Dialog-Based Online Argumentation

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Tobias Krauthoff

aus Wesel

Düsseldorf, Juni 2018

aus dem Institut für Informatik
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Jun.-Prof. Dr. Ing. Kalman Graffi

2. Univ.-Prof. Dr. Martin Mauve

Tag der mündlichen Prüfung: 31. Januar 2018

For my fiancée...

Abstract

Several decision processes – mainly at the local level – took place in Germany over the last years. These processes were executed according to the order of representative democracy, but the citizens were not satisfied with the outcome of these processes. Maybe the election of democratic representatives might not always be sufficient to gain support for political decisions by their citizens as soon as citizens want to participate actively on decisions, which affect them or their environment. With the rise of the Internet, people are now able to discuss online at any time and any place. Unfortunately, current tools for online discussions are not fit for large-scale online discussions and provide many common pitfalls.

The common problems of online discussions are very essential and can basically be separated into three main problems: (1) The structure of participants' contributions often lack structure and leads to an overload of information or even filter bubbles; (2) Current structured tools are not usable by unskilled participants, who are not familiar with the basics of argumentation theory as well as online argumentation; (3) Current software approaches do not scale with the amount of users and do not prevent polarization. The lack of structure enhance heated or ill-informed as well as polarized discussion. This is why we propose a dialog-based approach for large-scale online discussion as a novel approach.

Although many tools have been provided and suggested for large-scale online discussions, none of them have had a practical impact so far. Therefore it is part of this thesis to introduce and enable dialog-based online argumentation for any user with or without prior knowledge or training, to participate efficiently in a large-scale online argumentation and to develop software which avoids the problems mentioned above like the lack of structure and scalability issues.

In conclusion, we propose a software for the novel approach of dialog-based online argumentation, that is suitable for large-scale online discussions, avoids the common pitfalls and is usable for unskilled participants. The software aims to enter a dialog with new users by recycling already given arguments of other participants. It is possible to enter new statements during this dialog, which are connected to already given arguments and to submit feedback, which is correct in terms of argumentation theory, but wrapped in everyday language. These feedback options provide a helpful color scheme to be understood and used by unskilled participants. We have conducted a field experiment with the course of studies in computer science in May 2017, where 318 distinct visitors took part. 47 of them created an argumentation map with 235 arguments in total by conducting the dialogs. In addition, we have tested a decentralized moderation system, which worked well in this benign setting and invited every participant for a questionnaire about our prototype and the discussion. In summary, the outcome of the online discussions as well as the usability of our software are very persuasive.

Zusammenfassung

In den letzten Jahren fanden in Deutschland viele relevante politische und administrative Entscheidungen statt. Obwohl diese immer im Sinne der repräsentativen Demokratie durchgeführt wurden, zeigte sich, dass sich BürgerInnen gerne aktiv an Entscheidungen beteiligen möchten, die sie direkt betreffen. Mit dem Internet haben BürgerInnen eine Möglichkeit, schnell, einfach und unabhängig von Ort und Zeit miteinander in Kontakt zu treten. Allerdings sind die bisherigen Lösungen für Online-Diskussionen unzureichend und weisen allseits bekannte Schwachstellen auf.

Die grundlegenden Probleme im Bereich von Online-Diskussionen lassen sich in drei Teilbereiche gliedern: (1) Die Darstellung aktueller Beiträge ist nicht übersichtlich genug, was zu Überladung von Informationen oder Filterblasen führen kann; (2) Die Software kann nicht ohne Weiteres von ungeschulten BenutzerInnen genutzt werden, da nicht alle BenutzerInnen sich mit argumentationstheoretischen Ansätzen auskennt; (3) Die Software skaliert nicht anhand der Anzahl der BenutzerInnen und forciert Balkanisierung. Letzteres führt oft zu einem Mangel an Struktur, zu Fehlinformationen und stark polarisierenden Diskussionen. Obwohl in den letzten Jahren viele Ansätze für großflächig angelegte Diskussionen vorgeschlagen wurden, wird bisher keines in der Praxis verwendet. Daher ist es Teil dieser Dissertation, den neuartigen dialogbasierten Ansatz vorzustellen und damit ungeschulten BenutzerInnen großflächig angelegte Online-Diskussionen zu ermöglichen. Des Weiteren soll eine Software entwickelt werden, die die obigen Probleme umgeht beziehungsweise Lösungen dafür bietet.

Zusammenfassend stellen wir eine Software für den neuen dialogbasierten Ansatz vor. Dieser Ansatz umgeht die üblichen Fallstricke, ist ausgelegt für großflächig angelegte Online-Diskussionen und auch für ungeübte BenutzerInnen schnell und einfach zu benutzen. Dabei führt die Software einen Dialog mit der/dem BenutzerIn, wobei sie die bisherigen Argumente anderer NutzerInnen wiederverwertet und neuen BenutzerIn präsentiert. Während des Dialoges ist es jederzeit möglich, neue Aussagen anzugeben, die dann passend mit den bisherigen Argumenten verknüpft werden. Zwar basieren die von uns vorgeschlagenen Rückmeldungsmöglichkeiten auf Argumentationstheorie, sind aber alltagssprachlich verpackt und farblich codiert, sodass diese schnell und einfach genutzt werden können. Im Mai 2017 haben wir ein Feldexperiment mit dem gesamten Informatik-Studiengang durchgeführt, wobei 318 unterschiedlich Nutzende unsere Plattform besucht haben. Dabei haben 47 NutzerInnen durch die Dialoge eine Karte mit 235 Argumenten erstellt, wobei diese so treffend sind, dass sie in die zukünftige Gestaltung des Studiengangs einfließen. Zusätzlich haben wir ein dezentrales Moderationssystem getestet und alle TeilnehmerInnen zu einer Umfrage eingeladen. Abschließend können wir sagen, dass die Qualität der entstandenen Diskussion, als auch die Benutzerfreundlichkeit der Software sehr überzeugend sind.

Contents

Chapter 1.

Introduction

It is a fundamental need for people to participate on decisions which affect them or their environment. Over the last years, there were many cases where citizens demanded for an increased participation. Two notable decision processes took place in Germany: the construction of a new train station in Stuttgart (2011) – Stuttgart21 – and the (failed) school reform in Hamburg (2010). Next to these both processes, there are many more controversial decisions, mainly at the local level. All of these processes were executed according to the order of representative democracy, but the election of democratic representatives might not always be sufficient to gain endorsement for political decisions by their citizens. This suggests that citizens want to participate actively on the decisions of their political representatives and that they are searching for new alternatives to participate on these decisions.

With the rise of the Internet, we now have a basic infrastructure, which makes participation easier for everyone and independent of space and time. Online participation offers many possibilities for users as well as operators, it has an impact on political and administrative decisions. Additionally, it has the potential to strengthen the acceptance and the quality of them. Thereby we understand it as an activity among participants where they contribute voluntarily to discussions.

A central part of online discussion is the exchange of opinions among the participants. These exchanges can be very heated, if the topics or opinions are controversial. In addition to this, online discussions enjoy an increasing popularity amongst citizens, especially when they are related to political talks [Bimber, 2003, Chadwick, 2006]. Thus discussions are getting larger and new challenges for the structure of the participant's contributions arise. At the current state of the art, platforms for online discussions have a lack of structure, so that operators of these platforms lose the overview of the contributions, let alone the participants. This makes online participation processes even more difficult, because these lacks of structure enhances

heated or ill-informed as well as polarized discussion [Sunstein, 2009, Sunstein, 2002, Van Alstyne et al., 1996]. That is why we enhance the current technical solutions for online discussions and propose a novel approach for online discussions in the setting of online participation in this thesis.

1.1. Motivation and Problem Statement

Participation in online discussions about decisions which affect the participants can create a very informative space, but can also lead to an information and argument overload [Lampe et al., 2014]. Many discussions in the field of online participation are based on the rational exchange of arguments given by individual participants, but not only arguments are part of these online discussions, also positions, reasons and justifications are exchanged. Our aim is to support participants during these exchanges. In order to support participants, it has to be noted that maybe only a few of them are skilled to express their opinion or contribution in well-defined forms. Most of the time they do not know if they are expressing a personal belief or making a conclusion based on hard facts [Tannen, 2012]. Therefore we take a closer look at the field of online argumentation and we want to propose a novel approach for large-scale online argumentation to support these participants. In order to propose a novel approach for online argumentation, we have to focus on the main properties an online discussion tool should offer. These properties can be divided into three parts: 1) offer clarification of given contributions, because discussions can lead to an overload of information, 2) be usable by unskilled participants, because not every participant is familiar with the basics of argumentation theory or online argumentation and 3) scale with the amount of users and prevent polarization, because this lack of structure enhances heated or ill-informed as well as polarized discussion [Sunstein, 2009, Sunstein, 2002, Van Alstyne et al., 1996, Wyss and Beste, 2017]. In order to guarantee these properties, we will take a closer look at them and their related problems:

Clarity of contributions: When we look closely at current online discussion tools or written records of large scale discussions, we note a considerable disadvantage: they lack structure. To structure the arrangement of text contributions, the use of hierarchical lists can be considered, where the first contribution is the one with the earliest time stamp and new contributions are connected below. As a consequence, the first and last contributions are always privileged, because they are seen more often than contributions in the middle of the argumentation. In contrast to a list representation, argumentation maps seem to be more rational, but even harder for everyday use by unskilled participants. Therefore one aim is to develop a system, which connects the participants contributions regarding their logical consequences, not based on their

creation date, but without any needs of expert knowledge.

Unskilled participants: Many of the users of online participation processes are not familiar with argumentation theory, this means that they are unskilled in formulating and handling well-formed arguments or expressing their opinion in a rational and not emotional way. In order to satisfy both groups of users, systems can be separated in two different styles for argument representations: Contributions can be represented either textual or graphical. A textual representation has the advantage that there is less room for interpretation, but these systems are often hard to understand due to long text parts. However, graphical interfaces could be harder to interpret, because each element of a graphical interface has its own meaning. Hence we want to take the advantages of both parts for a novel interface. Furthermore participants have to give feedback towards given contributions and arguments. According to [Scheuer et al., 2010] feedback can be given at different times and with different graphical tools, but users (of current discussion tools) are accustomed to give feedback in natural language or rather within a text box. In summary, we have to develop an approach, where the users have a small list of different feedback options. These options have to base on argument theory, but are wrapped in natural language, so they can be used by unskilled participants with ease. The predefined feedback options have the advantage that the system can carefully react to the feedback, but we have to ensure that the range of useful and formally correct options is covered by our system.

Scalability and Balkanization: Depending on the design and the kind of navigation during discussions, systems often have to deal with the well-known problems of poor scalability and high level of polarization [Klein, 2010, Spada et al., 2014] as well as polarization [Sunstein, 2009, Sunstein, 2002, Van Alstyne et al., 1996]. It can be seen in practice that many tools do not scale with the number of involved participants and prevent well-formed argumentation, because they will get biased, irrelevant, emotionally heated and even ill-informed [Krauthoff et al., 2016]. That is why tools for online argumentation have to present arguments in a neutral way – without any judgmental comments – to prevent these problems. This means that the argumentation should be mainly argument-centric and not time- or topic-centric. To achieve scalability in the amount of arguments respectively text contributions of participants, we have to: (1) reduce the amount of statements with the same meaning and (2) propose a navigation through the discussion, so that participants will see every statement regarding their opinion while staying motivated and having fun during the argumentation. The first task will help to reduce duplicates, which will increase the signal-to-noise ratio [Klein and Iandoli, 2008]. The second task will help users for getting a better understanding of the discussion, which will help produce statements with a higher quality. In addition, this will decrease the amount of

duplicates, too, because users should have seen everything regarding their own point of view. Therefore a novel approach for online discussions and argumentation should offer mechanisms to reduce duplicates and to provide a clear navigation through the discussion, which implies a new form to bootstrap an ongoing discussion by new users.

Although the state of the art offers software with solutions for each of these problem statements, no current tool is able to deal with all three of them at the same time. Hence it is part of this thesis to enable online argumentation to any participant without previous knowledge or training, to participate efficiently in a large-scale online argumentation and to develop a software, which avoids the problems mentioned above.

1.2. Research Questions

In the following, we discuss the research questions which are derived from the problem statements described in Chapter 1.1. Our main goal is to develop a novel approach for online argumentation, which does not require any prior knowledge nor training in argumentation theory or related topics at all. In addition participants should have fun during online argumentations, so that they stay motivated even within complex topics. Further the already mentioned shortcomings have to be avoided while the approach has to allow complex argumentation based on unrestrained input. The main idea is to guide users through the arguments provided by other participants, whereby the arguments are connected in a *Web of Reasons* (WoR). In a WoR arguments are connected by their logical consequences, so that we can perform a time-shifted dialog between the current user and those, who have already participated, whereat the selection of the next argument is based on the user's previous actions. These actions are manifested in several feedback options, which are written in everyday language, but based on argumentation theory. The basic principle of this approach is shown in Figure 1. Based on the given feedback for a shown argument, the system can select a new argument, which is connected with the previous one and can present this new argument towards the user again. Thereby the system itself should be driven by a formal data structure capturing the full complexity of argumentation in the backend, whereby participants interactions take place in the frontend and have the structure of a regular dialog as it is performed in everyday life. It is the task of the system – and not of the participants – to translate between those two views. Based on this, we call our novel approach a *dialog-based approach*. In order to develop software for this approach, this thesis clarifies, which problems have to be solved and which questions have to be answered. Therefore we can constitute the following questions:

- Current software tools for online argumentation present complete sets of arguments as

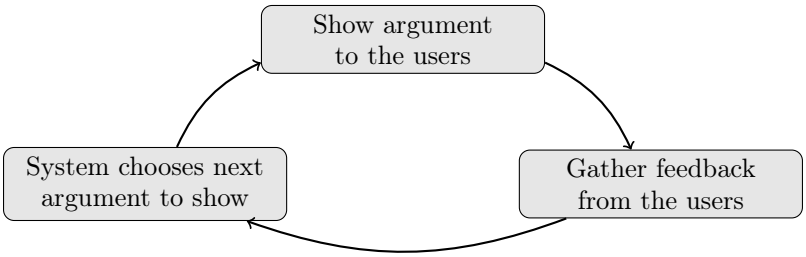


Figure 1.: Visual representation of a dialog-based approach.

tree-structured lists or maps. These representations contain too much information at once. In addition many of the presented arguments are not related with the participant's point of view and are tedious for the current user. Hence we need a way to display only relevant arguments the participant is interested in. So the question is: What is an effective and attractive way to present dialog-based online argumentation to make it work for skilled as well as unskilled participants?

