

# Dialog-Based Online Argumentation: Findings from a Field Experiment

Tobias Krauthoff, Christian Meter, and Mauve Mauve

Department of Computer Science, University of Düsseldorf, Germany  
{krauthoff,meter,mauve}@cs.uni-duesseldorf.de

**Abstract.** In this paper we report on the results of a field experiment where more than 300 participants used dialog-based online argumentation. The participants were computer science students discussing how to improve the computer science course of studies. At the beginning of the argumentation the participants were informed that the results would be carefully considered by the computer science department in order to revise the course of studies. Thus this was a real-world experiment and not an artificial lab setting.

Over the course of two weeks the online argumentation received 255 user-submitted statements, leading to 235 arguments. After the argumentation was concluded we carefully analyzed the resulting content and asked the participants to answer a questionnaire. Our findings indicate that dialog-based online argumentation can result in a high-quality exchange of arguments without the need of anyone involved being an expert on formal argumentation. Furthermore we identified several areas where dialog-based online argumentation and our specific implementation could be improved significantly.

**Keywords:** dialog-based, argumentation, field experiment, large-scale discussion

## 1 Introduction

Dialog-based online-argumentation is an online argumentation scheme, where participants are guided through the arguments provided by other users, so that they perform a time shifted dialog with those that have participated before them. It does not require any prior knowledge or training from the users and avoids the shortcomings of forum-based systems, in particular balkanization and lack of scalability. Dialog-based online-argumentation is driven by a formal data structure capturing the full complexity of argumentation. The user interaction, however, has the structure of a regular dialog as it is performed in everyday life.

We have introduced the idea of dialog-based online-argumentation in [9]. In that paper we discussed the challenges and potential solutions required to build a dialog-based online-argumentation system and presented a first prototype, called *Dialog-Based Argumentation System* (D-BAS)<sup>1</sup>, which is available on *GitHub* as

---

<sup>1</sup> <https://dbas.cs.uni-duesseldorf.de/>

open source software<sup>2</sup>. Since then, we have improved and extended D-BAS into a fully fledged system for dialog-based online argumentation, so that we are now able to leave the lab and lab-experiments behind and instead deploy and evaluate D-BAS in real world settings.

In this paper we describe the findings from a real world use of dialog-based online argumentation, where all students of our computer science department were invited to propose and discuss improvements to the computer science studies program. In particular this includes an analysis of how the users participated in the discussion, an investigation of the user-based review system provided by D-BAS, information on the resulting arguments and their structure as well as information from a user survey. Furthermore we provide free access to the resulting argumentation data both in the native language of the argumentation (German) and an English translation. Both language versions are downloadable<sup>3</sup> as data sets for further study and are included in the live version of D-BAS, so that anyone interested can review the discussion in detail.

This paper is structured as follows. In Sec. 2 we give a brief overview of related work in the area of online argumentation. The general idea of dialog-based online argumentation and its implementation in D-BAS is summarized in Sec. 3. Section 4 describes the setting of the field experiment. Section 5 has a closer look at the peer-based review system and how it was used by the participants of the discussion. The quality of the resulting online-argumentation is investigated in Sec. 6. The results from a survey taken by the participants of the discussion is presented Sec. 7. We conclude the paper with a brief summary and an outlook to future work in Sec. 8.

## 2 Related Work

Tools for asynchronous online-discussion can be separated into forum-based approaches, pro and contra lists and tools for argument mapping. Although forum-based approaches received quite a lot of criticism in the past [7], it is, by far the most commonly used approach to support online argumentation in practice.

It has been suggested to use online pro and contra lists to aid collective decision making processes like *ConsiderIt* [10]. These lists work very well for evaluating a given proposal, but they are not suitable to deal with more general positions and alternatives since they do not support the exchange of arguments and counter arguments.

Online systems for argument mapping enable participants to structure their arguments and the relation between them in an argument map. While those systems do avoid the shortcomings of forum-based approaches, they require the users to become familiar with their notations and the semantics of formal argumentation. Examples are *Carneades* [4, 3], *Deliberatorium* [8] and *ArguNet* [11]. Therefore, in practice, they are used by skilled users, who are familiar with logic

---

<sup>2</sup> <https://github.com/hhucn/dbas>

<sup>3</sup> [https://dbas.cs.uni-duesseldorf.de/static/data/fieldtest\\_05\\_2017.tar.bz2](https://dbas.cs.uni-duesseldorf.de/static/data/fieldtest_05_2017.tar.bz2)

of argumentation rather than by average participants that want to take part in an online argumentation.

