Detlef Krömker und Ulrik Schroeder (Hrsg.): Die 16. E-Learning Fachtagung Informatik,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018

Restructuring Linear Discussions in Mind Maps by Crowdsourcing

Julius Elias[1], Felix Dietze[1], René Röpke[2] and Ulrik Schroeder[2]

Abstract: With the rapid growth of the Internet, mobile computing and e-learning, more and more communication is happening online. With a large number of participants in e.g. MOOCs, discussions in linear structures can quickly become complex. Drawbacks of linear discussions motivate our approach to structure such discussions in mind maps. To control the restructuring process, we have created a series of user tasks that can be executed with the help of crowdsourcing. A two-stage study supported the conception and evaluation of suitable tasks. In addition, the study assesses the tasksô potential for crowdsourcing and whether users need to know the entire discussion to process the tasks. The results indicate that all tasks are suitable for a crowdsourced approach.

Keywords: Discussions, Restructuring, Mind Maps, e-learning, Crowdsourcing

1 Introduction

Today, online communication is growing rapidly and paper-based communication is replaced by email or chat software. Infrequent personal long-distance communication is nowadays achieved using instant messengers and social media. In the educational context, discussions are held in learning management systems or MOOCs. With a shift to distance and online learning, communication is taking place all over the world and the number of participants is steadily increasing. With a growing number of participants, todayôs communication and online discussions show various limitations.

Since most communication is persistently stored, new questions arise about information retrieval in linear communication streams. With the recent trend in remote work, many project teams communicate exclusively via online chat rooms or email. Due to the complexity and length of online discussions, information is difficult to retrieve from the conversation history.

Current learning systems such as MOOC platforms provide forums or chats where users can ask questions and discuss course content. This works well for small user groups. However, given the number of users in MOOCs, these communication channels quickly get overloaded. The same problem exists in self-hosted online discussion forums and e-learning platforms across different fields of study.

[1] RWTH Aachen University, Ahornstr. 55, 52074 Aachen, {vorname.nachname}@rwth-aachen.de
[2] RWTH Aachen University, Learning Technologies Research Group, Ahornstr. 55, 52074 Aachen, {roepke|schroeder}@cs.rwth-aachen.de

The lack of scalability in the number of users is a general problem of online communication systems. It is currently not possible, e.g. to have 10,000 people discuss a topic at the same time while keeping an overview of the entire topic. As the first part of a two-step study, we confirm that there is much more content than any participant could read, and the retrieval of information in large conversations is challenging.

To address the problems of linear discussions, popular Internet platforms such as Reddit, Hacker News, Facebook, and multiple blog comment systems offer a tree topology of comments, so that each comment responds to exactly one other comment. This adds more structure to discussions, but still does not provide an overview in very large ones. As an individual, it seems impossible to keep an overview.

This paper describes a two-stage study on the structuring of linear discussions in a mind map with a series of tasks using crowdsourcing. In the first part of our study, we confirmed that in discussions with many participants, it becomes more difficult to keep an overview because users have problems finding information and understanding relationships between contributions. To overcome these problems, we suggest encoding the discussion in a graph topology which users understand as mind maps. The conducted study follows a post-mortem approach which is independent of the structuring capabilities of the original platform. We have developed a series of basic tasks to transform chronological message histories into the mentioned graph topology. In the second step of our study we assessed the comprehensibility, the effort and the meaning of these tasks. Our aim was to determine whether it is necessary to know the entire discussion in order to structure it with the given tasks. If this knowledge is not crucial, the tasks are suitable for crowdsourcing.

The remaining paper is structured as follows: Section 2 presents related work on mind maps, argument maps, information retrieval from graphs, virtual teams, e-democracy and crowdsourcing. In section 3, we briefly present the concept of restructuring linear discussions in mind maps. Section 4 describes results of the conducted study. In Section 5 we conclude this contribution and introduce future work interests.

2 Related Work

Motivated by discussions with a growing number of participants in e-learning contexts, we refer to work in other disciplines such as information retrieval, e-democracy and virtual teams that identify similar problems.