- One basic part of (online) argumentation is to gather feedback from an interlocutor and the possibility for a system to understand given feedback in an unambiguous way. Therefore there are at least two emerging research questions when considering the foundation of dialog-based online argumentation: In which form should the systems feedback be displayed which is presented to the user? And how should the list of responses look like that the participants can choose from?

- In the beginning of every argumentation, the user needs a starting point, e.g. some initial argument or claim to start with. This is challenging since the system has no information on the user yet and does not know what the user is interested in. Thus, how do we bootstrap an ongoing discussion with many arguments or even more fundamental: how do we initialize an empty discussion?

- When an user has joined the discussion and gave feedback for the first argument, the system has to select the next argument that should be presented. This selection can be based on many sources of information, which can be collected during the previous argumentations of the current user or learned by the behavior of previous participants, like the history of actions that someone has performed. But what is a first approach to navigate through our WoR, so that the user does not have the feeling that she only sees some random arguments instead of a conclusive discussion consisting of arguments?

- An online argumentation can only grow, if the system accepts new arguments given by participants. Thereby it must be observed that the participants are not familiar with

argumentation theory and most of them are not able to define a well-formed argument for the WoR. Hence we need mechanisms to accept new input during the argumentation. In addition, the new input has to be compared to existing statements in the WoR, so that duplicates are avoided. In order for users to be able to form well-defined arguments, we need a variation of sentence-opener interfaces, which do not explicitly expose categories but which scaffold new contributions through predefined sentence-starting phrases. Typically, these interfaces are based on an underlying model of desired communication acts and processes, for instance, dialog games [Mcalister et al., 2004]. So, how do we accept new statements as well as arguments and connect them with our existing WoR?

- Transparency is highly important for discussions based on a WoR. None of the participants should get the feeling that their argumentation is guided for another one's interest. Therefore collected information about users as well as the arguments have to be accessible for all. How can we display the WoR in an intuitive way and how can we present the information based on the gathered feedback of the participants?

- Finally it should be possible to revise every statement and argument of the argumentation. For this, we need an approach to review given links and statements with respect to their spelling or their meaning. The execution of reviews should be possible without any interventions of moderators or operators, so the community should be able to administrate the content on their own, but the use of a moderator should be always possible. Then the question is: How can the community review data on their WoR?

In general, we provide a novel software approach called *Dialog-Based Argumentation System* (D-BAS), which provides answers to the questions mentioned above. In addition it is very important that participants have fun using the novel approach. Therefore we present a first glimpse into D-BAS in Chapter 2 to provide answers for all except the last two questions. In the following chapter we propose an extended version with improvements for the first two questions and we are pleased to answer the last questions in Chapter 4. Finally, in Chapter 5, we present a way to embed D-BAS in arbitrary blogs or newspaper websites.

1.3. Contributions

We released the achievements in this dissertation in four publications which had the purpose to implement, test, evaluate and spread a novel approach for large-scale online discussions with a dialog-based approach as well as a new way to embed our approach into arbitrary websites. Indeed, we first implemented the software tool and tested it at our department. Secondly, we

focused on improvements for the simplicity and visual acceptability of the software. Afterwards we performed a field experiment with our course of studies in computer science. And finally, we focused on a way to embed novel approaches for structured discussion in arbitrary websites like newspapers and/or blogs.

1.3.1. Dialog-Based Online Argumentation

In Chapter 2 the findings on the novel approach for dialog-based online discussions are summarized. In the publication [Krauthoff et al., 2016] we address a dialog-based and argument-centric way for online-discussions with a large number of participants and provide first solutions for three problems. First, we introduce a way to navigate the WoR, whereby we bootstrap the dialog by simply asking the participant for an interesting position she wants to talk about. Based on the first decision, the system selects an attacking argument. Second, the participant has to provide feedback on this attacking argument, whereby the feedback options have to be based on argumentation theory. For an easier use of these options, they are wrapped in everyday language. Third, the WoR is created by the participants themselves during the discussion, not by experts based on a transcript. Therefore participants have the ability to enter new premises for their opinions in every step of the discussion, where new statements will be connected to the existing WoR.

After the publication and presentation of our position paper [Krauthoff et al., 2016], we added a description of the theoretical foundation as well as details of the implementation of D-BAS and its WoR in an extended version. The publication [Krauthoff et al., 2017a] is described in Chapter 3. The main advantage of our WoR is the practical orientation during the development. This helped us to develop a data structure separated from any argumentation theory, so that D-BAS' data structure can be interpreted as either deductive argumentation [Betz, 2010], probabilistic reasoning [Skyrms, 2000, Hahn and Hornikx, 2015, Verheij, 2014] or defeasible reasoning [Dung, 1995a, Pollock, 1987].

1.3.2. Dialog-Based Online Argumentation: Findings from a Field Experiment

In Chapter 4 the findings on the first field experiment at the course of study in computer science at the University of Düsseldorf from the publication [Krauthoff et al., 2017b] are summarized. For this paper we invited all computer science students to take part in our field experiment with D-BAS to discuss how to deal with the increased number of students in our/their course

of study. Although these students are very tech-savvy, they are not skilled in argumentation theory and the dialog-based approach is new to them. During the two weeks of our experiment, more than 300 students joined our invitation, contributed 255 statements, which lead to an WoR with 235 arguments. Afterwards we invited all students to take part in our questionnaire about D-BAS and the discussion in D-BAS. The results showed that our approach works for this large-scale setting and the community was able to moderate the statements on their own.

1.3.3. discuss: Embedding Dialog-Based Discussions into Websites

The findings on embedding structured online discussion in arbitrary websites from the publication [Meter et al., 2017] are summarized in Chapter 5. In this paper we present a web application, which embeds structured online discussion tools, such as D-BAS, into arbitrary websites like newspapers and blogs. This embedding has the advantage that popular tools, which provide place for readers feedback, like simple comment sections in form of linear and topic-centric lists, can be replaced by structured applications. In the case of D-BAS the structured application has a common backend, so that the same discussion can be accessed from several sites to promote the divergence of audience, which leads to a higher quality and outcome of the discussion [Hoffman and Maier, 1961].

1.4. Related Work

In order to support people during online argumentations, many tools have been implemented and proposed as solutions over the last decades [Scheuer et al., 2010]. In part, it was also noted that not many people are skilled to argue in or rather with well-defined forms most of the time [Tannen, 2012]. So they do not know whether they express a personal opinion or make a conclusion based on hard facts. In order to resolve the problem, different graphical user interfaces for online discussions and argumentations were developed. According to [Scheuer et al., 2010] they can be divided into: (1) linear and (2) threaded, (3) graph-based as well as (4) container and (5) matrix representations. In the following we are going to introduce a possibility to compare these different styles as well as present the properties of each representation style.

Content	Usability	Effect / Impact
Completeness	Effectiveness	Understanding
Quality	Efficiency	Learning Progress
Connections	Satisfaction	Opinion Change
Replication	User Experience	Fun Factor
Relevance		
Well-Justified Premises		
Reference		

Table 1.: Metrics for a comparison of a discussion and its software.

1.4.1. Comparison of Representation Styles

Comparison of different representation styles, especially for online argumentation tools, is not an easy task. Next to the different styles, we also take a look on the discussions, which are clearly affected by their representations. Therefore we assume that a discussion and the representation style supporting it, can be measured in at least three dimensions. The first dimension is about the content of the discussion. It focuses on the (text) contributions of the participants. The second dimension deals with the usability of the software. We want to understand (and measure), what requirements must be met by software supporting online discussions in order to be usable. The third dimension is the effect and impact of the software/discussion on the participants. For example, in a good discussion the participant should be able to understand contributions by other participants, their motivations and so on. Only if given contributions are understood, a participant is able to give proper feedback. For each dimension, we suggest a list of different criteria, which are listed in Table 1. The complete list including specific metrics and experiment types can be seen in Appendix A. It can be seen that the analysis of the discussion's content is based on a framework for the analysis of deliberation by [Friess and Eilders, 2015]. Next to this, usability is a defined concept of the standard [ISO, 1997]. The effect and impact of a discussion and its style can be inspected by the subjective evaluation of the participants by questions like: "Did you understand every presented contribution of the other participants?".

We have to stress that the objective comparison of different styles and their effects and impacts on the resulting content of a discussion is a very complex proceeding and extremely interdisciplinary. Therefore we focus on the main aspects like the question "What is a good quality of style?" [Kitchenham and Pfleeger, 1996] in this dissertation. It can be stated that good quality is subjective and means different things to different people. One could compare different representations by their ease of use and satisfaction for a set of participants completing specific tasks in a specific environment. But to give a holistic picture, we think that this is only one point of the style comparison. In order that a comparison between the different representation

styles and their accompanying implementation can be given in a simple way, we compare them in a Kiviat diagram with seven categories. The diagram can be seen in Figure 2. Every category has a scale from 1 to 5. With a valuation of 1 (very weak), we assume that the current implementations have substantial difficulties to fulfill the requirements of the related category, whereas the valuation of 5 (very strong) reflects the best case that we can envisage until now. However, the values themselves indicate only an orientation value based on our convictions. Besides the category of this section – the ease of use – we classify in total online argumentation tools in the seven categories:

1. Online argumentation tools should be easy to *evaluate*. Ideally, the participant's text contributions are linked together, so that the system can recognize interferences between the contributions. One famous example for a huge amount of text contributions is Google's Project 10ˆ100, where Google asked as many people as possible to share ideas for changing the world by helping others [Twohill, 2008]. In this project, people from more than 170 countries submitted more than 150,000 ideas [Twohill, 2010]. Even if a huge amount of contributions may be sorted out with the help of natural language processing, there will be still an enormously amount of posts, which has to be manually analyzed. Therefore argumentation tools should have a high degree of evaluability.

2. The previous item illustrates that the amount of contributions increases significantly with the amount of involved participants. In [Iandoli et al., 2017] 640 participants already created enough content that the post-processing took about 160 men-hours of work. Thereby we think that a high degree of *clearness* is very helpful for argumentation tools to simplify analyses as well as the participants point of view, because a more clear tool will help participants to contribute arguments with a higher quality and analysts can inspect the argumentation easier.

3. Talking about online discussion leads us to online argumentation as well as argumentation theories, which will inevitably lead us to argumentation frameworks like [Stephen, 1958, Dung, 1995a, Modgil and Prakken, 2014]. Unfortunately the average user is not skilled in argumentation theory, therefore online argumentation platform should be *unrestrictive* in the frontend, so the usage of the frontend must not require any skill in argumentation theory. On the other hand the backend can be highly structured, if the frontend is able to map the data without conflicts.

4. Starting with [Twohill, 2008], we could ask, if every post was unique or have there been any duplicates? Often several participants have the same idea, but either the system has no algorithm for detecting duplicates or the list of contributions is just too long to read, so participants contribute their idea anyway. However, in the best case a system should

be able to reduce or even eliminate the redundancy of the contributions. Hence it should *scale* in the amount of user's contributions.

5. Even if a novel approach has high rankings in the four categories listed so far, the operators have no guarantee that their approach will be used in practical environments. Hence they may have to keep the special requirements of suitability for daily use in mind, so that their software is made for the *practical use* (at the local level), not at universities or by experts only.

6. Any system or device designed for the use by people should be easy to use, easy to learn, easy to remember, and helpful for users. Then, according to [Gould and Lewis, 1985], it has a high degree of *usability*. Although the field of usability and their tests is very broad, end users often have a very good gut feeling, when they would call a software efficient to use, easy to learn and satisfying during the use. Therefore software has to have a good *usability* to be accepted by participants.