The idea of engaging in a formalized dialog to exchange arguments is used by dialog games, where participants follow a set of rules to react to each others statements [12]. In contrast to our work, dialog games look at the real-time interaction between users in order to learn something about a subject at hand. They do not seek to provide better instruments for online argumentation.

In addition to the main classes of ideas presented above, there is an individual system that is related to our work: *Arvina* [1]. *Arvina* allows a user to conduct a dialog between robots and humans. As a basis, it uses an existing discussion specified in a formal language [2] where the positions and arguments of some real-world persons are marked. A robot can use this information to argue with human participants. The participants can query the robots and each other. In contrast to the system we envision, *Arvina* is driven by the questions of the users. Thus there is no need for the users to react to replies from the system by providing their own arguments.

### 3 Dialog-Based Argumentation System

The goal of dialog-based online argumentation is to enable any user to participate efficiently in a large-scale online argumentation. At the same time it seeks to avoid, or at the very least reduce, the problems that occur in unstructured online argumentation such as a high level of redundancy, balkanization, and logical fallacies.

In the following, we briefly describe terms that will be used to explain the main aspects of dialog-based online argumentation. Based on these terms, we then introduce the main concepts of dialog-based online argumentation.

Each discussion is a set of *statements*, which are the most basic primitives used in an online discussion. The negation of a statement is itself a statement. Individual participants might consider a given statement to be true or false. A *position* is a prescriptive statement, i.e., a statement which recommends or demands that a certain action can be taken. Furthermore we need to distinguish between first-order and second-order *arguments*. A first-order argument consists out of a premise group — a set of at least one statement — and a conclusion, i.e. a statement. Both are connected by an inference, which is either supporting or attacking, so that the premise group is a reason for or against the conclusion. A second-order argument has the same kind of premise group, but the conclusion is the inference of an argument. With this we can argue about the validity of another reason-relation. Together, the arguments of a debate form a (partially connected) *web of reasons*.

The core idea of dialog-based online argumentation is a loop consisting of three steps: (1) presenting a single argument; (2) gather feedback from the user based on a list of alternatives and (3) the system selecting the next argument that is shown to the user based on the response and, possibly, the data gathered from the responses of other participants [9]. In this way the user and the system

perform a *dialog* where the system selects arguments that are likely to be of interest to the user and where the user provides feedback on those arguments.

The first thing that the system needs to do when a new user wants to participate in the online discussion is to choose an initial argument. This is challenging since the system has no information on the user, yet. One fairly straightforward solution is to simply ask the participant for an initial position she is interested in (see Fig. 1). After she has chosen or provided her position, she is asked to select or provide a statement explaining her choice (see Fig. 2 and Fig. 3). This statement is used as the premise, whereas the position forms the conclusion.

I want to talk about the position that ...

- the university should increase its capacities.
- the course standards should be increased.
- all lectures should be recorded and put online.
- Neither of the above, I have a different ideal!

What do you think about all lectures should be recorded and put online?

- I agree.
- I disagree.
- I have no opinion yet, show me an argument.

**Fig. 1.** Choosing an initial position. **Fig. 2.** Choosing attitude towards a position.

☆ You disagree with: all lectures should be recorded and put online.

What is your most important reason why all lectures should be recorded and put online is not a good idea? Because...

- in this case the copyright and exploitation rights are difficult to clarify and defend, especially if third-party material is used.
- the lecturers will otherwise have no or less audience in the lectures who can give feedback.
- thereby procrastination is encouraged during the acquisition of the lecture material.
- None of the above! Let me state my own reason!

**Fig. 3.** Selecting a premise for the initial argument.

Once a user is confronted with an argument (see Fig. 4), she can provide feedback on the argument. The options have to be usable by unskilled participants, but also have to be logically correct. We propose the following: (1) Reject the premise. (2) Accept the premise and, as a consequence, the conclusion. (3) Accept the premise but disagree that this leads to accepting the conclusion. (4) Accept the premise but state that there is a stronger argument that leads to rejecting the conclusion. (5) Do not care about the argument. Depending on the choice of the user, she can provide a statement supporting her feedback on the

presented argument. This may be taken from a list of existing statements (see Fig. 5) or she may enter a new one (see Fig. 6). While entering a new statement, the system scans for similar statements that have already been provided by other users and displays them in a ranked list. In this way it is easy to reuse existing statements while avoiding duplication of statements in the web of reasons. Any new statement added by the user will be inserted in the web of reasons.