When reading continuous text, the argument structure needs to be inferred linearly through the text. Faridani et al. [Fa10] describe that comment lists do not scale and reinforce extreme opinions. They present a user interface called Opinion Space which visualizes comments based on different ratings and compare it to a list and grid interface. They confirm that users preferred the presented grid and space interface with a list interface to navigate.

Various studies have shown the benefits of working with argument maps, as the critical thinking ability of students increases significantly [Ku14] and as well as their recall of arguments [Da11, Dw10, Sh14]. The idea of structuring arguments for analysis and transparency is rather old, e.g., the model of Toulmin [To03] for argument analysis or IBIS [KR70] for tackling wicked problems. More modern implementations of argument mapping software exist such as DebateGraph, which was actively used by The Independent newspaper and the White House. Cosley et al. [Co06] discovered that oversight increased both the quantity and quality of contributions while reducing antisocial behavior, another benefit of argument maps.

Davies [Da11] argues, that argument mapping leads to higher information retention, while teaching can be improved by using mind maps [Ha08]. Students report that concept mapping helped them understand, clarify accounting concepts and enhanced their interest in accounting [Ch08].

Fu et al. compare the usability of indented tree and graph visualizations of ontologies. They find that tree visualization is more approachable and familiar for novice users. Other subjects reported the graph visualization to be more tractable and intuitive, because of less visual redundancy, especially for ontologies with multiple inheritance [Fu13]. Additionally, Fu et al. study the usability with eye tracking and find that indented lists are more efficient at supporting information searches while graphs are more efficient at supporting information processing [Fu17]. In previous work, we explored different discussion topologies based on their expressiveness and positively evaluated the usability of a hypergraph-based model [Di17].

Services like Amazon Mechanical Turk allows one to pose simple tasks to humans and reward them with money. With such a service, it is possible to automate solving of problems where humans currently perform much better than computers, like image description, optical character recognition and iterative text improvement [Vo08, Be15, Li10].

The scalability of large deliberation discussions is an open problem of e-democracy. Hilbert points out that democracy still has a tradeoff between group size and depth of argument [Hi09]. The same insight comes from virtual teams. It is found that the most effective teams have sizes of less than 10 people [Fe14]. Kriplean confirmed this problem in political discourse and implemented a system in which online discussion participants can write a summary of what they understood next to comments [Kr11].

Overall, classical discussion formats and argument mapping has been studied by various authors. The increasing number of participants has a crucial influence on the complexity of a discussion and the linear structure makes it difficult to obtain information and gain an overview.

3 Concept

We hypothesize that most online communication problems originate from (1) their linear structure and (2) a high number of contributions. Firstly, there are no structural relationships except for manual quoting in internet forums. The lack of relationships leads to entangled discussion threads in a single chronological history. The entanglement was addressed by tree-like comment systems, which are mostly used for temporary conversations, e.g. comments on news articles. However, their structure cannot serve as a complete replacement for chat rooms. As a result, several chat systems are currently exploring thread features. Secondly, large numbers of participants produce many contributions. At a point, when a discussion becomes too large for a single participant to read every contribution, redundant content will be created. Old, but important messages get lost in the history and can only be found if users know exactly what they are looking for.

In the first step of our study, we aim to confirm that users experience the described issues in their everyday textual online communications. We refer to these issues as ñPain Pointsò to emphasize their negative effects on the communication.

Earlier research indicates that information can be better memorized, processed and structured when organized in mind maps or argument maps. Hence, the topology of information storage and visualization is essential. We propose to store discussions in a graph-based structure and visualize them similar to a mind map. A graph topology can encode the intended relationships of the participants and make redundancies visible.

The main idea of this work is to restructure a chronological textual online discussion by transforming its linear structure into a graph-based one, similar to a topic map [Pa02]. In the second step of our study, we investigate the transformation using simple tasks, which are either performed by participants of the discussion or crowd workers with limited context knowledge. These tasks express simple, atomic units of change, which should be easy to understand and execute, but lead to a complex and sophisticated restructuring process in their composition. As an example, a user is asked to tag a contribution or to decide whether two contributions contain the same information.

4 Case Study

In the first part of the two-stage study, we wanted to confirm our hypotheses about existing pain points in current online discussions and collect additional pain points we were not aware of. Based on the results, we have developed a series of tasks that address the limited structuring capabilities of current discussion platforms.