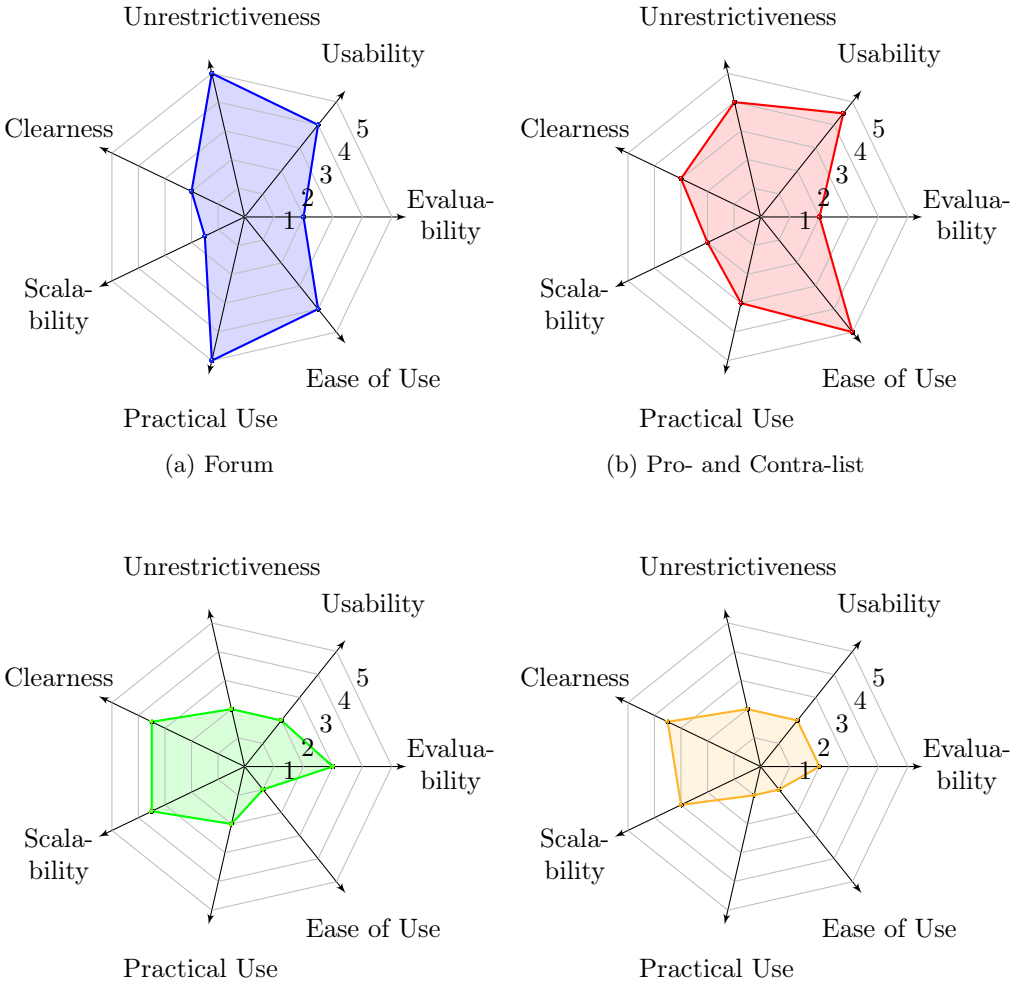
7. Even if the implementation of a specific style provides a very good usability, the ease of use may be unusual, strange or it needs getting used to. Hence the *ease of use* should be user-oriented and very high. According to [Scheuer et al., 2010] many software tools have a broad range of functionality according to online argumentation, but a low ease of use. Usually these tools are for experts only, who are trained or skilled enough, but tools for online argumentation and participation are useless, if they cannot be used by every kind of adult. Therefore we want to develop software for unskilled participants with a high degree of ease of use.

Based on the possibility to compare representation styles in seven different categories, let us now take a closer look on these styles as well as the related implementations, which can be found in everyday use:

- In Chapter 1.4.2 we will take a closer look on the most primitive form for online argumentation: Linear and threaded approaches in the form of chats and forums.

- Graph-based approaches are way more complex than forums and Chapter 1.4.3 is dedicated to these tools.

- Chapter 1.4.4 deals with matrix and container approaches, which are way more exotic than the first two are.



(a) Forum

(b) Pro- and Contra-list

(c) Argumentation Map

(d) Matrix- and container approach

Scale: 1 ≡ very weak, 2 ≡ weak, 3 ≡ average, 4 ≡ strong, 5 ≡ very strong

Figure 2.: Kiviat diagrams for different properties of online argumentation tools.

1.4.2. Linear and Threaded Approaches

The most primitive form to represent online argumentations is a *linear* list of text contributions, whereby each contribution can be marked by time stamp and author's name. This style is very easy to use and simulates a very common way of discussions between two or three people, but with increasing count of participants, the degree of incoherence significantly increases, too [Mcalister et al., 2004]. The most well known tool of linear representation is the *Internet Relay*

Chat (IRC), based on [Oikarinen and Reed, 1993]. Nevertheless the IRC and its representation were never intended to support large-scale discussions or even argumentations. As can be seen in the illustration of this style in Figure 3, it is not sufficient for large-scale argumentation.

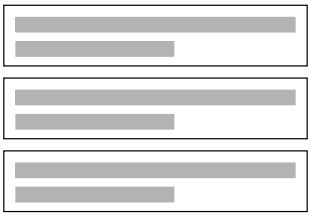


Figure 3.: Exemplary representation of a linear tool.

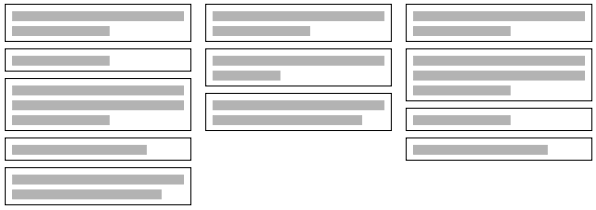


Figure 4.: Exemplary representation of a threaded tool.

In contrast *threaded* styles try to reduce the incoherence of linear lists by offering multiple lists at once, so that participants (or the systems) decide, where new contributions should be appended. The most famous representatives of threaded styles are internet forums – as presented in Figure 4 – or also called message boards. Their difference to regular chat rooms is that messages are often longer and every message in a forum deals with the same topic. Nowadays every forum software offers more than one discussion space, so called threads. Therefore forums have linear and threaded representation styles. Building on these representation style, different forums distinguish themselves through common features or their rules and policies. But in general, every forum allows participants to exchange arguments by means of a sequence of text contributions. In the past those approaches have encountered much criticism: in particular they are believed to lead to a high degree of redundancy and polarization while scaling poorly with the number of involved participants [Klein, 2012]. However, in practice they are, by far, the most commonly used approach to support online argumentation. Based on this, it is very obvious that a forum is not restrictive, it is very often practically used and the usability is very high due to its prominence. Unfortunately it is just a list of several text contributions. Hence they lead to a high degree of redundancy and polarization while scaling poorly. It is therefore evident that – based on the unrestrictiveness – the forum has the lowest ranking in the category of evaluability as well as clearness and a bad score for the scalability. Only a blob of text would be worse. The details can be seen in Figure 2a.

The implementations for pro- and contra-lists try to achieve a better scalability and clearness, but contain only two threads for one proposal, so they are more restrictive than forums. A popular implementation for this is the pro- and contra-tool *ConsiderIt* [Kriplean et al., 2012]. These approaches work well for evaluating a single proposal and are very easy to use, but they are not suitable to deal with more general alternatives since they do not support the exchange of (counter-)arguments. There have been attempts to support argumentation by

tagging the different argumentation types, like in ZENO [Gordon and Karacapilidis, 1997] and its successors Hermes [Karacapilidis and Papadias, 2001] as well as CoPe_it! [Karacapilidis et al., 2009], but they have a lack of structure during large-scale argumentations as well. The summary of the valuations can be seen in Figure 2b.

1.4.3. Graph-Based Approaches

The first mentioned representation styles are insufficient for an easy representation of arguments, because they are only textual. *Graph-based* styles are better for the representation of argumentation due to their graphical assistance. Instead of pure textual representations, this approach displays boxes or nodes for single components of arguments, like claims, facts and statements. The edges or arrows between these boxes are inference rules or relations of the arguments, e.g. they are supportive or attacking, but the logical structure is entirely clear and unambiguous, assuming one understands the conventions. In this class of representation style a huge diversity and amounts of tools can be found, because there are only minimal definitions required. Some tools do not allow cycles, other are not able to represent attacks on links or they have no possibility to represent a group of mutually dependent facts.

The first generation of argumentation tools, respectively online argumentation frameworks, were graph-based, e.g. *Issue-Based Information System* (IBIS) [Kunz, Werner; Rittel, 1970] and *graphical IBIS* (gIBIS) [Conklin and Begeman, 1988]. Until now, there are many tools like *Carneades* [Gordon and Walton, 2006, Gordon, 2017], *Deliberatorium* [Klein and Iandoli, 2008, Klein, 2012] and *Araucaria* [Reed and Rowe, 2004], which are all working with or rather on argumentation maps. In general, these maps should be easy to use, because they were designed to present the structure of reasoning [van Gelder, 2003]. However they only work well for unskilled participants, if the count of nodes is low like the illustration in Figure 5, but if the number of participants and nodes increases, the complexity will be much higher, is stated in the publication of Chapter 4. That is the reason why their practical use and ease of use is very low, whereby the clearness and scalability achieve a higher score. An exception here is the Deliberatorium, where people do not work with the argumentation map, but with a tree-based structure [Klein, 2012]. Although a comparison between the Deliberatorium and a common online forum was done in [Iandoli et al., 2017], both were designed to appear as similar as possible to reduce the chance that dissimilarities in the user interface would influence performance. This shows that the user interface has a high impact on the results, but that is why we only take a look at the general implementation details of each style.

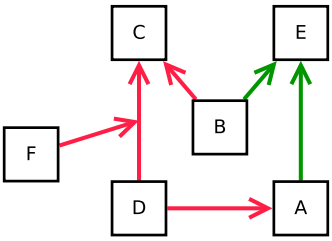


Figure 5.: Exemplary representation of a graph-based tool.

Further systems for argument mapping enable participants to structure their arguments and the relation between them in these maps. The list of argumentation tools is long and many works are listed in [Scheuer et al., 2010]. Based on the fact that these tools are developed from a scientifical point of view and often base on argumentation theory, they are highly restrictive and not easy to use, which decreases the degree of usability. On the other hand, argumentations scale in larger contexts and are more clear than huge lists of text. Based on the fact, that operators still have to read the content of every box or node, the degree of evaluability is still not the best, see the summary in Figure 2c.

1.4.4. Matrix and Container Approaches

Another representation style for online argumentation can be done with the help of matrices and containers. The big advantage of *matrix*-representation is that missing links can be seen very easy and fast, because the arguments' components are the rows and columns of a matrix, like in Figure 6. The cells represent the links, like in *Belvedere* [Suthers et al., 1995, Suthers, 2001,Suthers, 2003]. This representation is more abstract than a graph-based approach, but has an even clearer structure. However, link-to-link relation are not easy to maintain, because either we need a three-dimensional matrix, so that each cell has another, connected cell or we need a new matrix for each cell, which would blow up the space needed for a single argumentation.

	A	B	C	D	E	F
G	+	-	/	+	o	o
H	/	/	+	-	+	-
I	-	+	-	/	o	-
J	o	o	+	o	-	+

Figure 6.: Exemplary representation of a matrix tool.

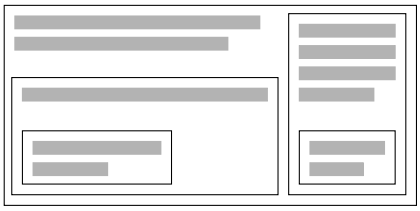


Figure 7.: Exemplary representation of a container tool.

A more simple way to represent arguments can be done with the use of *containers*, where each frame represents a claim or fact that is supported by the framed elements contained in that frame. An example scheme can be seen in Figure 7. This style supports the recognition of argument components belonging together, because their relations are expressed by the residing elements like in SenseMaker [Bell, 1997, Linn et al., 2003]. Nevertheless, the expressiveness is limited to implicit relations only, e.g. tree like structures.

Nevertheless both representation styles are far away from a natural representation of dialogs and unusual for everyday use cases, especially in online participation settings. Their only advantage is a good scalability, because they have a very clear structure, but obvious they lack in all of the other five categories. The estimation for the seven categories can be seen in Figure 2d.

1.4.5. Conclusion

As mentioned above, each style has its own advantages as well as disadvantages. The layered Kiviat diagram for every style can be seen in Figure 8. So far there is no tool, which has a high valuation in every category except in evaluability. This category may be difficult as long as online argumentation systems have to deal with free form text inputs by unskilled participants. Furthermore the possibilities for good evaluations are a research field of the natural language processing rather than online argumentation. Due to these insufficient approaches introduced in the previous chapters, we want to simulate a real-world argumentation in a time-shifted dialog with other participants based on argumentation theory. In our point of view, a dialog-based online argumentation system is a novel and very promising approach for large-scale discussions. This web-based application provides a way to let users investigate or participate in a structured argumentation in a step-by-step fashion. Its main characteristics are: (1) It uses web-technology to provide its services online; (2) Internally it uses a formal model to represent an argumentation. This typically includes at least formal models for arguments, positions and relations; (3) It leads the users through the argumentation in a step-by-step fashion, i.e. it provides a dialog between the system and the user. The user is not required to know or learn about formal argumentation (models). Hence a dialog-based representation and implementation has many, obvious advantages compared to tools so far, but it is not used in practice at the moment. This is why we developed a system, which reuses the arguments given by other participants as base for a time-shifted dialog, whereby only one argument is displayed at a time and the arguments are saved in a WoR, which can be understood as argumentation map.