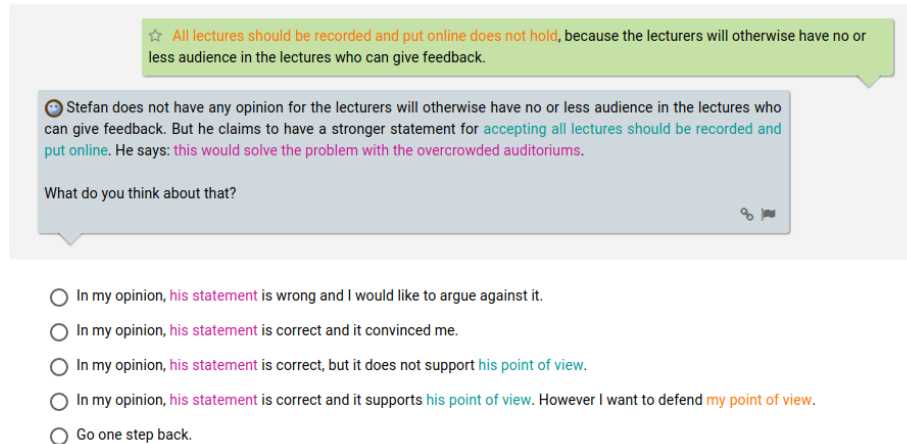


Fig. 4. Challenging the user’s argument and getting feedback from the user.

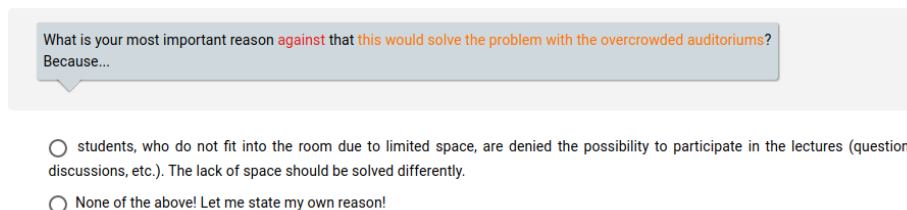
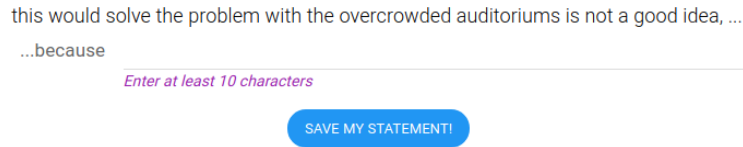


Fig. 5. Justification of the opinion in D-BAS.

## 4 Setting of the Field Experiment

The field experiment, we report about in this paper, took place at the computer science department of the Heinrich-Heine-University Düsseldorf. It targeted a topic that was relevant to the students of the department: how to deal with the increased number of students. The number of students has more than doubled in



**Fig. 6.** User interface for entering a new statement.

the past three years leading to numerous problems such as overcrowded lectures and a lack of places where students could sit down and study either in groups or by themselves. In order to avoid that participants are confronted with an “empty” system, we initialized D-BAS with two positions as well as two pro and two contra statements for each of those positions.

The students of the department were then invited via mail on behalf of the dean of the faculty of mathematics and natural sciences on May, 9<sup>th</sup> of 2017. Furthermore the teaching assistants of the department were invited, as well. The participants were asked to discuss how the course of study can be improved and how the problems caused by the large number of students can be reduced. The discussion was open until May, 28<sup>th</sup> of 2017. In total, there were 318 unique visitors and 47 users logged in to the system. Logging in is required to enter a new statement while conducting a dialog with the systems can be done anonymously. Out of the 47 users who logged in 11 were female and 36 were male. This roughly reflects the distribution of male and female students in the department. In total the participants added 22 positions and 255 statements (including the 22 positions). The resulting argumentation map is shown in Fig. 7<sup>4</sup>.

In order to allow others to analyze the discussion, it is available for download<sup>5</sup> as a dump of a *PostgreSQL* database and is licensed under the Creative Commons License *CC BY-NC-SA*<sup>6</sup>. The archive contains three versions: the original dataset of the discussion in German, a dataset which includes some corrections (those corrections are described in detail in Sec. 6) in German and a translation of the corrected dataset translated to English.