Restructuring with small tasks is a new approach to improve the overview of discussions. These tasks were designed to be simple questions so that the user does not have to worry about the underlying concept of the restructuring mechanism. Since there may be a considerable number of tasks that users need to perform until the discussion converges, a crowdsourcing approach seems appropriate. Therefore, we focused on finding out if it is sufficient to provide only a reduced context for the discussion, instead of forcing every user to read the whole conversation. The approach of providing limited contextual information should avoid the need for reading the entire conversation and facilitate the use of crowdsourcing.

4.1 Pain Points

This preliminary study is based on interviews with seven participants and aims to uncover and confirm problems that make it difficult to gain an overview of a discussion. The age range was between 19 and 32 years, five of the seven subjects were male and two female.

The interview was divided into two parts. In the first part, we interviewed the subjects on general online discussion issues. To set an anchor for further questions, each respondent was told to imagine a group of 30 people in an online chat. The questions ranged from general problems arising from the number of participants in the chat to the question of whether it is possible to find a particular message easily. In the second part, we presented the subject with an example discussion and asked more specific questions about it, based on the previous sensitization for discussion pain points.

Many subjects reported that the main problem in large discussions is the sheer amount of information. It may therefore be of interest to consistently create subthreads in order to organize the existing amount of information. Another hypothesis is that the overload of information results from a high level of information redundancy, since not every participant is able to read every contribution and therefore may produce redundant content.

The subjects often mentioned immature search options. In many cases, only exact phrases lead to proper results. This problem is partially resolved when more categories and tags are created.

Another problem in linear discussions is the understanding of the relationships between the contributions and between authors. The participants explained that it is challenging and exhausting to keep track of (1) related contributions and (2) the corresponding authors. They state that the relevance of a contribution is closely related to the author. Both points can be addressed by respective connections in a graph-based model.

4.2 Restructuring Tasks

Based on the results of the preliminary study, we constructed restructuring tasks to target mentioned pain points. The original task descriptions, as well as die example discussion were in German to avoid language barriers. The example discussion was taken from a real internet forum. As already mentioned, all subjects (age 20-28; 13 female and 28 male) were divided into two different groups. The first group consisted of 20 participants who had to read the entire discussion before they could perform the tasks. In addition, these participants were able to pause and look up information in the discussion. We will refer to this group as the Full Knowledge (FK) group. The second group, consisting of 21 participants, did not know the discussion in advance. These subjects only knew the title of the discussion and the context of the forum (a fitness forum). We refer to this group as the Partial Knowledge (PK) group.

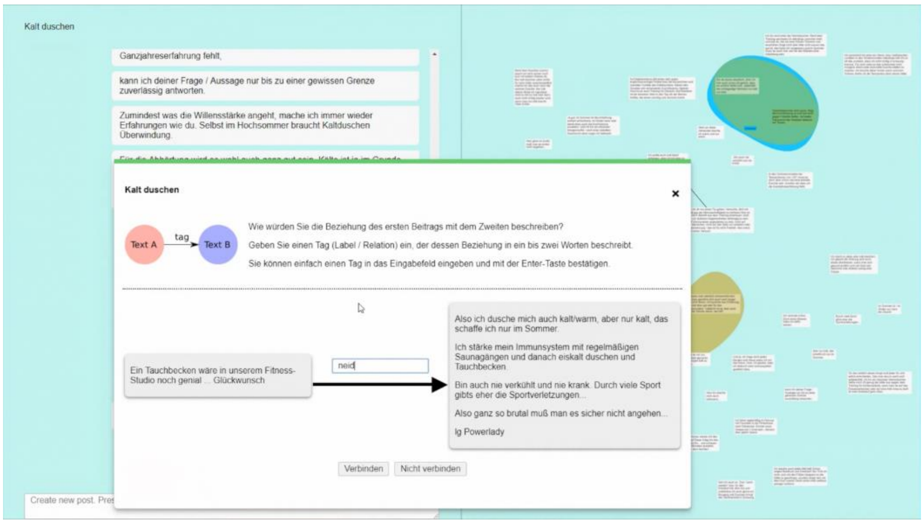

Fig. 1: Example of the AddTagConnection task. The task headline contains the discussion topic. On the left side, there is a symbol of the task. On the right side is a detailed task instruction. Below is the user interface to execute the task. In the background, the linear discussion and the mind map is visualized (FK group)

