The ”Concept Cloud”: Supporting Collaborative Knowledge Construction based on Semantic Extraction from... Go-Lab View project The ”Concept Cloud”: Supporting Collaborative Knowledge Construction based on Semantic Extraction from Learner-generated Artefacts

Sven Manske, Ulrich Hoppe

The " Concept Cloud " : Supporting Collaborative Knowledge Construction based on Semantic Extraction from...

Sven Manske
University of Duisburg-Essen

The "Concept Cloud": Supporting Collaborative Knowledge Construction based on Semantic Extraction from Learner-generated Artefacts

Sven MANSKE, H. Ulrich HOPPE
University of Duisburg–Essen, Faculty of Engineering, Duisburg, Germany
manske@collide.info, hoppe@collide.info

Abstract—**Explicit visual representations of domain knowledge have the potential to support students engaged in scientific inquiry learning activities on an epistemic level. This can be facilitated using computational methods for the extraction of concepts from student generated knowledge artefacts such as hypotheses, concept maps, or wiki articles. We propose an application of this approach in the context of inquiry learning with online science laboratories. As a cognitive awareness tool, the "concept cloud" presents domain concepts and key phrases to the learners in order to help them reflect on their own learning and knowledge building. As part of a learning analytics toolset, the concept cloud also supports teachers in supervising their students' knowledge building from an epistemic perspective. The approach has been tested in a classroom scenario with 84 secondary high school students.**

Cognitive awareness; reflection support; learning analytics; inquiry-based learning; online experimentation

I. INTRODUCTION

The on-going European project Go-Lab provides access to a large collection of existing online laboratories for student experimentation based on a unified interface and platform [1]. The availability of a large variety of labs in various domains of science makes this approach widely usable and scalable. Thus, Go-Lab can help to overcome existing deficits in science education [2, 3]. The Go-Lab approach supports student-centred and supportive teaching methods aiming to convey key competencies through active, experimentation-based learning in contrast to traditional deductive and teacher-centric approaches [4]. Here, students adopt the role of researchers to investigate the big ideas of science.

The Go-Lab approach combines both regular classroom activities and in-class online learning. It provides teachers with facilities to create customized learning spaces around a given virtual or remote laboratory [1]. These spaces are organized as subsequent phases and orchestrate an inquiry-based learning (IBL) process [5]. Learning spaces can be enriched with resources and supportive apps. Specific apps enable the students to generate knowledge artefacts such as wiki articles or concept maps. Due to the heterogeneity between different artefacts of these types, the interpretation of learner-generated content is challenging for both students and teachers, which underlines the need of an aggregated representation of the content to support learning and knowledge building on an epistemic level [6, 7, 8].

To meet this objective, and to support students and teachers, we propose the use of computational methods of semantic extraction to better understand and reflect on the activities in the Go-Lab online learning environment. We present an application in the context of Go-Lab that creates a structured visual representation of semantic concepts ("concept cloud") based on a group-oriented model of (shared) student knowledge. The "concept cloud" aims at supporting students in reflecting their own knowledge building as well as teachers in supervising their students in online inquiry-based learning through the Go-Lab portal[1].

To evaluate this approach, we have conducted an experiment in which students carried out a learning activity through the Go-Lab portal using (among others) the concept cloud application. Usually it is expected that after reflection phases, students re-visit or review formerly created artefacts and modify them to improve the artefact quality. The use of the concept cloud app should enforce students' reflective behaviour and should have a beneficial effect on the overall learning outcome. We evaluated this hypothesis in a real classroom setting where the students had to carry out an inquiry-based learning-task. We used a triangulation approach and captured both artefact- and behaviour-related data of the learners and teachers.

In the next sections we present a short overview of the state of the art, the experimental setting of the study, and our conceptual framework. We present the analysis and evaluation of the results, a discussion on the finding of the study, and we conclude with the outcome and future work.

II. BACKGROUND AND RELATED WORK

The notion of Learning Analytics (LA) stands for the usage of computational methods of analysis on learning data to inform various stakeholders with the aim to improve learning processes and environments [9]. In this sense, the concept cloud can be conceived as a LA tool open to be used by learners and teachers. We can distinguish three types of computational methods used in LA: (1) The analysis of content using text mining or other techniques of artefact analysis, (2) the analysis of processes using methods of sequence analysis, and (3) the analysis of (social) network structures. In the context of Go-Lab, network analysis (3) has minor relevance due to its design principles and the lack of explicit (social) relations. The analysis of learning Processes (2) following a model of learning phases in IBL has been applied using methods of sequential pattern mining [10].