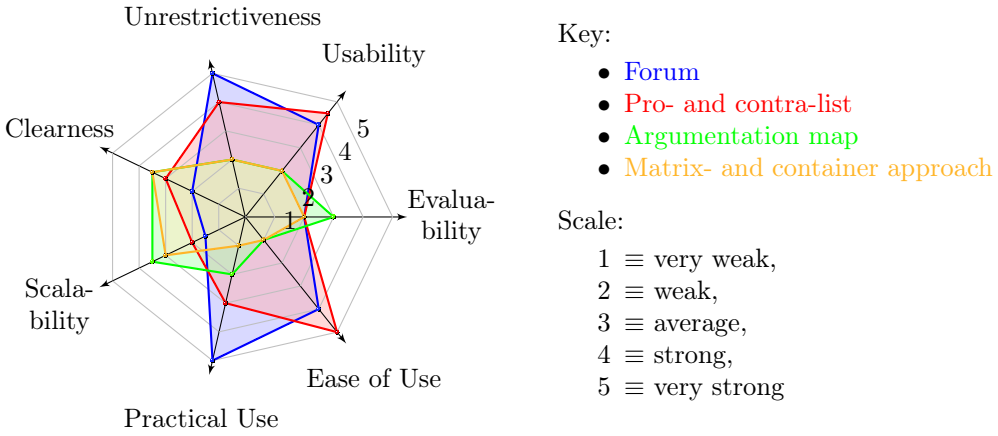


Figure 8.: Layered Kiviat diagrams for different properties of online argumentation tools.

1.5. Thesis Organization

In this chapter, we introduced and motivated our novel approach for a dialog-based argumentation system, related research questions as well as related work. In the next chapters, our contributions are presented, where this thesis is organized as the following sequence:

Chapter 2 mentions the first position paper about our implemented prototype for dialog-based online argumentation [Krauthoff et al., 2016]. We give solutions for the first problems of a new prototype, which are: (1) the presentation of how we envision that dialog-based online argumentation should work, (2) the identification of the main research challenges of providing feedback, navigating the WoR and accepting new arguments, (3) a description of first ideas on how they can be addressed. We have deepened our work on D-BAS, so that Chapter 3 is the journal version based on the position paper [Krauthoff et al., 2016]. We have enhanced the underlying data structure, so that the dialogs we model can be interpreted in three quite different ways: (1) deductive argumentation like [Betz, 2010], probabilistic reason as described in [Skyrms, 2000, Hahn and Hornikx, 2015, Verheij, 2014] and defeasible reasoning according to [Dung, 1995a, Pollock, 1987]. Additionally we give more details about the description of a full prototype implementation of dialog-based online argumentation as well as the results of a first experiment about D-BAS feedback options.

We present the findings of a field experiment in Chapter 4. In May 2017 we tested D-BAS in the computer science courses of the University of Düsseldorf, where we invited all students of this course to take part at an argumentation about the increased number of students in their courses. Further we wanted to find out, whether our approach for a decentralized moderation

system also works in this setting, where unskilled participants are involved into the novel approach.

Although, D-BAS is a fully functional online argumentation tool, we see huge potential for structured online argumentations on arbitrary websites with the help of an embeddable module based on D-BAS. Therefore Chapter 5 deals with the work to embed structured argumentations into arbitrary websites [Meter et al., 2017]. In this process we introduced new possibilities for bootstrapping as well as using statements from any desired websites as proof for given statements in our WoR. With this approach we are able to link, e.g., articles from newspapers with dialogs in D-BAS.

Finally, Chapter 6 contains a summary as well as interesting and challenging research plans for future work in the field of dialog-based online argumentation.

Chapter 2.

Dialog-Based Online Argumentation

This chapter summarizes the contributions and gives a verbatim copy of our paper [Krauthoff et al., 2016]:

> Krauthoff, T., Baurmann, M., Betz, G., and Mauve, M. (2016). Dialog-Based Online Argumentation. In *Proceedings of the 2016 Conference on Computational Models of Argument*. iOS Press. Acceptance Rate: 66.6%.

For several centuries, argumentation – the identification, the analysis, and the evaluation of arguments as well as the persuasion of another participant regarding the truth of a claim [Mohammed, 2016, Gilbert, 1996] – is a vital tool for discussions of two or more participants. With the help of the Internet, discussions now take place independent of time and place and are easier to access than ever before. Unfortunately, these advantages did not result in high quality argumentation for the participants yet, but rather the opposite might be true. Current online discussion tools result in unstructured argumentations, lack structure (see Section 1.4) and it seems that they even encourage polarization as well as redundancy [Klein, 2010, Spada et al., 2014].

Hence, we see dialog-based online argumentation as an approach to solve not only this problem, but also the problems mentioned in Chapter 1.1. Our implementation D-BAS simulates a real-world discussion, where participants are guided through a dialog and are confronted with arguments, stated by other users who already participated. During every step of the dialog, a user has the option to give individual feedback, e.g. state a new justification for her opinion or to choose an already existing justification entered by another participant. In this process the several proposed feedback options are based on argumentation theory, but wrapped in everyday

language, so that even unskilled participants are able to give formally correct answers without having knowledge about technical expressions like *undercuts* or *rebuts*.

2.1. Related Work

The idea of tools for structured argumentation is not new and has been researched on for at least three decades. Current practical solutions can be separated into three categories: Forums, the representatives of linear and threaded lists (see Chapter 1.4.2), allow participants to exchange their opinions via simple text contributions. But it was shown that forums lead to a high degree of redundancy as well as polarization and lack structure [Klein, 2010,Spada et al., 2014]. Nevertheless, they are the most used tool in practice. Pro- and contra lists are representatives of the linear representation styles, outlined in Chapter 1.4.2). They are good to collect statements regarding one proposal, but therefore they do not support the exchange of (counter-)arguments and they lack structure as well. On the contrary, the group of graph-based approaches, see Chapter 1.4.3, deals with the formalization of dialogs and argumentation. Many tools have been proposed for a huge variety of problems, like the *Structured Consultation Tool*, which allows a government agency to elaborate and present a justification for a given action [Bench-Capon et al., 2015] or the *Carneades Opinion Formation and Polling Tool* [Gordon, 2013] to allow participants to provide structured, questionnaire-style feedback on a given argumentation consisting of multiple arguments and positions. The most interesting software is *Arvina* by [Bex et al., 2013]. It allows an user to conduct a dialog between robots and humans. The argumentation map is based on the *Argument Interchange Format* (AIF) [Reed et al., 2008], a formal data structure for the exchange of arguments online, and created by experts before an user can enter the dialog. Nevertheless the user cannot enter new statements on her own during the dialog rather than just using the argumentation map for a limited dialog.

2.2. Contribution

Our contributions in this paper are the following. First, we introduce the foundation of dialog-based online argumentation. We show how a time-shifted dialog based on participants opinions and arguments can be enabled. Thereby the given arguments are recycled by the system and presented to other participants. Second, we outline the main challenges and their solutions of this paper, which are separated in the task of providing feedback for untrained participants, navigating the WoR, which includes bootstrapping as well as selecting the next argument and accepting new arguments, which will be mapped in our WoR.

2.3. Personal Contribution

Tobias Krauthoff, the author of this thesis, contributed to this paper by proposing a data structure for the WoR, which is influenced by the IBIS-model. In difference to IBIS-based approaches, our data structure is able to represent reason-relations like an *undercut*, which are not available in IBIS [Kunz, Werner; Rittel, 1970]. Furthermore, he implemented the prototype and proposed several solutions for the main challenges as well as for the feedback options, which were discussed with Gregor Betz and Martin Mauve. The latter two as well as Michael Baurmann had the initial idea for a dialog-based system. The first detailed development is based on research between Martin Mauve and Tobias Krauthoff only.

2.4. Importance and Impact on Thesis

In this paper we propose not only the implementation of a dialog-based online argumentation system, but provide novel solutions for: (1) bootstrapping a discussion, e.g. finding the starting point in the WoR; (2) selecting the next argument to be presented, whereby the selection is based on the participants' last decision; (3) a way to accept new arguments given by the users and to map them into our WoR and (4) how to provide precise and meaningful feedback for already given arguments. Hence, we present a first glimpse on how dialog-based argumentation can take place online and lay a foundation for further research with our novel approach. With this approach we are able to answer some of the research questions from Chapter 1.2:

- What is an effective way to present dialog-based online argumentation to make it work for untrained participants?
- How should the systems' feedback, that is presented towards the user, be displayed? And how should the list of responses look like that the participants can choose from?
- How do we bootstrap an ongoing discussion or how do we start an empty discussion?
- What is a first approach to navigate through our WoR?
- How do we accept new statements and connect them with our existing WoR?

With our first prototype we are able to answer all but the last two questions of Chapter 1.2. It is clear that more work has to be done so that D-BAS can be used for a first field test outside the setting of our university. Therefore, we have taken the last two questions into account to provide a first field test at our course of study in Chapter 4.

Dialog-Based Online Argumentation

Tobias KRAUTHOFF [a,1], Michael BAURMANN [b], Gregor BETZ [c] and Martin MAUVE [a]

[a] *Department of Computer Science, University of Düsseldorf, Germany*
[b] *Department of Sociology, University of Düsseldorf, Germany*
[c] *Department of Philosophy, Karlsruhe Institute of Technology, Germany*

Abstract. In this position paper we propose a novel approach to online argumentation. It avoids the pitfalls of unstructured systems such as asynchronous threaded discussions and it is usable by any participant without training while still supporting the full complexity of real-world argumentation. The key idea is to let users exchange arguments with each other in the form of a time-shifted dialog where arguments are presented and acted upon one-at-a-time. We highlight the key research challenges that need to be addressed in order to realize such a system and provide first solutions for those challenges.

Keywords. online argumentation, dialog-based approach, computer science, collaborative argumentation, collaborative work, dialog games

1. Introduction

Argumentation, the rational exchange of positions, reasons and justifications, is a vital tool whenever a group of two or more persons needs to decide on a course of action, to determine what to accept as truth, to agree on a set of shared values or to simply reach a common understanding of what the positions of the members of the group are. The Internet has provided the basic infrastructure to enable argumentation for all kinds of groups, no matter how large these groups are, where the members of the group are located or at what time they choose to participate.

Unfortunately, this basic infrastructure has not yet led to the hoped for spreading of rational exchange of arguments. In fact, quite the opposite seems to be true. A quick glance at the discussion section of online-news-media as well as blogs and social media sites shows that the expression of opinions, disputes and controversies in the Internet are often anything but rational. They lack structure and clarity, suffer from frequent repetition of similar arguments, conflate diverse aspects of a subject or are biased, irrelevant, emotionally heated and ill-informed. Furthermore they encourage the balkanization of the participants and they do not scale to large numbers of users. It has been argued [1, 2] that this may be due to the predominant use of forum-based systems which rely on the input of free text.

As a consequence there have been several attempts to provide better support for online argumentation. However, so far, none of them has had really significant practical

[1] Corresponding Author: Tobias Krauthoff, Department of Computer Science, Universitätsstraße 1, 40225 Düsseldorf, Germany; E-mail: krauthoff@cs.uni-duesseldorf.de

impact. One important reason for this may be due to the fact that forum-based systems offer something that other systems do not: they allow for a highly complex exchange of arguments and counter-arguments with an intuitive statement-reply scheme. Other approaches to online argumentation either do not capture the full complexity of argumentation (e.g., pro/con lists) or they require that the user is trained in operating a rather complex technical tool (e.g., the cooperative creation of an argument map).

In this paper we describe a novel approach to support online argumentation, that does not require any prior knowledge or training from the user and avoids the shortcomings of forum-based systems while still allowing complex argumentation. The main idea is to guide participants through the arguments provided by other users so that they perform a time shifted dialog with those that have participated before them. The system 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. It is the task of system – and not of the participants – to translate between those two views. We call this approach dialog-based online argumentation.

A full realization of dialog-based online argumentation requires the solution of several hard problems. We cannot hope to solve all of them in one pass. Therefore the contributions of the paper are limited to: (1) the presentation of how we envision that dialog-based online argumentation should work, (2) the identification of the main research challenges that need to be addressed to realize this approach, (3) a description of first ideas on how they can be addressed.

This paper is structured as follows. In Section 2 we give a brief overview of related work in the area of online argumentation. The general idea of dialog-based online argumentation is presented in Section 3. Section 4 defines the terms that we use in the remainder of the paper. The key challenges that need to be addressed in order to realize dialog-based online argumentation are described in Sections 5, 6 and 7. We concluding the paper with a summary and an outlook to future work in Section 8.

2. Related Work

Forum-based approaches, also called asynchronous threaded discussions, allow participants to exchange arguments by means of a sequence of text contributions. In the past those approaches have encountered much criticism: in particular they are believed to lead to a high degree of redundancy and balkanization while scaling poorly with the number of involved participants [1]. However, in practice they are, by far, the most commonly used approach to support online argumentation.

Online systems for argument mapping enable participants to structure their arguments and the relation between them in an argument map. Examples are Carneades [3,4], Deliberatorium [5] and ArguNet [6]. While those systems do avoid the shortcomings of forum-based approaches, they require the users to become familiar with their notation and the semantics of formal argumentation. Therefore, in practice, they are used by experts or students who want to learn about the logic of argumentation rather than by average participants that want to take part in an online argumentation.

It has been suggested, e.g. ConsiderIt [7], to use online pro and contra lists to aid collective decision making processes. These lists work very well for evaluating a given proposal. However, they are not suitable to deal with more general positions and alternatives since they do not support the exchange of arguments and counter arguments.