For the evaluation setup, we selected a subset of tasks in order to avoid overstraining the subjects:

1. AddTagConn(ection): Define a tag (label) that describes a relationship between two contributions (see Figure 1).
2. AddTagToPost: Define a tag (label) to categorize a contribution in the context of the discussion.
3. DeletePost: Decide whether the displayed contribution is relevant for the current discussion, so that irrelevant ones, such as spam or empty phrases, can be deleted.
4. MergePosts: Determine whether two contributions contain the same information. This serves as a tool to collect redundant content and reduce the number of contributions.

5. SubtopicPosts: Decide whether the first of the presented contributions initiates a new subthread to which the second belongs.

6. SplitPost: Divide a contribution into any number of units that contain separate statements. This can be used to split large contributions in order to refine the categorization of its parts.

Every task is repeated three times, resulting in a total of 18 tasks per subject. In addition, we have designed the tasks by selecting specific contributions that are suitable for the tasks. The tasks are presented in a fixed order. This order was created by a random permutation of the task list.

The study was conducted using a web-based application and the subjects were able to perform the tasks either during a video call using screen sharing or in person. We recorded the screen and logged the answers for each task. All subjects were encouraged to complete the tasks on their own and make reasonable decisions. We only gave explanations when parts were unclear.

After completing the tasks, we asked the subjects questions about each task and the overall concept itself. For each task, we presented a screenshot of the task with unrelated content to remind the subject of the task itself. The following questions were asked and the participant had to choose between the listed answers to get a four-dimensional result vector. Participants were also asked to explain their responses in order to gain more insights for improvement:

1. Did you understand the task description in the way it was intended?

2. Does the task make sense to enhance the overview?

3. In relation to the other tasks, how do you estimate the effort of the task?

4. Only in PK group: Did the displayed contributions in the task provide enough context to complete it?

The result vector includes the dimensions (1) comprehensibility, (2) sense of purpose, (3) effort and, only for the PK group, (4) sufficient context. The comprehensibility was explicitly assessed separately from the sense of purpose, in order to avoid that a wrong understanding leads to a misstatement about the sense of purpose. Accordingly, we explained each task in detail and asked the subjects whether they understood the task as explained or not. We asked the PK group explicitly whether the context is sufficient to complete the task to find out whether the task is suitable for crowdsourcing.

Afterwards, we presented two restructured versions of the discussion as mind maps. One was slightly restructured, the other in more depth. These graphs were structured manually by the authors until a sufficient overview was created. However, it would also be possible to create the graphs by repeatedly executing tasks.

5 Findings and Results

In the following, each dimension of the result vector is examined in more detail and compared in both (FK and PK) groups.

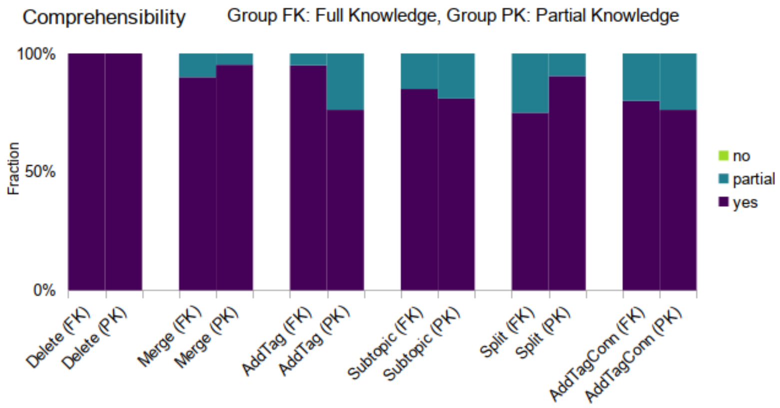

Fig. 2: Pairwise comparison of the comprehensibility of each task between the FK and PK groups

Overall, comprehensibility performs similar in both (FK and PK) groups. In all cases, at least 75 % of the subjects understood the task as intended and the remaining subjects did understand the task partially as intended (see Figure 2). Conspicuous is that subjects of the PK group stated that the AddTag task was not completely clear since they did not know which kind of tags or categorization is suitable. Concerning the Split task, the FK group stated that the contributions were mostly self-contained and because of that, they were not entirely sure how to execute the task.