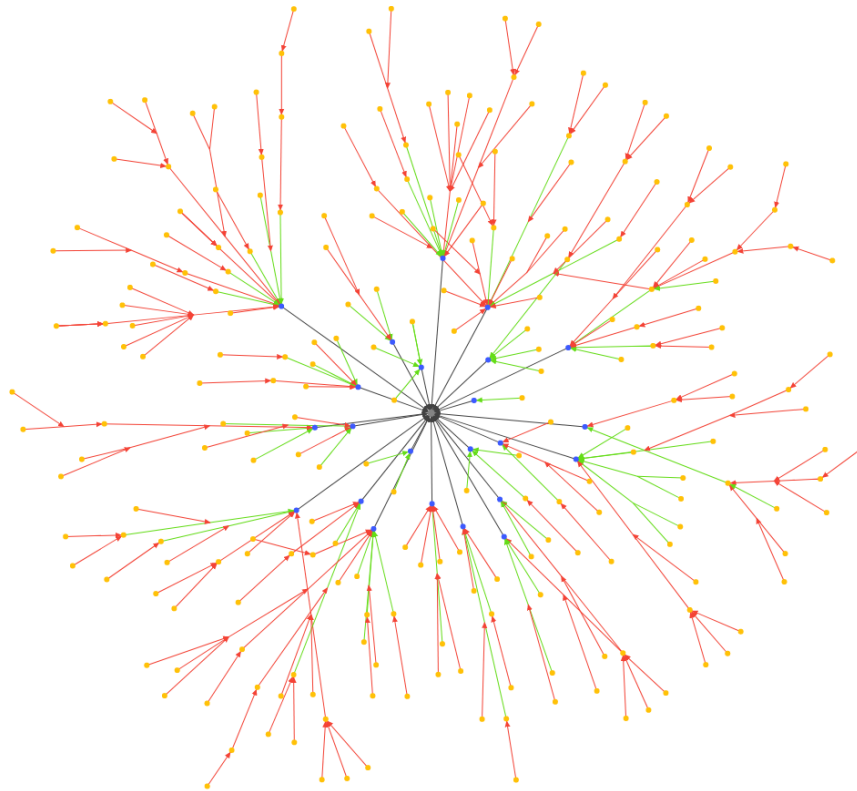
## 5 Decentralized Moderation

Dialog-based Online Argumentation relies on statements provided by the users in order to construct arguments that are then used in the dialog with other participants. In order to encourage users to provide well-formed statements, D-BAS provides a specific context when statements are entered, for example “Lectures should be recorded and released on a streaming platform because ...”.

<sup>4</sup> <https://dbas.cs.uni-duesseldorf.de/discuss/improve-the-course-of-computer-science-studies#graph>

<sup>5</sup> [https://dbas.cs.uni-duesseldorf.de/static/data/fieldtest\\_05\\_2017.tar.bz2](https://dbas.cs.uni-duesseldorf.de/static/data/fieldtest_05_2017.tar.bz2)

<sup>6</sup> <https://creativecommons.org/licenses/by-nc-sa/3.0/>



**Fig. 7.** Argumentation graph created by participants in D-BAS. The grey dot is the root of the discussion, the blue dots are positions and the yellow dots are statements that are not positions. Green arrows denote supporting arguments and red arrows denote attacking arguments

This will usually nudge the user towards entering a statement that completes the sentence in a meaningful way. Of course, this cannot completely prevent errors or malicious behaviour. It is therefore necessary to have a means for moderating the content provided by the users.

This could have been done by providing an interface where dedicated moderators would be able to alter or delete the statements provided by the regular users. If those moderators are skilled in argumentation and familiar with D-BAS, they could even make sure that statements are well formed for the use in D-BAS. We did not choose to take this approach. Instead we wanted to see if a decentralized moderation by the (untrained) participants themselves could work as well. This would be an important finding, since it would show that dialog-based online argumentation can take place and lead to a complex formal argumentation structure without anyone involved knowing anything about formal argumentation.

The decentralized moderation system implemented in D-BAS has been inspired by *Stack Overflow*<sup>7</sup> and works as follows. Every participant can flag content. She can either provide an improved version of the flagged content or simply report it as “The statement needs to be revised” or “This statement is off-topic or irrelevant” or “This statement is harmful or abusive” or “This statement is a duplicate”. Flagged content is not changed immediately. Instead it is entered into one out of several review queues, depending on how it was flagged. For example if a statement is flagged as harmful or abusive it is entered in the “Delete” review queue. Other users can go through those queues and either vote on the action to be taken or provide an alternative version of the flagged statement. Once a sufficiently clear-cut collective opinion has been reached, the appropriate action is taken, e.g. the statement might be replaced or deleted or the flagging might be discarded. The review queues maintained by D-BAS are as follows:

*Delete:* This queue contains statements, which have been flagged as off topic, irrelevant, harmful or abusive. If positive collective consensus is reached, this statement will be deleted.