[1] Go-Lab portal: http://golabz.eu (retrieved 3rd Feb, 2016)

Content-based analyses (1) have so far received less explicit attention from an LA perspective, although computational linguistics techniques have been successfully applied to the analysis of collaborative learning processes [11]. The OpenEssayist-System [12, 13] uses linguistic approaches to analyse text-based artefacts generated by learners. While the outcome of the system aims to adapt learners to write essays with a specific structure, it underrepresents epistemic aspects of learning analytics. Students are forced to think about why the system produces a certain outcome, but it does not provide any help by delivering explanatory models.

To support the revision or the evaluation of students' learning outcomes effectively, learning analytics applications might need to process learner-generated content automatically. The processing of text-based representations, e.g., wiki articles, highlights the importance of semantic technologies for LA. Several systems use semantic technologies to represent knowledge. The Xerox Incremental Parser (XIP) Dashboard uses approaches of Natural Language Processing to aggregate salient sentences of scholarly papers, providing a variety of analytics to the students [14]. The Debate Dashboard [15] focused on distributed knowledge management. A central aspect is an argument mapping tool with a visual component to support collaborative work and collective sense-making.

To effectively provide scaffolds for the students' interaction with learner-generated content, the design of LA interventions as proposed by Wise [16] can serve as a point of reference. An LA intervention can be defined as a frame in which analytic tools, data, and reports are gathered and used. Wise formulates four design principles for the successful integration of LA tools: The smooth *integration* of LA results into the learning environment; the *agency* in interpreting and responding to analytics in terms of supporting and not detracting learners; setting up a *frame* to give learners a comparison point when interpreting analytics results; and the chance to discuss and negotiate the LA results in a *dialogue*. These statements have been extended by Harrer et al. [17] adding the two principles *scope*, focusing on contextually relevant information for the learner, and *representational consistency*, i.e. adapting the interface to the learning environment.

Apart from the question of what data to visualise, how to support pedagogical interventions and present them to learners, the nature of knowledge and its implications to learning is relevant. Epistemic aspects are framing research in learning analytics [18]. Dimensions of epistemic beliefs highlight the learner's view on what is to be learnt, particularly classifying knowledge on part of the students [19, 20].

The design of the concept cloud was influenced by Wise's framework for learning analytics interventions and incorporates Mason's epistemic beliefs and aspects to foster reflective behaviour on part of the learner and connect the visual items to students' knowledge. Similar to the OpenEssayist system, it uses a tag-cloud-like visualisation to display key concepts used in the learning spaces, particularly the learner-generated content. Such representations support the user in monitoring a large number of items and thus provide a medium the learner can manipulate and interact with [21], reducing navigational paths [22].

III. Semantic Extraction: Applications for Learning Analytics

In this section, we describe the conceptual framework for our approach of guidance based on semantic technologies in the context of IBL. As stated in the background, learning analytics aims at improving learning. Semantic extraction as a means of LA is utilised to provide deeper insights into knowledge structures on part of students and teachers. Thus, such models that externalise knowledge and representations of these structures can be used to make better-informed decisions or place tailored pedagogical interventions. Typical applications of learning analytics focus on performance- and activity-related data. Examples are descriptive statistics about time spent on specific production tools or the number of resource accesses. Besides the problems of interpretability, these statistics focus on supporting the ex-post analysis phase on part of the teachers. Guidance mechanisms targeting students are typical scaffolds on a micro-level, e.g., prompts inside a specific tools. These mechanisms demand a high degree of adaptation and encode very narrow domain knowledge or match simple and hardcoded interaction patterns. Our work aims to provide more general mechanisms to support students and teachers by enforcing critical thinking, initialising reflective processes and using representations of knowledge structures that are connected to the actual learning productions and thus to the learning outcomes. Scaffolds in Go-Lab enable the learners to create artefacts as externalisation of their knowledge structures [23] and to provide guidance through the inquiry learning process. The importance of these scaffolds has been pointed out for the engagement [24] and the support [1] of the learner as well as for fostering critical thinking and the development of 21st century skills [25].

As part of this framework, we present an application that is using an explicit representation of knowledge items.

A. Conceptual Framework

The underlying conceptual framework of our work is presented in this section. Figure 1 shows the layers starting from the pedagogical framework of inquiry learning to the data model of a concept cloud.