In addition, the subjects were asked whether each of the tasks makes sense for restructuring discussions (see Figure 3). Here the FK group found that the AddTagConn task was not very helpful, as the discussion presented was simple and the task may be more useful in a more complex discussion. The AddTag task, which tags and categorizes a contribution, and the Subtopic task, which moves one contribution into another, were well received. Approximately 90 % stated that these tasks made sense and the remaining part thought that the task made at least some sense. In the FK group, each participant indicated that the Delete task makes sense. In the PK group, only 85 % reported this. One reason may be that it is difficult to decide whether a message contributes to the discussion without knowing the whole discussion. Merge seemed to make sense for only 55 % in the FK (55 % yes, 40 % partial, 5 % no) and 52 % in the PK (52 %, 43 %, 5 %) group, mainly due to the fact how the redundant contributions are processed. In the study, the Merge task created a new contribution by concatenating the contents of two contributions. Many subjects criticized this behavior because it did not solve the redundancy and suggested to keep only one contribution and delete the other. Another strategy could be to add an indicator that shows how many contributions have been merged into one contribution and e.g. display all contributions by clicking on this indicator. The split task made sense for 70 % of the subjects in the FK and 60 % in the PK group. Subjects of both groups explained that by splitting a contribution the context is lost, which can lead to difficulties in understanding the contribution.

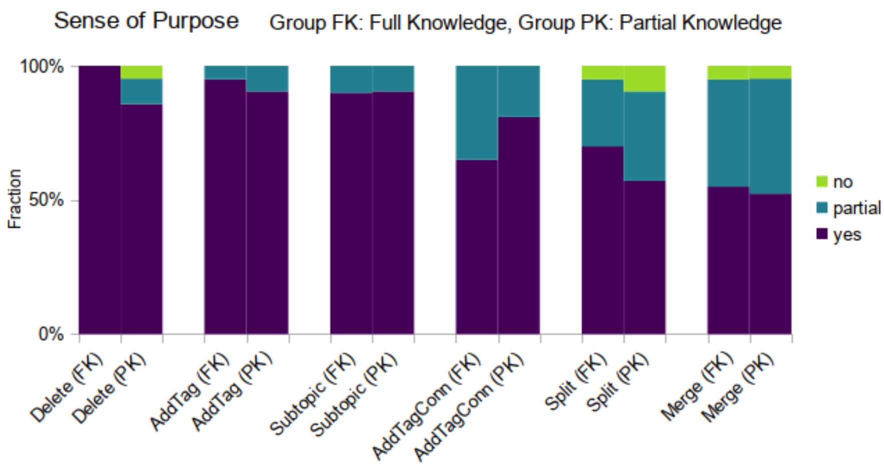

Fig. 3: Pairwise comparison whether a task is considered to make sense for each task between the FK and PK group

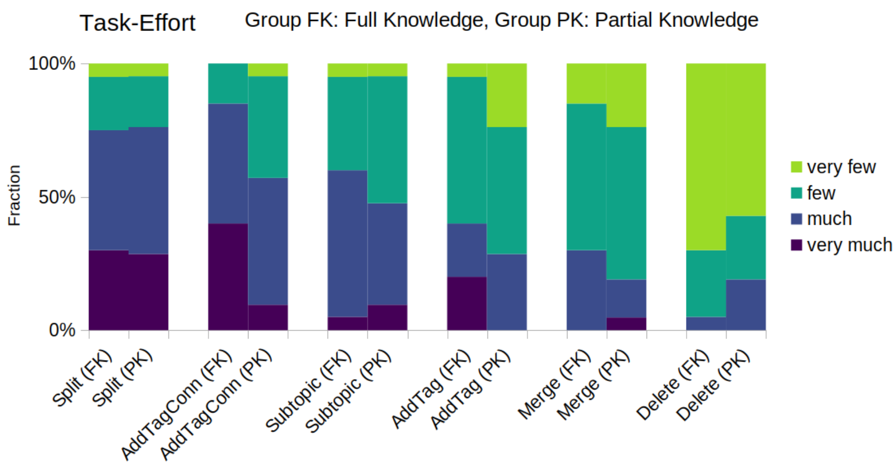

Fig. 4: Comparison of the estimated task effort for each group, sorted by total effort. The split task is considered as the most challenging task if both groups are considered equal, although the AddTagConn task requires more effort in the FK group