The idea of engaging in a formalized dialog to exchange arguments is used by so called dialog games. In these games the participants follow a set of rules to react to the statements of each other, e.g. [8]. They are commonly used as a teaching method and were originally developed without any computers in mind. However, with the ability to let computers enforce the rules, they gained significant attention. A good overview of the current state of development of digital dialog games is given in [9]. 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 are three individual systems that are related to our work. The first one is the *Structured Consultation Tool* (SCT) [10]. Its primary goal is to allow a government agency to elaborate and present a justification for a given action. Members of the public can then evaluate that reasoning in a step-by-step fashion. While the SCT explicitly seeks feedback on the arguments provided by the government agency it does so in a questionnaire kind of way. This is valid for gathering feedback on government proposals, but it is unsuitable for an online argumentation, where the dynamic exchange of arguments is the main focus.

The Carneades Opinion Formation and Polling Tool [11] is part of the Carneades argumentation mapping system. It allows participants to provide structured, questionnaire-style feedback on a given argumentation consisting of multiple arguments and positions put forward by - potentially - many agents. This tool can be regarded as a generalization of the SCT. As with the SCT the questionnaire-style feedback is well suited for an evaluation of government activities by citizens but it does not fit the idea of an online argumentation amongst peers.

The third system that is related to our work is Arvina [12] and its predecessor MAgtALO [13]. Both systems allow a user to conduct a dialog between robots and humans. As a basis they use an existing argumentation specified in a formal language [14] 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 and MAgtALO are 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. Large Scale Online Discussions as a Dialog between a System and many Users

The primary goal of dialog-based online argumentation is to enable any user without prior knowledge or training, to participate efficiently in a large-scale online argumentation. At the same time dialog-based online argumentation avoids or at the very least reduces the problems that plague unstructured online argumentation such as a high level of redundancy, balkanization, and logical fallacies.

The foundation of dialog-based online argumentation is a novel way to navigate an existing set of arguments pertaining to a given subject. Instead of presenting many arguments at once – in maps or lists of arguments – the user is shown only a single argument at a time. It is then possible to select a response from a list of alternatives. Based on this response and, possibly, the data gathered from the responses of other participants, the system selects the next argument that is shown to the user. 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 the user provides feedback on those arguments.

Both, the user and the system, profit from the dialog. The user is efficiently guided towards those arguments that are particularly relevant for her. If done right, this should also eliminate redundancy and balkanization and reduce the occurrence of logical fallacies. The system, on the other hand, will increase its knowledge base with every response from a participant. This can then be used to improve the selection of arguments for the next user and to provide a summary of the online argumentation at hand.

There are at least two obvious research questions when considering the foundation of dialog-based online argumentation: How should the next argument be determined that is presented to the user? And: How should the list of responses look like that the participants can choose from? We will touch upon both questions later.

Dialog-based online argumentation, as described so far, requires a fixed set of arguments that is pre-constructed by experts. In many application scenarios this will be entirely sufficient. It will allow users to form their own opinion regarding the presented arguments and ultimately make a decision on which position to support. The system, on the other hand, will be able to learn about the popularity and perceived interdependency of arguments and positions.

However, for a genuine online argumentation the system has to allow participants to add their own arguments. This raises the question how user input can be integrated in a way that enables the navigation of arguments to operate on user-provided arguments. After all, the users are not schooled in argumentation (software) and will not articulate their views in a formally standardized language. This is the third main challenge for realizing dialog-based online argumentation.

The following sections will give an overview of the terms as well as the three challenges and potential solutions. We are are currently in the process of developing a first prototype of dialog-based online argumentation which can be accesses at `https://dbas.cs.uni-duesseldorf.de/`.

4. Terms and Data Structure

In the following we define the terms that will be used to describe the main aspects of dialog-based online argumentation. Their definition also describes the underlying data structure of our implementation.

Every online argumentation is identified by a **topic**. An example of a topic could be: “Our town needs to cut spending. Please discuss ideas how this should be done”. **Statements** are the most basic primitives used in an online discussion. Examples for statements are: “We should shut down university park”or “Shutting down university park will save $ 100,000 a year”. 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 be taken. “We should shut down university park” is an example for a position.

While experimenting with an early prototype we realized, that we need a somewhat unusual definition of the term “argument”. First of all, there is argumentation for or against statements. This leads to the well-known premise-conclusion-structure of an argument, where both premises and conclusions are statements or negations of statements.

For example: “Shutting down university park will save $ 100.000 a year, therefore we should shut down university park” would be an argument, where “Shutting down university park will save $ 100.000 a year” is the premise and “we should shut down university park” is the conclusion. With this structure it is straight forward to support attacks and rebuttals. An attack is an argument with a conclusion that is the negation of a premise of another argument, while a rebuttal is an argument with a conclusion that is the negation of the conclusion of another argument.

Furthermore, there are arguments that target the validity of other arguments by undercutting attacks. An undercutting attack is an argument that does not reason about statements in another argument but question that a certain statement really supports a conclusion. An example would be: “Yes, drug dealers are using the park to sell drugs but this is not a good reason for shutting down university park since we should not give in to criminals”. In this example the premise is “We should not give in to criminals” while the conclusion is the negative form of the argument “We should shut down university park because criminals use university park to sell drugs”.

As a consequence we use the following definition: an **argument** consists of one or more statements (or their negations) that form the **premise(s)** and one statement or another argument (or the negation of any of those two) that form the **conclusion**.

Together, arguments and statements form a (partially connected) *web of reasons* (WoR).

5. Challenge: Providing Feedback

The most basic building block of dialog-based online argumentation is gathering feedback from a user regarding a given argument. This is done by asking a question derived from the statements pertaining to the argument in the WoR. For example, if the premise of the argument is “Criminals use University Park to sell drugs” and the conclusion is “We should shut down University Park” the question generated by the system could be “What do you think about the following argument: ‚We should shut down University Park‘ because ‚criminals use University Park to sell drugs‘?”

The system then offers a set of answers from which the user can choose. This set has to be constructed in a way that enables an untrained user to provide precise feedback on the argument. A simple choice between: “I agree with this argument” and “I do not agree with this argument” could certainly be made by an untrained participant. However, both statements are not precise and have little significance. For example “I do not agree with this argument” might refer to several distinct scenarios: the user might disagree with the premise, the user might think that the conclusion is not supported by the premise or the user might consider this to be a valid argument but that it is weaker than other arguments supporting the negation of its conclusion.

In order to get precise and meaningful feedback from the user, the system has to differentiate between the scenarios by means of a set of answers that the user can choose from. Experiments with a prototype system that allowed users to react to arguments of a pre-constructed online-argumentation led us to two observations: (1) We need to add alternatives that are not commonly mentioned in argumentation theory, such as “I don’t care about this argument.” (2) We have to take into account that giving feedback on an argument is a two step process. The first step is mandatory and requires just a single

click from the user to determine his initial reaction to the argument. For example, the user could choose: "Yes, criminals use University Park to sell drugs, but I do not think that this is a valid reason to close down University Park." The second step is selecting or entering a statement that supports the choice taken in the first step. For the given example this might be "Because we should not give in to criminals." The second step is only available if the selection in the first step allows for a follow up statement and the user can choose to skip it. Separating the two steps facilitates very fast feedback and a clean user interface design.

On the basis of this general approach, the options can be examined that a participant should have in the first step of providing feedback. 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.

Once the user has selected an answer the system can use this to update the internal information of the WoR and to select the next argument that is presented to the user.

6. Challenge: Navigating the Web of Reasons

The second major challenge for dialog-based online argumentation is how the system should select the arguments that are presented to the participant. We believe that addressing this challenge will have to be a competitive process between different approaches. Any navigation, however, will consist of two parts: (1) bootstrapping the dialog by identifying the first argument that should be presented to a given user and (2) selecting the next argument based on the prior actions of the user.

6.1. Bootstrapping

The first thing that the system needs to do when a new users wants to participate in the online discussion is to choose an initial argument to present to the user. 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 he is interested in. This is the starting point in the WoR. The user is then invited to indicate his attitude towards this position: he can support or attack the position or investigate existing arguments regarding this position.

If the supporting or the attacking option is chosen, the user is asked to select or provide a statement explaining his choice. This statement is used as the premise and the position (or its negation) is the conclusion. Thereby the first argument is complete and bootstrapping is finished. If the user chooses to investigating existing arguments, the system instead selects an initial argument from the WoR where the position (or its negation) under consideration is the conclusion. We have implemented this approach in our prototype and it works surprisingly well.

6.2. Selecting the Next Argument

The selection of the next argument that is presented to the participant can be based on many sources of information. In particular it could rely on the history of actions that

this specific participant has performed and the knowledge gained by the actions of other users. Different kinds of selection criteria could operate on this basis. Furthermore, the selection of arguments might be influenced by the need of the system to learn more about specific arguments or the desire to keep the participant interested in continuing the dialog.

At the moment we use a very simple approach which, nonetheless, illustrates the potential of our idea. We look at the participation history of a user to identify the most recent argument that she provided or supported. Then we search the WoR for an argument of prior users which challenges (attack, rebut or undercut) the argument of the current user. This argument is shown to the current user who then has the opportunity to react to it and thereby provides the next argument. This process continues until the WoR contains no counter argument to the argument of the current user. The overall intention is to simulate a real discussion where participants challenge the arguments of other participants.

7. Challenge: Accepting New Arguments

The key to incorporating new arguments in dialog-based online argumentation is to nudge the users to provide arguments themselves and to connect them with existing arguments in an appropriate way. Currently, we use four mechanisms for this purpose. First, users can enter their own statements only within the dialog. This ensures that whatever statement the user enters, it is automatically connected to the WoR in an appropriate fashion. Second, we apply sentence openers to frame the statements of the users. In this way the user is guided towards making structured and well-formed statements. Third, we automatically match the text entered by users with existing statements in the WoR by means of the Levensthein distance [15] and display the users the top results while they enter their statement. Users can then select one of these results instead of completing their own statement. Finally, whenever users employ the keyword "and" in a premise, the system asks the user if this is in fact a single statement or a sequence of statements. The reply to this question helps the system to identify arguments that have multiple premises.

8. Conclusion and Future Work

In this paper we have presented the idea of dialog-based online argumentation as a time-shifted dialog between the individual users participating in an online argumentation. We have identified the three main challenges that need to be solved in order to realize this idea: providing feedback on existing arguments, selecting the next argument that should be presented to the user and incorporating user input. For each challenge we have provided an initial solution and we have developed a first prototype implementing them.

While the work presented in this paper is sufficient to provide a first glimpse at dialog-based online argumentation, there is a multitude of further research questions that have not yet been addressed. We believe that in particular the selection strategies for the next argument provides a lot of research opportunities. New solutions and inspirations in this area might be derived, e.g., from argumentation theory, the studies on bounded rationality and fallacies of group deliberation, "wisdom of the crowd approaches" or the research area of recommender systems. We also expect that novel ways to embed dialog-

based online argumentation in regular web-content such as blogs or online newspapers will be part of the future work in this area.

Finally, and possibly most importantly, it will be pivotal to perform an empirical evaluation of dialog-based online argumentation in real-life settings.

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. We would like to thank the reviewers for their very helpful comments. Since this work was accepted as a short paper, we were not able to extend it in the way they suggested. We will prepare a full version of this paper that includes both those changes and a full description of our prototype implementation.

References

[1] M. Klein, "Using Metrics to Enable Large-Scale Deliberation," *A workshop of the ACM Group 2010 Conference*, 2010.

[2] P. Spada, M. Klein, R. Calabretta, L. Iandoli, and I. Quinto, "A First Step toward Scaling-up Deliberation: Optimizing Large Group E-Deliberation using Argument Maps," 2014.

[3] T. F. Gordon and D. Walton, "The Carneades Argumentation Framework – Using Presumptions and Exceptions to Model Critical Questions," *6th Comput. Model. Nat. argument Work. (CMNA), Eur. Conf. Artif. Intell.*, vol. 6, pp. 5–13, 2006. [Online]. Available: http://cgi.csc.liv.ac.uk/~floriana/CMNA6/CMNA06Gordon.pdf

[4] T. F. Gordon, "Carneades - tools for argument (re)construction, evaluation, mapping and interchange," http://carneades.github.io/, 2015, [Online, Last access 2016-03-13].

[5] L. Klein, Mark; Iandoli, "Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium," 2008.

[6] D. C. Schneider, C. Voigt, and G. Betz, "Argunet – A software tool for collaborative argumentation analysis and research," 2006. [Online]. Available: http://cmna.csc.liv.ac.uk/CMNA7/papers/Schneider.pdf

[7] T. Kriplean, J. Morgan, D. Freelon, A. Borning, and L. Bennett, "Supporting Reflective Public Thought with ConsiderIt," in *Proc. ACM 2012 Conf. Comput. Support. Coop. Work - CSCW '12*. ACM Press, 2012, pp. 265–274. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2145204.2145249

[8] S. Wells, "Supporting Argumentation Schemes in Argumentative Dialogue Games," *Studies in Logic, Grammar and Rhetoric*, vol. 36, no. 1, pp. 171–191, 2014.

[9] T. Yuan, D. Moore, C. Reed, A. Ravenscroft, and N. Maudet, "Informal logic dialogue games in human–computer dialogue," *The Knowledge Engineering Review*, vol. 26, no. 02, pp. 159–174, 2011.