*Edit:* This queue contains proposals where users have submitted and revised version of an existing statement. If positive collective consensus is reached, the old statement will be replaced by the new one.

*Duplicate:* It may happen that two separate statements are provided by users even though those statements have the same meaning. In this case it would make the argumentation more straight forward if those statements were merged. Those duplicate statements can be reported in the following way: one statement is marked as a basis and then another statement is selected as the duplicate. If positive collective consensus is reached, the duplicate will be deleted and the original statement will replace it.

*Optimization:* Finally, statements may be flagged because they need to be revised. Users going through the optimization queue can provide an alternative version of a statement from the optimization queue. This revision is then submitted to the edit queue for review.

In order to motivate users to participate by providing statements or by taking part in the review system, they gain reputation by helpful actions and in order to deter them from abusing the system, they lose reputation if their actions are considered unhelpful. The actions that a user can take in D-BAS, in particular which review queue he can use, depends on the reputation of the user.

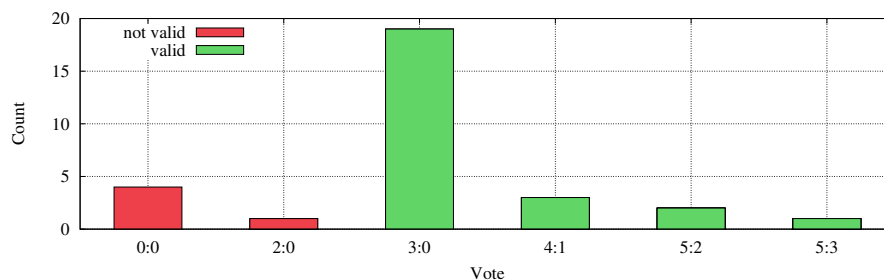
During the discussion at hand, 47 statements were flagged: no deletes, 25 edits, 5 duplicates and 17 requests for optimization. Figure 8 shows the results of the voting on the flagged statements. This excludes requests for optimization since those will not result in a vote but in an updated statement which is then

---

<sup>7</sup> <https://stackoverflow.com/review>



submitted to the edit queue. The vast majority of flagged statements is decided upon unanimously with three votes in favour of positive consensus. Only very few decisions required more than three votes to reach a decision, whereby the limit is five. The two instances marked in red were not decided upon at the end of the discussion, since they have not received a sufficient number of votes. This happened since they were flagged close to the end of the discussion.



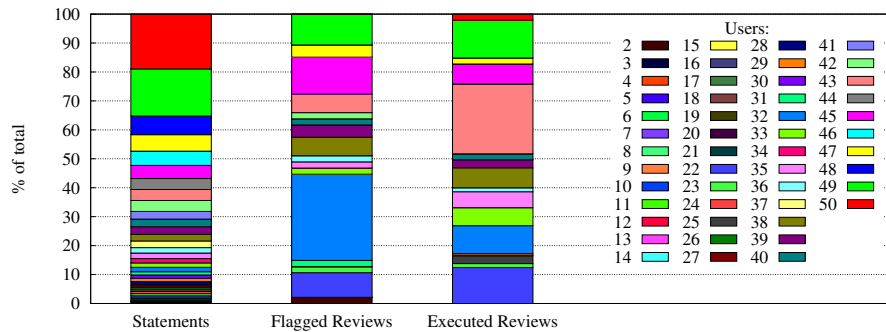
**Fig. 8.** Overview of voting in the D-BAS review system.

In the discussion, positive consensus was reached in every single case where any consensus was reached at all: all actions proposed by the user flagging the content were taken and all proposals for updating statements were accepted. We checked manually, if those decisions were plausible and found that this is, in fact, the case. All statements flagged as duplicates were true duplicates and every single edit corrected at least some mistake in the original statement. Also, there were no duplicates remaining that have not been flagged. However, some of the edits introduced new (mostly spelling) errors. This might also explain the non unanimous votes.

We were interested in how participation was distributed among the participants of the discussion in the review system. Figure 9 shows the share of each user for contributing statements, flagging statements and actions taking in the review system. It is quite obvious that for each type of action there are some power users. However, those are not the same across all action types. It seems that distinct users enjoy different aspects of contribution to the discussion.