In IBL scenarios such as Go-Lab, students create artefacts across the different inquiry phases. Typically, these activities are pre-structured in a way that each inquiry phase offers specific production tools or scaffolds to support learning activities and processes. Examples of these scaffolds are apps for concept mapping, hypothesis creation or text editors (wiki texts). Most of the learner-generated content that has been created through these scaffolds is text-based or in a format which is very similar and can be reduced to a textual representation without a major loss of information. Each artefact consists of a set of knowledge items that refer to specific concepts in the domain of the inquiry learning space. These concepts are extracted from the artefacts and added to each student model, which is a collection of all concepts used in all phases by a student. Thus, such a student model contains all knowledge items of a student that are relevant to the connected science domain. Semantic technologies such as DBPedia use ontologies to disambiguate terms dependent on the context of terms in a certain (scientific) domain [26]. The concept cloud data model contains all student models for the whole learning group connected to a single learning space. The data model of the concept cloud can be used to further generate external representations of the knowledge items, e.g., through the concept cloud app.

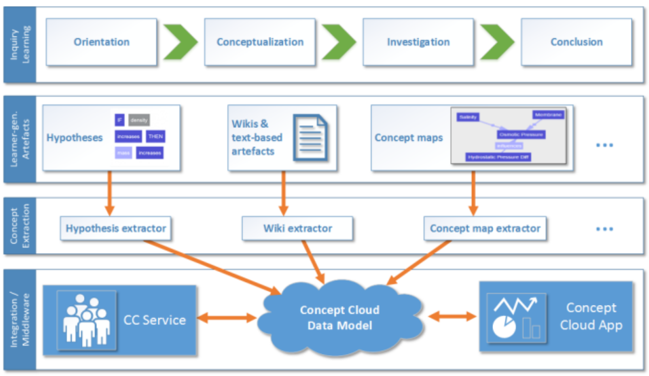

Figure 1. Conceptual Framework for the concept cloud.

B. Data Model and Extractors

Depending on the conceptual model above, each artefact type (e.g., concept map, text, etc.) needs its own mechanism for the extraction of relevant concepts. The general approach for a concept extractor is to map an artefact to a list of concepts that represent its inherent knowledge items. All artefacts of all students are aggregated into a unified data model, which needs to follow the approach of a common denominator sacrificing structural aspects, particularly relations between concepts (in a concept map) for the purpose of unification.

Hypotheses contain operators, relations, quantifiers, and parameters. Only the parameters relate to important concepts, these are extracted from the artefacts. Relations are removed as they cannot be represented in the loose structure of a tag cloud. The same applies for the relations in concept maps, where only the list of concepts is used. The extraction of concepts is more challenging for text-based artefacts such as wiki articles. For this purpose, we facilitate existing semantic technologies: Currently, two different extractors exist: a) DBPedia Spotlight, and b) AlchemyAPI [27]. Both have different characteristics: Each concept extracted in DBPedia Spotlight is matched with an underlying ontology and corresponds to a domain concept. This restricts the output to specific terminology used, which might be connected to declarative knowledge of the learner. The AlchemyAPI extractor outputs phrases, which leads to a more diverse aggregation, but also displays knowledge items that cannot be expressed in a single domain concept. Such phrases might be connected to procedural and metacognitive knowledge.

C. Concept Cloud App

The main idea of this app is to use an external and explicit representation of the data model to be presented to teachers and learners. The app is embedded into the Go-Lab environment and can be used in different occasions depending on the instructor. In the evaluated scenario, the app has been used by the students during an IBL activity with Go-Lab. Figure 2 shows the concept cloud from a student's perspective with data from a learning space. The different colours refer to the usage of concepts across the different inquiry phases by the student. Red means that the student has not used the concept while others did so. Green stands for a consistent use of the concept across different phases and production tools, while yellow stands for inconsistency in this sense.

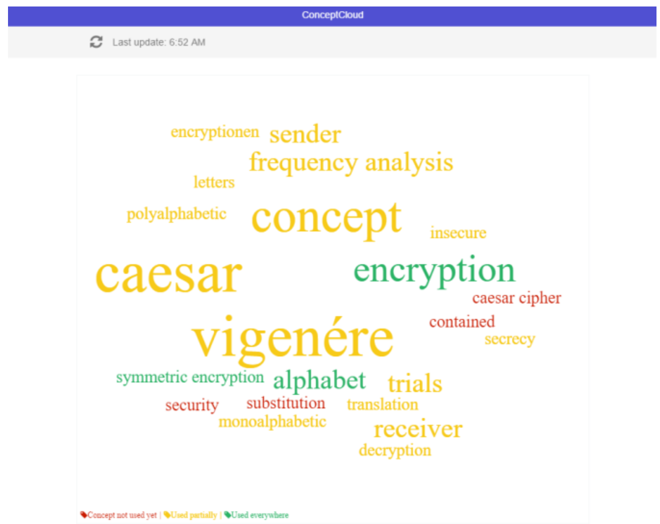

Figure 2. Conceptual Cloud App: the data is (translated) from a student of experimental condition C2.