We also evaluated the estimated effort of each task (see Figure 4). Interestingly, in both tasks where the user had to add a tag, the subjects of the FK group tend to evaluate the tasks with more effort. This is because the subjects of the FK group put a lot of effort into the search for a more accurate tag and the subjects of the PK group only tried to find a roughly fitting tag. Surprisingly, both groups chose similar tags of the same quality and accuracy, which contradicts these statements. However, it can be different in discussions that are more complex. In addition, many participants pointed out that tag suggestions, e.g. from previous users or machine learning, could be useful. In contrast to these tasks, the participants in the FK group evaluated the Delete task with less effort than the participants in the PK group. This may result from a lack of knowledge about the whole discussion. The Split task was evaluated equally in both groups.

In summary, the effort tends to be higher in the FK group but is similarly distributed in both groups with the exception of the AddTagConn task. This task is evaluated with considerably more effort in the FK Group.

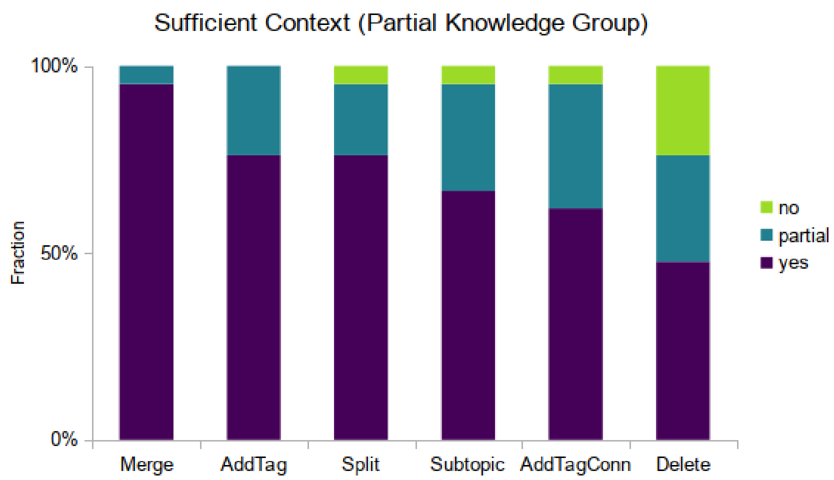

Fig. 5: Taskwise comparison in the PK group whether the displayed contributions in the tasks provide enough context to complete the task

As a final step, we asked the participants in the PK group whether the contributions presented in a task were sufficient to complete the task (see Figure 5). In this evaluation, the deletion task proved to be the worst (48 % yes, 29 % partial, 24 % no). As already mentioned, it was difficult to decide whether the contributions presented are relevant if the discussion itself is unknown. The tasks AddTagConnection (62 %, 33 %, 5 %) and Subtopic (67 %, 29 %, 5 %) turned out to be worse than the tasks AddTag (76 %, 24 %, 0) and Split (76 %, 19 %, 5 %), but still in a similar range. These tasks can be carried out without full knowledge of the discussion, but a better context knowledge can lead to decisions that are more accurate. The merge task seems to be executable independently of the level of knowledge (95 %, 5 %, 0).

6 Conclusion and Future Work

In our preliminary study, we confirmed that a large number of contributions lead to a loss of overview, which in turn evolves more quickly with more participants. Based on these results, we have developed simple, easy-to-understand and human-executable tasks to clean up and restructure arbitrary discussions. The tasks were evaluated in a simple discussion with two subject groups. One group read the entire discussion before the study, the other did not. The results indicate that all tasks are suitable for a crowdsourced approach. The knowledge gap between the two groups does not justify the omission of a certain task. We also found that many tasks can be improved to achieve a faster convergence to a useful mind map, e.g. by suggesting previously entered tags. Additionally, motivation to execute the tasks is an important factor to consider, be it intrinsically by active participants or extrinsically by crowd workers. After the first execution of a task, we noticed a speedup in subsequent executions of the same tasks. This suggests that the execution time will be much lower than observed in the study. Ideally, the discussion platform should allow to create a suitable structure while communicating.