Clearly, the discussion took place in a benign setting. A more controversial topic discussed by a less homogeneous group might stress the distributed review system to a significantly larger extent. However, what our findings clearly show, is that regular users will participate in the review system and that they are able to collectively improve the quality of individual statements and the overall discussion.

From observing the discussion we also learned, that there should be two more review queues. One for statements that should be split into several distinct statements. This would come in handy if an inexperienced user includes both premise and conclusion or multiple distinct premises in a single text contribution.



**Fig. 9.** Distribution of the users activity in D-BAS.

Another one for handling the opposite case, i.e., restoring a statement that has incorrectly been split into multiple parts. The specific observations that led us to those conclusions will be discussed in more detail in the following section.

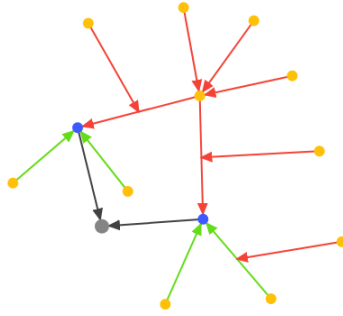
## 6 Quality of the Argumentation

One key question we wanted to answer with the field experiment was whether dialog-based online argumentation works and can, in fact, lead to a good online argumentation. Obviously, there is no simple metric that one could use to decide whether this is the case or not. However, it is possible to investigate individual characteristics of the argumentation that, taken together, provide a strong hint regarding its quality.

First, we take a look at the positions that were proposed by the participants. Positions are statements that can be executed. In this specific argumentation they represent ideas on how the computer science studies program can be improved. Altogether the participants added 22 positions to the argumentation. As mentioned above, additionally, two positions were provided by us at the start of the field test. All of the positions added by the participants are meaningful in the sense that they are actions that could potentially have an impact on the quality of the studies program. They all led to further reactions by other participants, indicating that they were of interest to others. Furthermore, there were no duplicate positions. This is an important prerequisite for scalability. While it is not possible to prove that no other means of online argumentation might lead to more or better positions, the absolute number indicates that the argumentation was extremely successful at gathering meaningful positions.

Next, we investigate how interactive the online argumentation was. The argumentation consists of 265 statements, including the 24 positions. In order to investigate interactivity, it is important to understand how the results of the argumentation look like. Essentially, each position is the start of a sub-graph of arguments. Since statements can be reused, the sub-graphs of the positions

are interconnected. From the perspective of the individual positions they overlap. An example for two overlapping subgraphs from the discussion is shown in Fig. 10<sup>8</sup>.



**Fig. 10.** Connected subgraph during a discussion.

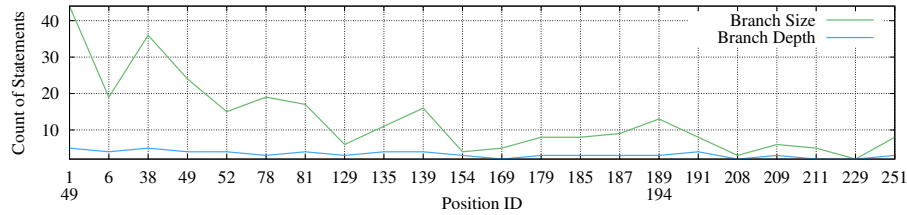
In order to determine the interactivity of the argumentation, we can now look at the number of statements that are directly or indirectly connected to each position. Furthermore we can investigate the maximum length of chains of arguments that are connected to each position.

Both the number of statements related to each position and the length of argument chains for each position are shown in Fig. 11. Most positions attracted more than ten arguments with the maximum at around 45 arguments for one position. Also, each position led to an average argument chain of length three or four. This clearly shows that this was a very interactive argumentation. Furthermore, the argumentation does not contain any (obvious) duplicate statements. Again, this is an important prerequisite for scalability. However, this is due to the review system and not an inherent attribute of dialog-based online argumentation: the participants themselves detected and removed five duplicated statements over the course of the argumentation using the review system.

One important aspect regarding the quality of an argumentation is whether the participants are able to react to arguments of others in an appropriate way. Given an argument consisting of a set of premises and a conclusion, D-BAS allows for the reactions described in Sec. 3 and shown in Fig. 4. Based on each participants history, recorded by Piwik<sup>9</sup>, we analyzed the selected feedback options. During the field test users have selected 200 undermines, 44 supports, 137 undercuts, 56 rebuts, 19 times they wanted to see another attacking argument and 104 times they just wanted to go back. We manually investigated, if those reactions were used appropriately, that is, if the resulting argument makes sense in relation to the argument it was a reaction to. This holds true for every single

<sup>8</sup> <https://dbas.cs.uni-duesseldorf.de/discuss/improve-the-course-of-computer-science-studies/attitude/454#graph>

<sup>9</sup> Piwik is an open-source analytics platform: <https://piwik.org/>.