[10] T. Bench-Capon, K. Atkinson, and A. Wyner, "Using Argumentation to Structure E-Participation in Policy Making," vol. 8980, pp. 1–29, 2015. [Online]. Available: http://link.springer.com/10.1007/978-3-662-46485-4

[11] T. F. Gordon, "Structured Consultation with Argument Graphs," 2013. [Online]. Available: http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-2856073.pdf

[12] F. Bex, J. Lawrence, M. Snaith, and C. Reed, "Implementing the Argument Web." *Commun. ACM*, vol. 56, no. 10, pp. 66–73, 2013.

[13] S. Wells and C. Reed, "MAgtALO: an Agent-Based system for persuasive online interaction," in *AISB 2008 Convention Communication, Interaction and Social Intelligence*, vol. 1. Citeseer, 2008, p. 29.

[14] F. Bex, J. Lawrence, and C. Reed, "Generalising argument dialogue with the Dialogue Game Execution Platform." in *COMMA*, 2014, pp. 141–152.

[15] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, no. 8, 1966, pp. 707–710.

Chapter 3.

Dialog-Based Online Argumentation (Journal)

This chapter summarizes the contributions and gives a verbatim copy of our paper submitted as [Krauthoff et al., 2017a]:

Krauthoff, T., Meter, C., Baurmann, M., Betz, G., and Mauve, M. (2017a). Dialog-Based Online Argumentation. *Computer Supported Cooperative Work (CSCW)*. Impact Factor: 1.590

Note: This publication was rejected after the submission of the dissertation. Nevertheless we have rewritten this paper and it was published as [Krauthoff et al., 2018].

After the publication and presentation of our position paper [Krauthoff et al., 2016], where we proposed the general idea of D-BAS, we wrote an extended version. In this second paper we added more information on our data structure as well as implementation details. Further we performed a first laboratory experiment to verify whether our optimization regarding the feedback options did not deteriorate our system and whether we have included all relevant feedback options an unskilled participant would choose.

Next to our definition of the WoR, we outline that the data structure is not bound to a specific framework like [Gordon and Walton, 2006,Bench-Capon et al., 2015], but our model can be read in three different ways. First, D-BAS can be a tool for deductive argumentation [Betz, 2010], so that incomplete arguments can be reconstructed as deductively valid arguments if all their implicit assumptions are made explicit. Second, the reason-relations can be declared in terms of

probabilistic reasoning [Skyrms, 2000, Hahn and Hornikx, 2015, Verheij, 2014], where a second-order argument undermines or sustains another reason-relation. Third, defeasible reasoning [Dung, 1995a, Pollock, 1987] with a second-order argumentation is possible. Arguments in the second order defeat or attack another argument. This freedom in theory is one major strength of D-BAS compared to other systems.

The first version of D-BAS only had a simple version of the proposed feedback option, where we repeated the full premise and conclusion for each option every time. This inflated each option by a lot and the options had many redundant parts. As an improvement, we now use substitutions like "her point of view" and "my opinion". In addition each substitution and the associated statement represented in the same color. Furthermore, mouse over events highlight both, the term in the feedback option and the appropriate text in the original argument.

In [Krauthoff et al., 2016] we proposed several options to gather feedback from participants. In this paper, we conducted a comparative study with two experiments and 18 participants each to verify whether we proposed all relevant feedback options for everyday argumentation. Each participant of the groups got a snippet from a dialog in D-BAS, where the user stated her point of view and the system presented a counter argument. Afterwards they had to write down their answer on a piece of paper first, and second, the participants could select from of the proposed set of feedback options of D-BAS. One group used D-BAS without color-coded feedback options, whereby the other group used the improved version of D-BAS. The evaluation of this small experiment showed that color-coding improves the comprehensibility as well as clarity and every handwritten answer could be mapped to the associated feedback option.

3.1. Contributions

The contributions in this paper are based on the work of our first paper [Krauthoff et al., 2016]. We extended the version to give a full description of our data structure as well as the many opportunities with the data structure of D-BAS and its theoretical freedom. In addition we presented implementation details and results of an experiment about the completeness of our proposed feedback options and their improvements regarding color coding for several parts of an argument.

3.2. Personal Contribution

The contributing author of this paper, Tobias Krauthoff, supervised a bachelor's thesis, which conducted the experiments on the improvements of our feedback options. The idea for the color-coding and substitution find their origin in discussions between him and Christian Meter. However, the filigree view of argumentation regarding the data structure was proposed by Gregor Betz.

3.3. Importance and Impact on Thesis

In this paper we give further insights into D-BAS. First, it describes implementation details of our prototype implementation. Second, it gives a more detailed overview of the data structure and the theoretical background of D-BAS. Third, it presents the results of a first laboratory test to check the completeness of our feedback options as well as the improvements and fourth, the general ideas, based on the feedback option of [Krauthoff et al., 2016], are outlined.

Noname manuscript No.
(will be inserted by the editor)

Dialog-Based Online Argumentation

Tobias Krauthoff · Christian Meter · Mauve Mauve · Michael Baurmann · Gregor Betz

Received: date / Accepted: date

Abstract In this paper, we propose a novel approach to online argumentation. It avoids the pitfalls of unstructured systems such as asynchronous threaded discussions and it is usable by any participant without training while still supporting the full complexity of real-world argumentation. The key idea is to let users exchange arguments with each other in the form of a time-shifted dialog where arguments are presented and acted upon one-at-a-time. We highlight the key research challenges that need to be addressed in order to realize such a system and provide first solutions for those challenges. Furthermore, we describe a prototype implementing dialog-based online argumentation.

Keywords dialog-based approach, dialog games, collaborative argumentation, collaborative work, online argumentation

A note to the reviewers: this is an extended version of a position paper that we presented as a short paper at COMMA 2016 [1]. While the position paper proposed the general idea of dialog-based online argumentation, this extended version also (1) features the description of a full prototype implementation of dialog-based online argumentation, (2) presents the data structures and the theoretical background upon which the prototype is based, (3) contains the results of experiments conducted with the prototype and (4) extends the ideas presented in the short paper based on the feedback we received at COMMA 2016.

Tobias Krauthoff · Christian Meter · Mauve Mauve
Department of Computer Science, University of Düsseldorf, Germany
E-mail: {krauthoff,meter,mauve}@cs.uni-duesseldorf.de

Michael Baurmann
Department of Philosophy, University of Düsseldorf, Germany
E-mail: baurmann@hhu.de

Gregor Betz
Department of Philosophy, Karlsruhe Institut of Technology, Germany
E-mail: gregor.betz@kit.edu

1 Introduction

Argumentation, the rational exchange of positions, reasons and justifications, is a vital tool whenever a group of two or more persons needs to decide on a course of action, to determine what to accept as truth, to agree on a set of shared values or to simply reach a common understanding of what the positions of the members of the group are. The Internet has provided the basic infrastructure to enable argumentation for all kinds of groups, no matter how large these groups are, where the members of the group are located or at what time they choose to participate.

Unfortunately, this basic infrastructure has not yet led to the hoped for spreading of rational exchange of arguments. In fact, quite the opposite seems to be true. A quick glance at the discussion section of online-news-media as well as blogs and social media sites shows that the expression of opinions, disputes and controversies on the Internet are often anything but rational. They lack structure and clarity, suffer from frequent repetition of similar arguments, conflate diverse aspects of a subject or are biased, irrelevant, emotionally heated and ill-informed. Furthermore they encourage the balkanization of the participants and they do not scale to large numbers of users. It has been argued [2, 3] that this may be due to the predominant use of forum-based systems which rely on the input of free text.

As a consequence there have been several attempts to provide better support for online argumentation. However, so far, none of them has had really significant practical impact. One important reason for this may be due to the fact that forum-based systems offer something that other systems do not: they allow for a highly complex exchange of arguments and counter-arguments with an intuitive statement-reply scheme. Other approaches to online argumentation either do not capture the full complexity of argumentation (e.g., pro/con lists) or they require that the user is trained

in operating a rather complex technical tool (e.g., the cooperative creation of an argument map).

In this paper we describe a novel approach to support online argumentation, that does not require any prior knowledge or training from the user and avoids the shortcomings of forum-based systems while still allowing complex argumentation. The main idea is to guide participants through the arguments provided by other users so that they perform a time-shifted dialog with those that have participated before them. The system is driven by a formal data structure capturing the full complexity of argumentation. User interaction, however, has the structure of a regular dialog as it is performed in everyday life. It is the task of system – and not of the participants – to translate between those two views. We call this approach *dialog-based online argumentation*.

A full realization of dialog-based online argumentation requires the solution of several hard problems. We cannot hope to solve all of them in one pass. Therefore the contributions of the paper are limited to: (1) the presentation of how we envision that dialog-based online argumentation should work, (2) the identification of the main research challenges that need to be addressed to realize this approach, (3) a description of first ideas on how they can be addressed. (4) a prototype implementation of dialog-based online argumentation.

This paper is structured as follows. In Section 2 we give a brief overview of related work in the area of online argumentation. The general idea of dialog-based online argumentation is presented in Section 3. Section 4 defines the terms that we use in the remainder of the paper. The key challenges that need to be addressed in order to realize dialog-based online argumentation are described in Sections 5, 6 and 7. Section 8 is about the description of a full prototype implementation of dialog-based online argumentation. We concluding the paper with a summary and an outlook to future work in Section 9.

2 Related Work

Forum-based approaches, also called asynchronous threaded discussions, allow participants to exchange arguments by means of a sequence of text contributions. In the past those approaches have encountered much criticism: in particular they are believed to lead to a high degree of redundancy and balkanization while scaling poorly with the number of involved participants [2]. However, in practice they are, by far, the most commonly used approach to support online argumentation.

Online systems for argument mapping enable participants to structure their arguments and the relation between them in an argument map. Examples are Carneades [4, 5], Deliberatorium [6] and ArguNet [7]. 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. Therefore, in practice, they are used by experts or students who want to learn about the logic of argumentation rather than by average participants that want to take part in an online argumentation.

It has been suggested, e.g. ConsiderIt [8], to use online pro and contra lists to aid collective decision making processes. These lists work very well for evaluating a given proposal. However, they are not suitable to deal with more general positions and alternatives since they do not support the exchange of arguments and counter arguments.

In contrast to pro and contra lists the *Issue-Based Information System* (IBIS) [9] as well as its graphical implementation IBIS [10], both can be used on large, complex design problems and decision processes. However, participants directly work on the IBIS-graph and therefore the system is not usable for untrained participants.

The idea of engaging in a formalized dialog to exchange arguments is used by so-called dialog games. In these games the participants follow a set of rules to react to the statements of each other, see [11]. They are commonly used as a teaching method and were originally developed without any computers in mind. However, with the ability to let computers enforce the rules, they gained significant attention. A good overview of the current state of development of digital dialog games is given in [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 are three individual systems that are related to our work. The first one is the *Structured Consultation Tool* (SCT) [13]. Its primary goal is to allow a government agency to elaborate and present a justification for a given action. Members of the public can then evaluate this reasoning in a step-by-step fashion. While the SCT explicitly seeks feedback on the arguments provided by the government agency, it does so in a questionnaire kind of way. This is valid for gathering feedback on government proposals, but it is unsuitable for an online argumentation, where the dynamic exchange of arguments is the main focus.

The Carneades Opinion Formation and Polling Tool [14] is part of the Carneades argumentation mapping system. It allows participants to provide structured, questionnaire-style feedback on a given argumentation consisting of multiple arguments and positions put forward by – potentially – many agents. This tool can be regarded as a generalization of the SCT. As with the SCT the questionnaire-style feedback is well suited for an evaluation of government activities by citizens but it does not fit the idea of an online argumentation amongst peers.

The third system that is related to our work is Arvina [15] and its predecessor MAgtALO [16]. Both systems allow a user to conduct a dialog between robots and humans. As a basis, they use an existing discussion specified in a formal language [17] 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 and MAgtALO are 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 Large Scale Online Discussions as a Dialog between a System and many Users

The primary goal of dialog-based online argumentation is to enable any user without prior knowledge or training in argumentation theory, to participate efficiently in a large-scale online argumentation. At the same time, dialog-based online argumentation avoids or at the very least reduces the problems that plague unstructured online argumentation such as a high level of redundancy, balkanization and logical fallacies.

The foundation of dialog-based online argumentation is a novel way to navigate an existing set of arguments pertaining to a given subject. Instead of presenting many arguments at once – in maps or lists of arguments – the user is shown only a single argument at a time. It is then possible to select a response from a list of alternatives. Based on this response and, possibly, the data gathered from the responses of other participants, the system selects the next argument that is shown to the user. 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 the user provides feedback on those arguments.

Both, the user and the system, profit from the dialog. The user is efficiently guided towards those arguments that are particularly relevant for her. If done right, this should also eliminate redundancy and balkanization and reduce the occurrence of logical fallacies. The system, on the other hand, will increase its knowledge base with every response from a participant. This can then be used to improve the selection of arguments for the next user and to provide a summary of the online argumentation at hand.

There are at least two obvious research questions when considering the foundation of dialog-based online argumentation: How should the next argument be determined that is presented to the user? And: How should the list of responses look like that the participants can choose from? We will touch upon both questions later.

Dialog-based online argumentation, as described so far, requires a fixed set of arguments that is pre-constructed by experts. In many application scenarios this will be entirely sufficient. It will allow users to form their own opinion regarding the presented arguments and ultimately make a decision on which position to support. The system, on the other hand, will be able to learn about the popularity and perceived interdependency of arguments and positions.

