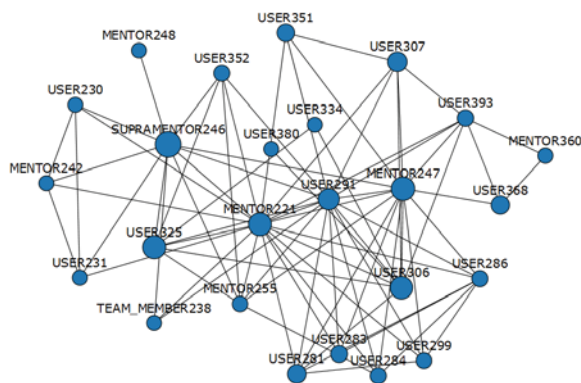

Figure 8: KOLEGEA community (one-mode actor network).

6. SUMMARY AND PROSPECTS
In this paper, we have shown how a generic and extensible analytics workbench can be integrated with an online learning platform for medical doctors in specialty training for performing analyses. We also have presented how the results of different analysis processes with different perspectives on the available data can be used to inform researchers developing online learning platforms and operators of such platforms. The analyses addressed miscellaneous levels of data aggregation ranging from basic log data to one-mode networks. There are relations between the analyses, e.g. the number of distinct users linked to a case can be extracted directly from the log file or as the degree centrality from a case-actor network. An actor-artifact network, which is based on a logfile, can be aggregated into an actor-actor network for getting a more condensed view on user interactions.

In our future work, we plan to extend the work presented here in several ways. One direction will be the extension of the workbench with further analysis modules focused on the analysis of activity logs. Another direction will be the application of more of the already implemented analysis ap-

proaches on the log data of online platforms, e.g. the available sequence analysis modules. We also plan to build a collection of reusable analysis workflows that can be applied to different online platforms and that can support different target groups. In addition to researchers developing new approaches for learning platforms and operators of such platforms, we also plan to support the users of the platforms directly with awareness features. A last goal for both the development of the analytics workbench and the KOLEGEA platform will be integrating the two in such a way that it is possible to directly view the results of the workflows in the platform and thus giving especially the users of the platform an easy access to the analysis results.

7. REFERENCES

[1] L. Beuster, M. Elkina, A. Fortenbacher, L. Kappe, A. Merceron, A. Pursian, S. Schwarzrock, and B. Wenzlaff. Lemo: An analytics tool for lmss as well as portals. Poster Presentation at the LAK 2013, 2013. Online available at http://lemo.htw-berlin.de/public/publication/LAK13_Beitrag.pdf, last checked on 2014-01-26.

[2] L. H. Distlehorst and H. S. Barrows. A new tool for problem-based, self-directed learning. *Journal of Medical Education*, (57):486–488, 1982.

[3] A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human-Computer Interaction*. Prentice Hall, 2004.

[4] D. Gelernter. Generative communication in linda. *ACM Trans. Program. Lang. Syst.*, 7(1):80–112, jan 1985.

[5] T. Göhnert, A. Harrer, T. Hecking, and H. U. Hoppe. A workbench to construct and re-use network analysis workflows - concept, implementation, and example case. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2013.

[6] C. Gutwin, G. Stark, and S. Greenberg. Support for workspace awareness in educational groupware. In *The First International Conference on Computer Support for Collaborative Learning*, pages 147–156, 1995.

[7] N. Malzahn, T. Ganster, N. Sträfling, N. Krämer, and H. U. Hoppe. Motivating students or teachers? challenges for a successful implementation of online-learning in industry-related vocational training. In *EC-TEL 2013*, pages 191–204, 2013.

[8] C. A. Steinkuehler, S. J. Derry, D. K. Woods, and C. E. Hmelo-Silver. The step environment for distributed problem-based learning on the world wide web. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, pages 217–226, 2002.

[9] S. Weinbrenner. *SQLSpaces – A Platform for Flexible Language-Heterogeneous Multi-Agent Systems*. PhD thesis, Universität Duisburg-Essen, 2012.

[10] S. Ziebarth, A. Kötteritzsch, H. U. Hoppe, L. Dini, S. Schröder, and J. Novak. Design of a collaborative learning platform for medical doctors specializing in family medicine. In *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings*, pages 205–208, 2013.