IV. Evaluation

The concept cloud aims at supporting students' reflection and epistemic aspects. We expect students to revisit and edit their generated artefacts and modify it after spending time on the concept cloud. We analyse the sequence of actions in order to find these *revision patterns*. We define additionally to the structure a condition to ensure a temporal closeness to the concept cloud visit.

Regarding these patterns, we also investigate the time spent on the concept cloud depending on the different experimental conditions as well as descriptive and activity-related statistics that might have implications on the learning activity. We tested several experimental conditions:

C1: no concept cloud (control group)
C2: concept cloud with DBPedia Spotlight engine
C3: concept cloud with AlchemyAPI engine
C4: concept cloud without a real-time engine (static)

A. Learning Scenario

The classroom experiment was conducted in three computer science classes and three didactical lessons each at a German secondary school including 84 students aged fourteen to eighteen years. The objective of the learning scenario for this study was to learn about basic encryption algorithms and the difficulty of decryption using online labs. During the activity, each student created four short wiki articles, one concept map, and a set of hypotheses. These artefacts were used for the assessment of performance characteristics. The class was split up into experimental groups C2 ($n = 20$), C3 ($n = 14$), and C4 ($n = 10$) and control group C1 ($n = 40$). In the test conditions C2, C3, and C4, the concept cloud appeared in the ILS for the students. In a pre-study we found out that it is necessary to force the students to actively engage with the app in contrast to have it on demand in an additional phase that is optional for them and they might skip it. The control group C1 used a similar ILS without the concept cloud. The teacher was advised to monitor the classroom activity.

B. *Results*

Figure 3 gives an overview about a few activity metrics that are relevant regarding our hypothesis of reflective behaviour, the corresponding patterns, and activity on specific artefacts. Of course, these metrics do not state out any benefits on the quality of the learning outcome of the students. The results indicate that the concept cloud app forces students to revisit their artefacts and fosters a higher engagement with the system, as implied by the activity metrics. In condition C2 in 92.86%, and in C3 in 70% the learners revised and modified their artefacts.

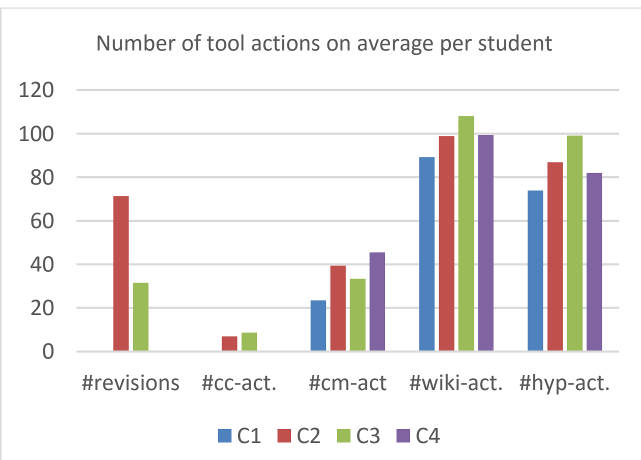

Figure 3. App-related user activity: number of revision patterns, concept cloud, concept mapper, wiki and hypothesis tool actions

An important aspect when dealing with these explicit knowledge representations is if students use this to cheat the system. For example, they could use all the important concepts that have been used by the majority to create artefacts such as concept maps that contain these key concepts without a deeper level of understanding or reflection. Therefore, we analysed students' action sequences to detect this behaviour. Indicators for this are certain revision patterns that involve productions. We investigated Revision patterns (a) without non-productive actions, (b) with production-only actions, (c) patterns separated per artefact, (d) time spent on each revision, (e) concept occurrences in the concept cloud and each artefact, and (f) sequences of added and revised concepts.

Although hypotheses play a major role in IBL scenarios, there are no relevant findings regarding the artefact analysis of hypotheses for this experiment. We could not observe any revision after using the concept cloud.