However, for a genuine online argumentation the system has to allow participants to add their own arguments. This raises the question how user input can be integrated in a way that enables the navigation of arguments to operate on user-provided arguments. After all, the users are not trained in argumentation (software) and will not articulate their views in a formally standardized language. This is the third main challenge for realizing dialog-based online argumentation.

The following sections will give an overview of the terms as well as the three challenges and potential solutions. All of those features, which are described in this paper, have been implemented in a first prototype called *D-BAS* and can be accessed at `https://dbas.cs.uni-duesseldorf.de/`.

4 Terms and Data Structure

In the following we define the terms that will be used to describe the main aspects of dialog-based online argumentation. These definitions also describe the underlying data structure of our implementation. It is important to understand that users in their actual dialog with the system are never directly confronted with the terms defined here – but they will have the additional option for a full access to the inner workings of the system if they wish.

Every online argumentation is identified by a *topic* that describes what the argumentation is about. An example of a topic could be: “Our town needs to cut spending. Please discuss ideas how this should be done”. *Statements* are the most basic primitives used in an online discussion. Exemplary statements are: “We should shut down University Park”, “Shutting down University Park will save $ 100,000 a year”, “Criminals use University Park to sell drugs” or “We should not give in to criminals”. Individual participants might consider a given statement to be true or false. The negation of a statement is itself a statement. A *position* is a prescriptive statement, i.e., a statement which recommends or demands that a certain action can be taken. As an example, we will focus on the position “We should shut down University Park”.

While experimenting with an early prototype, we realized that we need to distinguish between first-order and second-order *argumentation*. First of all, there is argumentation for or against statements. Here, some statement (the premise) is said to be a reason for another statement (the conclusion). This leads to a first-order argument, consisting in a premise-statement, a conclusion-statement and a reason

relation between both. For example: "Shutting down University Park will save $ 100,000 a year, therefore we should shut down University Park" would be a first-order argument, where "Shutting down University Park will save $ 100,000 a year" is the premise which represent a reason for the conclusion "We should shut down University Park".

With this structure it is straightforward to represent undermines and rebuttals. A first-order argument A attacks another first-order argument B iff A's conclusion is the negation of a premise of B; and A is a rebuttal of B iff A and B have contradictory conclusions.

Still, we must not presuppose that users advance only deductive reasons and valid arguments. The reason-relations claimed by users might be more or less cogent – and more or less evident for other users, which may trigger a discussion about the strengths of reason-relations. That is what we call second-order argumentation. Accordingly, we allow reasoning not only about statements but also about whether one statement really supports (attacks) another statement. A second-order argument consists of a statement (the premise) that is cited as a reason for why another reason-relation does not hold. Second-order arguments allow us to express undercutting attacks, namely as arguments against reason-relations pertained in other arguments. An example for this would be: "Yes, drug dealers are using the park to sell drugs but this is not a good reason for shutting down University Park since we should not give in to criminals". In this example the premise is "We should not give in to criminals" while the conclusion is the denial of the reason-relation "We should shut down University Park because criminals use University Park to sell drugs".

As a consequence we use the following definition: an argument consists of one or more statements, which form the premise(s); one statement or the second-order claim that a certain reason-relation does or does not hold, which forms the conclusion; and the reason-relation between premise and conclusion. Together, the arguments of a debate form a (partially connected) *web of reasons* (WoR).

We would like to stress that our data structure and the distinction between first-order and second-order argumentation in a user dialog does not commit us to a specific argumentation-theoretic framework. On the contrary, the dialogs we model can be interpreted in quite different ways:

- Deductive argumentation. The arguments we model can be understood as enthymemes, i.e. incomplete arguments, that can in principle be reconstructed as deductively valid arguments if all implicit assumptions are made explicit (see [18]). On this view, second-order argumentation would actually be argumentation about the plausibility of those implicit assumptions.
- Probabilistic reasons. The reason-relations can be explicated in probabilistic terms (see [19–21]). On this view, a second-order argumentation undermines or establishes the probabilistic reason-relation maintained in another argument.
- Defeasible reasons. The arguments we describe can be understood as defeasible reasons (see [22, 23]). On this view, a second-order argumentation defeats another argument (or attacks a defeater).

We conceive this theoretical openness of our argumentative dialog model as a major strength. Note that the data we generate can also be used to check how well the alternative paradigms of rational argumentation can cope with the discussions we protocol.

5 Challenge: Feedback

The most basic building block of dialog-based online argumentation is gathering feedback from a user regarding a given argument. This is done by asking a question derived from the statements pertaining to the argument in the WoR. For example, if the premise of the argument is "Criminals use University Park to sell drugs" and the conclusion is "We should shut down University Park" the question generated by the system could be "What do you think about the following argument: We should shut down University Park because criminals use University Park to sell drugs?". In the following, we first discuss the options that a user should be offered as a reply to the question. Then we investigate optimizations regarding the way the questions and the potential answers should be represented.

5.1 Feedback Options

After presenting the question regarding an argument, the system offers a set of answers from which the user can choose. This set has to be constructed in a way that enables an untrained user to provide precise feedback on the argument. A simple choice between: "I agree with this argument" and "I do not agree with this argument" could undoubtedly be made by an untrained participant. However, both statements are not precise and have little significance. For example "I do not agree with this argument" might refer to several distinct scenarios: the user might disagree with the premise, the user might think that the conclusion is not supported by the premise or the user might consider this to be a valid argument, but at the same time she might consider, that it is weaker than other arguments supporting the negation of its conclusion.

In order to get precise and meaningful feedback from the user, the system has to differentiate between the scenarios by means of a set of answers that the user can choose from. Experiments with a prototype system that allowed users to react to arguments of a pre-constructed online-argumentation

led us to two observations: (1) We need to add alternatives that are not commonly mentioned in argumentation theory, such as "I don't care about this argument." and (2) we have to take into account that giving feedback on an argument is a two step process. The first step is mandatory and requires just a single click from the user to determine her initial reaction to the argument. For example, the user could choose: "Yes, criminals use University Park to sell drugs, but I do not think that this is a valid reason to close down University Park". The second step is selecting or entering a statement that supports the choice taken in the first step. For the given example this might be "Because we should not give in to criminals.". The second step is only available if the selection in the first step allows for a follow up statement and the user can choose to skip it. Separating the two steps facilitates very fast feedback and a clean design of the user interface.

On the basis of this general approach, the options can be examined that a participant should have in the first step of providing feedback. 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.

Once the user has selected an answer, the system can use this to update the internal information of the WoR and to select the next argument that is presented to the user.

5.2 Optimizing the Representation of Questions and Answers

In an early implementation of our system we simply repeated the full *premise* and *conclusion* for each individual option that the user could choose. For example, the feedback option (3) looked like: "I accept that criminals use University Park to sell drugs, but I do not believe, that this is a good reason for we should shut down University Park.". First tests showed that this leads to very long feedback options where some text parts were repeated several times. Participants in those tests indicated that this was not acceptable since they lost their focus when reading all the feedback options.

As a solution to this problem, we use terms like "my point of view", "their statement" or "their point of view" instead of repeating the actual premise and conclusion of the argument. In order to make sure that the participants can easily determine what those terms refer to, both the terms and the premise or conclusion they refer to are colored in the same way. Furthermore, mouseover highlights both the term in the feedback option and the appropriate text in the original argument. An example for feedback option (3) now looks like: "Right, their argument is true, but I do not believe that this is a argument for their point of view".

5.3 User Evaluation

We conducted two experiments with 18 participants each. The goal of the experiments was to verify whether we have included all relevant feedback options, which (in-)experienced users would use in everyday argumentation. For the first experiment we used the initial feedback options without substituting the premises and conclusions. The second experiment then employed those substitutions including coloring and highlighting. In both experiments the number of male and female participants was about the same and the age of the participants covered a wide range.

Both experiments were designed as follows: the participant was shown two arguments. A first argument represented the opinion of the participant and a second argument was an attack on the first argument. The participant was then asked: "What is your opinion regarding the second argument?". The participant would then write down her answer in a simple text field. Afterwards, the participant was shown the feedback options described above and asked the question: "Which option would you choose?". Comparing the text entered with the chosen option allowed us to determine whether the user is able to indicate the desired reply by using the feedback options. Since we distinguish between three attack types (undermine, undercut and rebut), we separated the experiments into three groups with six users each. Additionally, we selected two users from each group in the first experiment and let them also complete the second experiment. After completing the second experiment, they where asked the following additional questions: (Q1) Are the feedback options clearly represented? (Q2) Is the coloring helpful? (Q3) Are the feedback option easy to understand? (Q4) Have the feedback options an acceptable length?

The first, and possibly most important, result of the experiment was that every single reply given in the text field could be matched to one of the feedback options we described above. This indicates that our feedback options are complete. We then investigated whether the participants provided consistent answers and thereby showing that the feedback options where presented in a way that the user would understand. To this end we recorded a 0, when the free text of a participant could be matched with her selected answer from the feedback options and a 1 otherwise. The averages from the experiments are: 0.278 for the first experiment and 0.167 for the second experiment. In general, we could map the user answers to the appropriate feedback option very often. Especially in the second experiment there were very few differences between the free text and the chosen feedback option. We believe that the improvement is caused by a reduction in the complexity of the feedback options when using color-coded replies. Furthermore, it could be that defined feedback options presented the participant with options that they did not think of before and therefore a difference be-

tween the free text and the selection of the feedback option might not only be caused by a mistake but could be quite deliberate.

The evaluation of the 7th group advocates the improvements, too. The users find the answers more clear to read with an average of 3.33 points out of a maximum of 5 points. Color-coding and substitutions can be seen as useful improvements (average rating 3.5 out of 5), the answers are easier to read (with an average of 3.67 out of 5). At last, with an average of 4 out of 5 points, the length of the answer is more comfortable with the improvements. Summarized, the results of the experiments led us to believe that we do have chosen the right feedback options and a good representation of those options.

6 Challenge: Navigating the Web of Reasons

The second major challenge for dialog-based online argumentation is how the system should select the arguments that are presented to the participant. We believe that addressing this challenge will have to be a competitive process between different approaches. Any navigation, however, will consist of two parts: (1) bootstrapping the dialog by identifying the first argument that should be presented to a given user and (2) selecting the next argument based on the prior actions of the user.

6.1 Bootstrapping

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 to present to the user. 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. This is the starting point in the WoR. The user is then invited to indicate her attitude towards this position: she can support or attack the position or investigate existing arguments regarding this position. The same structure of deliberation can be seen in [24].

After the supporting or the attacking option is chosen, the user is asked to select or provide a statement explaining her choice. This statement is used as the premise and the position (or its negation) is the conclusion. This completes the first argument and ends bootstrapping. If the user chooses to investigating existing arguments, the system instead selects an initial argument from the WoR, where the position (or its negation) under consideration is the conclusion. We have implemented this approach in our prototype and it works surprisingly well.

6.2 Selecting the Next Argument

The selection of the next argument that is presented to the participant can be based on many sources of information. In particular it could rely on the history of actions that this specific participant has performed and the knowledge gained by the actions of other users. Different kinds of selection criteria could operate on this basis. Furthermore, the selection of arguments might be influenced by the need of the system to learn more about specific arguments or the desire to keep the participant interested in continuing the dialog.

At the moment we use a very simple approach which, nonetheless, illustrates the potential of our idea. We look at the participation history of a user to identify the most recent argument that she provided or supported. Then we search the WoR for an argument of prior users which challenges (attack, rebut or undercut) the argument of the current user. This argument is shown to the current user who then has the opportunity to react to it and thereby provides the next argument. This process continues until the WoR contains no counter argument to the argument of the current user. The overall intention is to simulate a real discussion where participants challenge the arguments of other participants.

7 Challenge: Accepting New Arguments

The key to incorporating new arguments in dialog-based online argumentation is to nudge the users to provide arguments themselves and to connect them with existing arguments in an appropriate way. Currently, we use four mechanisms for this purpose. First, users can enter their own statements only within the dialog. This ensures that whatever statement the user enters, it is automatically connected to the WoR in an appropriate fashion. Second, we apply sentence openers to frame the statements of the users. In this way the user is guided towards making structured and well-formed statements. Third, we automatically match the text entered by users with existing statements in the WoR by means of the Levensthein distance [25] and display the users the top results while they enter their statement. Users can then select one of these results instead of completing their own statement. Finally, whenever users employ the keyword “and” in a premise, the system asks the user if this is in fact a single statement or a sequence of statements. The reply to this question helps the system to identify arguments that have multiple premises.