**Fig. 11.** Size of sub-graphs and length of argument chains for each position.

reaction. This is surprising since at least the undercut is a challenging type of reaction. While we were very pleased with this result, it should be noted that the participants were all computer science students. It is not certain that this result would remain unchanged with a different set of participants.

So far all aspects of the argumentation indicate that dialog-based online argumentation and the D-BAS implementation indeed support high quality online-argumentations. However, as we will show next, there have also been some problems that we could observe. All of them are caused by the current D-BAS implementation and all of them can be avoided in the future by adapting the implementation accordingly.

During the experiment we had to intervene three times in order to split a single contribution of a user into several separate statements. In each of these cases we feared that not intervening would lead to follow-up problems when other users would try to react to the contribution of the user.

The first two cases occurred while the user was entering a position. Instead of just entering a position the user also provided a justification for the position. This problem happened, because the respective participant did not know that right after entering a position she would be asked for a justification for the position. This problem occurred only twice, because as soon as one had used D-BAS for a very brief time, it would become obvious that one should enter only the position at this time. In the future we will prevent this problem by merging the two steps of providing a position and its justification so that a user immediately realizes that she can provide the justification for the position in a separate entry field.

In the third case a user provided several separate premises in one contribution. This is a problem, because it would then not be possible for other participants to address each premiss individually. Again, after getting familiar with D-BAS, it would be obvious that one should provide only separate statements. Since we can not completely prevent this from happening, however, we will add an option to the review system that would allow other participants to break down a contribution like this into separate statements. Since this functionality was not present in the version of D-BAS we used in the field experiment, we manually split the contribution.

Additionally, we discovered that one feature of our user interface was misleading, if the user did not pay close attention: we assumed that the usage of the keyword “and” in a statement would often mean that the user tried to connect multiple statements that would better be represented as separate statements. Whenever a participant used “and”, D-BAS therefore explicitly asked if it should split the statement. If the user, at this point, did not choose the correct answer, a single statement that included “and” would be split in two meaningless fractions of a statement. While in the vast majority of cases where “and” was used, the participant choose the right option, there were six occurrences where they did not. We did not correct those issues while the discussion was under way, since they did not significantly hamper the discussion itself. However, in order to make the resulting data more accessible, we corrected them later on. For transparency reasons, we also kept the original data set.

In order to avoid this problem in the future, we will simply allow users to recombine those statements using the review system. This will solve this issue, since the problem is really obvious as soon as D-BAS splits the statements.

Summarizing, while there have been minor problems caused by the current version of D-BAS, the field experiment clearly shows that it is possible to lead a high quality and redundancy free online argumentation by using dialog-based online argumentation and its implementation, D-BAS. In particular, it demonstrates in a real-world setting that participants with no background in formal argumentation are able to collectively argue about a topic in such a way that the resulting formal argumentation map is correct and very comprehensive.

## 7 User Feedback

As a follow-up to the online discussion, we invited all participants to take part in a survey about D-BAS. As an online survey tool we used *Unipark*<sup>10</sup>.

Figure 12 shows the attitude of the participants towards key statements regarding D-BAS. For each line, the number of participants that answered the question is given. Clearly, the participants that have answered those questions do have a positive attitude towards D-BAS. In particular, they seem to like the general approach taken by D-BAS and they would use D-BAS again. It is also noteworthy, that for every single statement the average attitude is at or above neutral.

We were also interested in the attributes that users would associate with D-BAS. As a means to investigate this, we used bipolar word pairs. The result of this is shown in Fig. 13. Again, the results show that users participating in the survey assign quite positive attributes to D-BAS. However, they also indicate, that there are areas where it could be improved. In particular this holds true for the orientation that users have during an ongoing dialog (clear vs. confusing and unpredictable vs. predictable). We will address this in future versions of D-BAS by displaying a miniature version of (a part of) the argumentation graph during