When investigating the produced artefacts, it turned out that students who used the concept cloud in the proper way created better concept maps regarding the coverage of key concepts. We argue that a meaningful usage pattern would incorporate time and certain consecutive actions. An ideal use of the concept cloud would be the following example: a student creates an artefact during the learning scenario that is later followed by a reflection period when she uses the concept cloud. After a certain time, she revisits the artefact and a) leaves it because she is satisfied with the relative quality compared to others, or b) edits the artefact and proceeds with the learning activity. A problematic pattern will be an oscillation between the artefact modification and the concept cloud, which indicates the aforementioned "cheating the system"-behaviour. We could not observe a single case using this pattern in the experiments.

When we compare the ideal use cases of the concept cloud with the non-ideal cases leaving out the "mock"-condition C4, (50 non-ideal, 14 ideal cases), we observe major differences in the artefact creation. Concept maps have been created after using the concept cloud. This implies that the results of the terms visualised in the app might influence the creation of the map. The average number of concepts per concept map is with 5.04 very statistically significantly lower compared to the average of 11.29 in the ideal group ($p < 0.01$, unpaired two-tailed t-test,). In all of the cases of ideal usages, the concept maps are only with a few exceptions composed of relevant terminology according to the learning design (defined by the teacher), which implies that the concept map quality in terms of coverage increases proportionally to the number of concepts. In conclusion, concept maps by "ideal" concept cloud users are likely to be better than the maps from students who did not spend any or not enough time on the concept cloud.

According to Wilson, the development of epistemic fluency – the ability to – is potentially supported through rich information sources such as the concept cloud, when learners participate actively [8]. This seems to be in line with the combination of concept cloud and concept mapper, as tools to support knowledge construction as of Wittrock [28]. Following this, the concept cloud is likely to be a useful scaffold in combination with other production tools.

V. Conclusion and Discussion

In this paper, we discussed the use of semantic technologies for learning analytics, particularly to support guidance of students and teachers. We introduced a conceptual framework for guidance based on semantic technologies, which aggregates knowledge items into a single data structure. The framework has been applied in the concept cloud app, which aims at supporting knowledge construction and reflection on the part of the students. For teachers, it gives insights into knowledge structures of the students in a learning group as a learning analytics tool for teachers.

The idea of content-related reflection is facilitated when it comes to the idea of teacher-led inquiry: teachers perceive and design their own teaching as an experiment, following an inquiry-based approach and further develop their teaching materials [29, 30] . The concept cloud app supports this idea: in contrast to process learner productions, visualising the extracted concepts from the teaching materials, the teacher might be (positively) influenced by reviewing the usage of specific concepts across the different inquiry phases, which will uncover potential inconsistencies about key concepts that are relevant for a didactical unit or across different lessons. Future evaluations need to prove the use of this application to support teacher-led inquiry and its impact on teaching quality.

The conducted experiment demonstrated how learning analytics applications in conjunction with inquiry apps support and scaffold learners in their knowledge construction, and teach them epistemic fluency on their way to become a 21st century learner.

Acknowledgment

The Go-Lab project is partially funded by the European Union (Grant Agreement no. 317601) under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (FP7). This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

References

[1] T. de Jong, S. Sotiriou, and D. Gillet, „Innovations in STEM education: the Go-Lab federation of online labs,“ *Smart Learning Environments,* vol. 1, pp. 1-16, 2014.

[2] J. Gago, J. Ziman, P. Caro, C. Constantinou, and G. Davies, „Europe needs more scientists,“ in *Brussels: European Community Conference Increasing Human Resources for Science and Technology*, 2004.

[3] M. Rocard, P. Csermely, D. Jorde, D. Lenzen, H. Walberg-Henriksson, and V. Hemmo, „Science education now: A renewed pedagogy for the future of Europe,“ 2007.

[4] M. Hannafin, J. Hill, S. Land, and E. Lee, „Student-Centered, Open Learning Environments: Research, Theory, and Practice,“ in *Handbook of Research on Educational Communications and Technology*, Springer New York, 2014, pp. 641-651.

[5] M. Pedaste, M. Mäeots, L. A. Siiman, and others, „Phases of inquiry-based learning: Definitions and the inquiry cycle,“ *Educational research review,* vol. 14, pp. 47-61, 2015.

[6] M. Lakkala, „How to design educational settings to promote collaborative inquiry: Pedagogical infrastructures for technology-enhanced progressive inquiry,“ 2010.