In this study, only a limited set of tasks was evaluated. One can imagine many more useful tasks, like trimming, ordering or simplifying formulations in contributions. Depending on the medium (Forum or Chat), different tasks may be more appropriate. Future research should also investigate iterative task result evaluation, where users decide whether changes made by a task actually improved the discussion overview. Since users may perform differently in different tasks or content, one could implement intelligent task routing to decide which user is most suitable for a given task, similar to [Co06]. Additionally, strategies for reaching a consensus of applied changes need to be considered.

If restructuring discussions via crowdsourcing tasks proves to be useful in practice, task selection heuristics and task execution could be automated using machine learning and natural language processing. Even if automated task execution does not work as reliably as human execution, it can still enhance task productivity for humans.

References

[Be15] Bernstein, M.S. et al.: Soylent: a word processor with a crowd inside. Commun. ACM. 58, 8, 85ï 94 (2015).

[Ch08] Chiou, C.-C.: The effect of concept mapping on studentsô learning achievements and interests. Innov. Educ. Teach. Int. 45, 4, 375ï 387 (2008).

[Co06] Cosley, D. et al.: Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 1037ï 1046 (2006).

[Da11] Davies, M.: Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? High. Educ. 62, 3, 279ï 301 (2011).

[Di17] Dietze, F. et al.: Thatôs so Meta! Usability of a Hypergraph-Based Discussion Model. In: International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. pp. 248ï 258 (2017).

[Dw10] Dwyer, C.P. et al.: The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. Think. Ski. Creat. 5, 1, 16ï 22 (2010).

[Fa10] Faridani, S. et al.: Opinion space: a scalable tool for browsing online comments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1175ï 1184 (2010).

[Fe14] Ferrazzi, K.: Getting virtual teams right. Harv. Bus. Rev. 92, 12, 120ï 123 (2014).

[Fu13] Fu, B. et al.: Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation. In: International Semantic Web Conference. pp. 117ï 134 (2013).

[Fu17] Fu, B. et al.: Eye tracking the user experience--An evaluation of ontology visualization techniques. Semant. Web. 8, 1, 23ï 41 (2017).

[Ha08] Hay, D. et al.: Making learning visible: the role of concept mapping in higher education. Stud. High. Educ. 33, 3, 295ï 311 (2008).

[Hi09] Hilbert, M.: The maturing concept of e-democracy: From e-voting and online consultations to democratic value out of jumbled online chatter. J. Inf. Technol. Polit. 6, 2, 87ï 110 (2009).

[Kr11] Kriplean, T.: Encouraging reflective discussion on the web. University of Washington (2011).

[KR70] Kunz, W., Rittel, H.W.J.: Issues as elements of information systems. Institute of Urban and Regional Development, University of California Berkeley, California (1970).

[Ku14] Kunsch, D.W. et al.: The use of argument mapping to enhance critical thinking skills in business education. J. Educ. Bus. 89, 8, 403ï 410 (2014).

[Li10] Little, G. et al.: Turkit: human computation algorithms on mechanical turk. In: Proceedings of the 23nd annual ACM symposium on User interface software and technology. pp. 57ï 66 (2010).

[Pa02] Park, J. et al.: XML Topic Maps: creating and using topic maps for the Web. Addison-Wesley Longman Publishing Co., Inc. (2002).

[Sh14] Shum, S.B. et al.: DCLA meet CIDA: collective intelligence deliberation analytics. In: 2nd International Workshop on Discourse-Centric Learning Analytics, LAK14: 4th International Conference on Learning Analytics & Knowledge. (2014).

[To03] Toulmin, S.E.: The uses of argument. Cambridge university press (2003).

[Vo08] Von Ahn, L. et al.: recaptcha: Human-based character recognition via web security measures. Science. 321, 5895, 1465ï 1468 (2008).