---

<sup>10</sup> <http://www.unipark.com/en/>

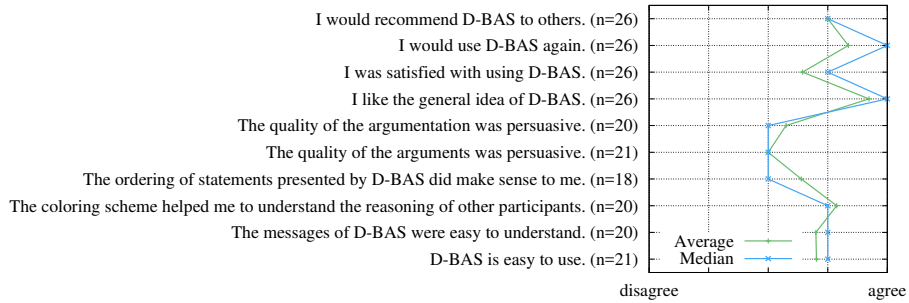


Fig. 12. Users evaluation of usability questions, based on *SUMI*[6].

the dialog. This should help the user to keep track of her position in the overall argumentation.

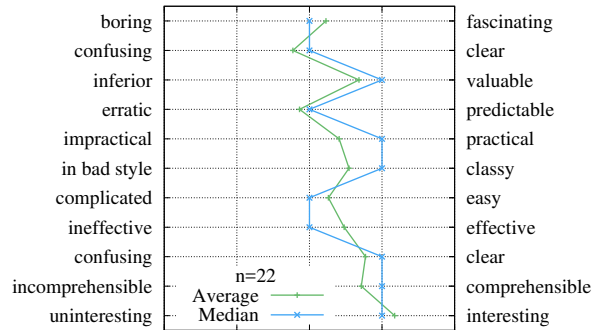


Fig. 13. Users evaluation of bipolar word pairs, based on *AttracDiff*[5].

## 8 Conclusion

In this paper we reported on the findings of a first field experiment using dialog-based online argumentation in a real world setting. The experiment confirmed, that this argumentation scheme is accessible by untrained participants and can result in a high-quality argumentation.

While the experiment provided us with a lot of information it is limited by the fact that this was only a single experiment with a very specific set of participants. In the future we will revise D-BAS according to the ideas presented here and make it available as a web-based service that anyone can use to host their

online argumentation. Our goal is to collect the data from a large number of argumentations so that we can then investigate dialog-based online argumentation on a much larger scale.

### Acknowledgements

This work was done in the context of the graduate school on online participation, funded by the ministry of innovation, science and research in North Rhine Westphalia, Germany.

### References

- [1] T. Bench-Capon, K. Atkinson, and A. Wyner. Using Argumentation to Structure E-Participation in Policy Making. *Transactions on Large-Scale Data-and Knowledge-Centered Systems XVIII*, 8980:1–29, 2015.
- [2] F. Bex, J. Lawrence, and C. Reed. Generalising argument dialogue with the Dialogue Game Execution Platform. In *Computational Models of Argument: Proceedings of COMMA*, pages 141–152, 2014.
- [3] T. F. Gordon. Carneades - tools for argument (re)construction, evaluation, mapping and interchange. <http://carneades.github.io/>, 2015. [Online, Last access 2017-06-27].
- [4] T. F. Gordon and D. Walton. The Carneades Argumentation Framework – Using Presumptions and Exceptions to Model Critical Questions. In *6th computational models of natural argument workshop (CMNA), European conference on artificial intelligence (ECAI), Italy*, volume 6, pages 5–13, 2006.
- [5] M. Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction*, 19(4):319–349, 2004.
- [6] J. Kirakowski and M. Corbett. Sumi: The software usability measurement inventory. *British journal of educational technology*, 24(3):210–212, 1993.
- [7] M. Klein. Using Metrics to Enable Large-Scale Deliberation. In *Collective intelligence in organizations: A workshop of the ACM Group 2010 Conference*, pages 103–233, 2010.
- [8] M. Klein and L. Iandoli. Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium, 2008.
- [9] T. Krauthoff, M. Baurmann, G. Betz, and M. Mauve. Dialog-Based Online Argumentation. *Proceedings of the 2016 conference on Computational Models of Argument (COMMA 2016)*, 2016.
- [10] T. Kriplean, J. Morgan, D. Freelon, A. Borning, and L. Bennett. Supporting Reflective Public Thought with ConsiderIt. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 265–274. ACM Press, 2012.
- [11] D. C. Schneider, C. Voigt, and G. Betz. Argunet – A software tool for collaborative argumentation analysis and research, 2006.
- [12] S. Wells. Supporting Argumentation Schemes in Argumentative Dialogue Games. *Studies in Logic, Grammar and Rhetoric*, 36(1):171–191, 2014.