[7] J. Zhang, M. Scardamalia, R. Reeve, and R. Messina, „Designs for collective cognitive responsibility in knowledge-building communities,“ *JLS,* vol. 18(1), pp. 7-44, 2009.

[8] B. G. Wilson, Constructivist learning environments: Case studies in instructional design, Educational Technology, 1996.

[9] R. Ferguson, and S. B. Shum, „Social learning analytics: five approaches,“ in *Proceedings of LAK*, 2012.

[10] S. Manske, I.-A. Chounta, M. J. Rodriguez-Triana, and others, „Exploring Deviation in Inquiry Learning: Degrees of Freedom or Source of Problems?,“ *In: ICCE,* 2015.

[11] C. Rosé, Y.-C. Wang, Y. Cui, J. Arguello, K. Stegmann, A. Weinberger, and F. Fischer, „Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning,“ *IJCSCL,* vol. 3(3), pp. 237-271, 2008.

[12] D. Whitelock, D. Field, S. Pulman, J. T. Richardson, and N. Van Labeke, „OpenEssayist: An automated feedback system that supports university students as they write summative essays,“ in *The 1st International Conference on Open Learning: Role, Challenges and Aspirations.*, 2013.

[13] D. Whitelock, A. Twiner, J. T. Richardson, D. Field, and S. Pulman, „OpenEssayist: a supply and demand learning analytics tool for drafting academic essays,“ in *Proceedings of LAK*, 2015.

[14] D. Simsek, S. B. Shum, A. De Liddo, R. Ferguson, and Sándor, „Visual Analytics of Academic Writing,“ in *Proceedings of LAK*, 2014.

[15] I. Quinto, S. Buckingham Shum, A. De Liddo, and L. Iandoli, „A debate dashboard to enhance on-line knowledge sharing,“ *VINE The Journal of information and KMS.,* vol. 42, p. 67–93, 2010.

[16] A. F. Wise, „Designing pedagogical interventions to support student use of learning analytics,“ in *Proceedings of LAK*, 2014.

[17] A. Harrer, and T. Göhnert, „Integrated representations and small data: towards contextualized and embedded analytics tools for learners,“ in *Proceedings of LAK*, 2015.

[18] S. Knight, S. Buckingham Shum, and K. Littleton, „Epistemology, pedagogy, assessment and learning analytics,“ in *LAK*, 2013.

[19] L. Mason, N. Ariasi, and A. Boldrin, „Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning,“ *Learning and Instruction,* vol. 21(1), pp. 137-151, 2011.

[20] L. Mason, A. Boldrin, and N. Ariasi, „Epistemic metacognition in context: evaluating and learning online information,“ *Metacognition and Learning,* vol. 5(1), pp. 67-90, 2010.

[21] G. Qiu, R. Laurini, C. Leung, and others, Advances in Visual Information Systems: VISUAL, vol. 4781, Springer, 2007.

[22] A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and others, „Getting our head in the clouds: toward evaluation studies of tagclouds,“ in *Proceedings of SIGCHI*, 2007.

[23] A. M. O'donnell, D. F. Dansereau, R. H. Hall, and others, „Knowledge maps as scaffolds for cognitive processing,“ *Educational Psychology Review,* vol. 14(1), pp. 71-86, 2002.

[24] S. P. Leeman-Munk, E. N. Wiebe, and J. C. Lester, „Assessing Elementary Students' Science Competency with Text Analytics,“ in *Proceedings of LAK*, 2014.

[25] S. Wheeler, P. Yeomans, and D. Wheeler, „The good, the bad and the wiki: Evaluating student-generated content for collaborative learning,“ *BJET,* vol. 39(6), pp. 987-995, 2008.

[26] P. N. Mendes, M. Jakob, A. García-Silva, and others, „DBpedia spotlight: shedding light on the web of documents,“ in *Proceedings of SEMANTiCS*, 2011.

[27] J. Turian, „Using AlchemyAPI for Enterprise-Grade Text Analysis,“ 2013.

[28] M. Wittrock, C. Clark, and P. Peterson, Vol. 3: Students' thought processes, New York [etc.]: MacMillan [etc.], 1990.

[29] V. Emin-Martinez, C. Hansen, R. Triana, and others, „Towards teacher-led design inquiry of learning,“ *eLearning Papers,* 36, 2014.

[30] W. Clark, R. Luckin, and C. Jewitt, „Deliverable D5. 1 Methods and Specifications for TISL Components V1,“ *NEXT-TELL Consortium, EU,* 2011.