8 D-BAS: Implementing a Dialog-Based Online Argumentation System

We have developed a prototype of dialog-based online argumentation called *Dialog-Based Argumentation System* (D-

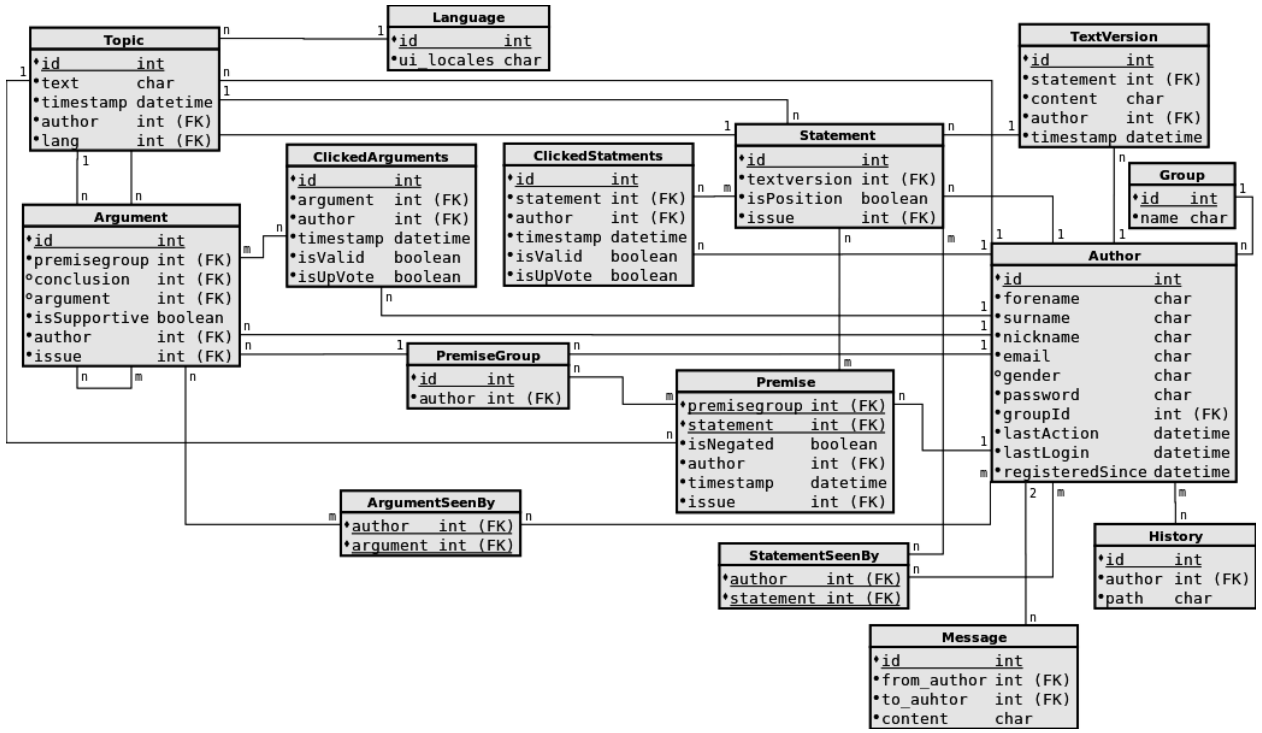


Fig. 1: Entity-relationship model of the core datastructure. Underlined rows with rhombuses are primary keys, filled circles are not nullable and empty circles are nullable.

BAS) which implements the ideas described above. In this section we describe implementation details and specifications about the underlying datastructure. Furthermore, we sketch additional functionality and provide a walk-through of a sample dialog.

8.1 Implementation and Data Structure

D-BAS' backend is written in *Python 3* with support of the *Pylons Pyramid web framework* [26]. We use *nginx* [27] as proxy, *uwsgi* [28] as webserver and *Chameleon* [29] as HTML template engine. Additionally we use *Node.js* [30] with *Socket.IO* [31] for asynchronous and bidirectional communications, e.g. notifications about specific events. D-BAS' datastructure is managed by the Python SQL oolkit *SQLAlchemy* [32]. The entity-relationship model of the core datastructure is displayed in Figure 1.

The key elements of our WoR data structure are topics, statements and arguments as introduced in Section 4. Each discussion has exactly one topic that is connected to all statements and arguments of this discussion. Statements have a text value that may change when a statement is edited. A history of all text versions of a statement is maintained in the table *TextVersions* while the text version of a statement simply maps the most current version of the text. Each version of a text has an author that may belong to one or more groups. A statement also contains a Boolean value that indicates whether this statement is a prescriptive statement, i.e. a position. *Premises* may include one or more potentially negated statements which are realized through the premise and premise groups table. Finally, arguments consist of a group of premises and a conclusion, where the conclusion is either a statement or an argument. The relation between the group of premises and the conclusion can be either a support or an attack, depending on whether the conclusion is a negated statement/argument or not. Next to the tables of our WoR, we have several self-explanatory tables like language, message and history. The tables *ArgumentSeenBy* and *StatementSeenBy* are listings of arguments respectively statements, which were seen by a user, whereby *ClickedArguments* and *ClickedStatement* contain all arguments/statements, which were selected during a discussion by a participant. Therefore we are able to analyze the user's behaviour such as attacking an argument or, later on, changing ones mind.

D-BAS' frontend is built upon a number of established technologies like *HTML*, *JavaScript* with *jQuery* [33], *Bootstrap* [34] and *SASS [35]*. Additionally, D-BAS uses *internationalization* (i18n) for the user interface.

8.2 Additional Functionality

Besides the functionality outlined in the sections above, D-BAS provides several additional features. This includes:

- user management (registration, login, password management, user specific settings);
- options for sharing shortened *uniform resource locators* (URLs) of the discussion via social media, e.g. Facebook, Twitter, and Google+, and mail;
- management interface for editing data directly on the website;
- private messages and notifications.

Also, users are allowed to report statements when they think that they are inappropriate. Our prototype distinguishes between registered and anonymous users. Anonymous users can navigate through the WoRs but they cannot provide new statements or change existing ones.

First experiments with our prototype showed that users frequently requested a history of the current dialog. They also often wanted to go back in the dialog and investigate alternative arguments. In order to support this, arguments provided by the user and the system are presented as a list of chat bubbles similar to those used in state-of-the-art chat clients. The user can simply click on one bubble to navigate directly to that point of the dialog.

Finally, D-BAS has been built in a way to make it is easy to add new functionality or replace existing functionality. For example, it would be trivial to replace the existing method for choosing the next argument that is presented to the user with a new scheme.

8.3 An Exemplary Dialog

The interaction of a user with D-BAS typically starts by logging in (Figure 2). This is an optional step, but currently only registered users are allowed to add their statements to the discussion. Anonymous users are only allowed to read and follow the discussion. Other options, e.g. let anonymous users participate in discussion, are imaginable, but currently not implemented.

Next, the user selects a topic and starts the discussion regarding this topic (Figure 3).

Then the user chooses an initial position she is interested in (Figure 4).

After the user has selected an initial position, she may choose whether to support or attack the position. Alternatively, she may also be undecided and therefore requests that she is shown a supporting argument regarding that position. The three options are display in Figure 5.

In our example the user has chosen to support the initial position. She is then requested to provide a premise for that position. This is shown in Figure 6.

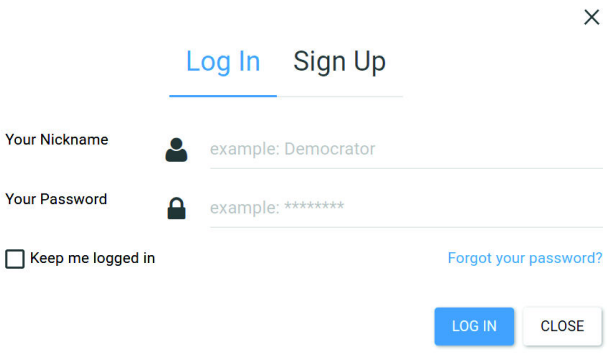


Fig. 2: D-BAS' log-in popup.

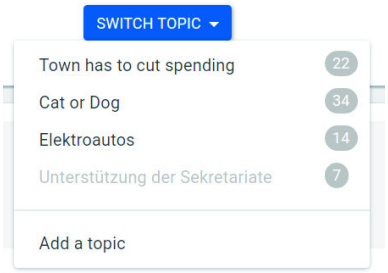


Fig. 3: Selecting a topic.

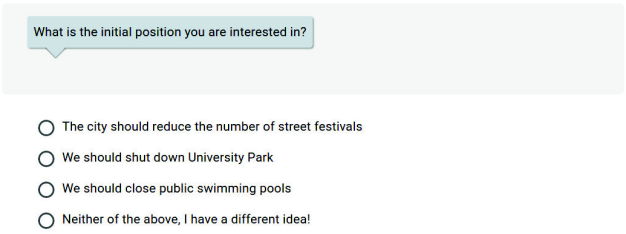


Fig. 4: Choosing an initial position.

Fig. 5: Choosing attitude towards a position.

At this time the bootstrapping is complete: the user has provided an initial argument based on premise and conclusion. As described above, D-BAS now chooses an existing argument that attacks the argument provided by the user. This may be an attack on the premises, the conclusion or the validity of the premises-conclusion relation of an argument. This attack is fetched from the WoR which consists out of arguments of other participants. The user can then

Fig. 6: Selecting a premise for the initial argument.

give feedback on that argument as described in Section 5. This is illustrated in Figure 7.

Fig. 7: Challenging the user's argument and getting feedback from the user.

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 or she may enter a new one. 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 WoR. Any new statement added by the user will be inserted in the WoR. Thus D-BAS knows the trace of the participant and the new statement can be connected in an appropriate way into the WoR. An example how the user chooses a new statement is displayed in Figure 8.

If the user employs the keyword *and* in her premise she is asked to give feedback whether the input is a single statement or a set of statements. In situations where the user

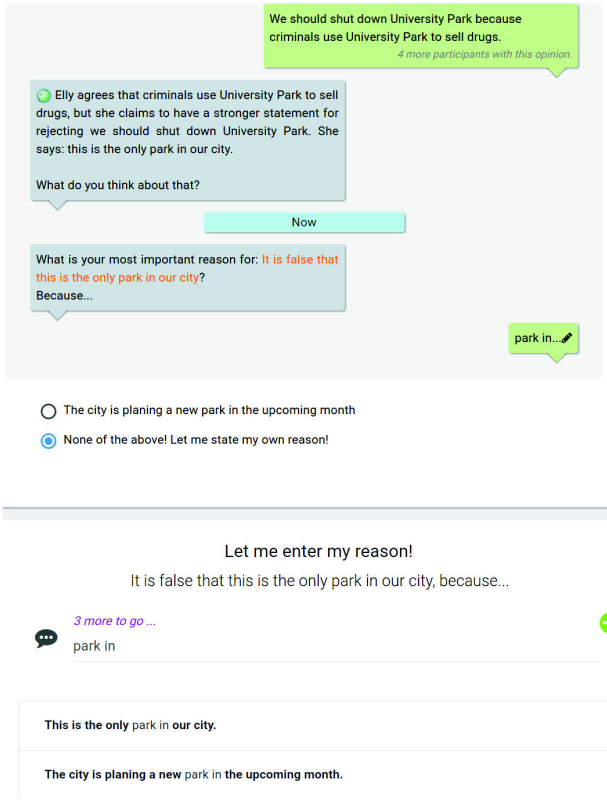


Fig. 8: User interface for entering a new statement.

enters more than one statement with the keyword she can browse through her statements in order to give feedback. An example for the sentence "The city will open a new park next month and we have forest around out small town" is displayed in Figure 9.

The feedback of the user results in a new argument which is then attacked again. This repeats until there are no counter arguments in the current WoR for the argument provided by the user. At this point the user can restart the discussion or go one step back as shown in Figure 10.

All steps above are summarized in the flow chart of Figure 11. Thick lines represent the main dialog of D-BAS, which is a circle of: (1) selecting an argument, (2) getting an attacking argument and (3) gathering feedback. The dashed lines represent additional actions like: adding new statements / positions or request for another attacking argument, because the user does not know how to react to the current one.

8.4 Application Programming Interface

For future applications we implemented an *application programming interface* (API), which provides the possibility to access D-BAS' backend via the Internet [36]. This makes the core functionality of a dialog-based argumentation ac-

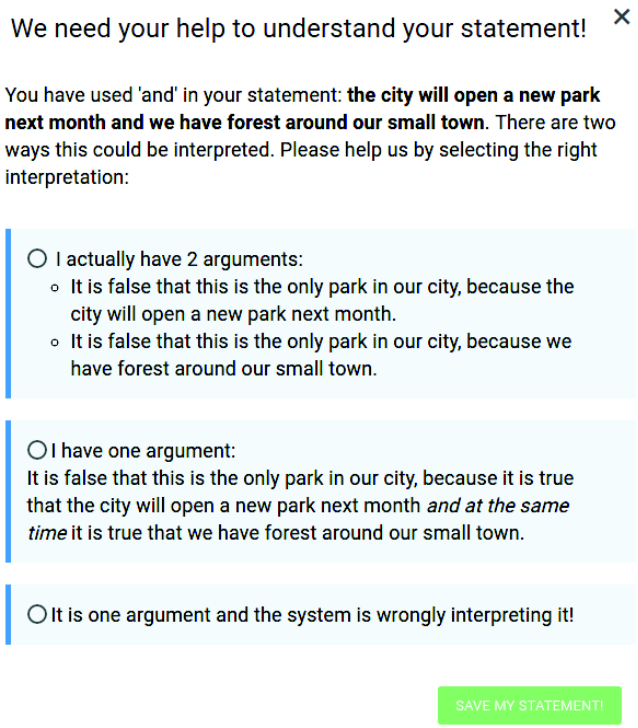


Fig. 9: User interface for feedback on her input.

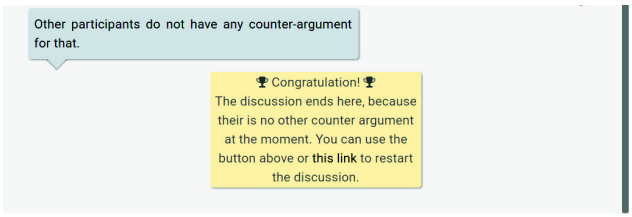


Fig. 10: End-dialog in D-BAS.

cessible to arbitrary applications. All necessary steps of the argumentation can be accessed via a common JSON interface, which are formalized and describe the next steps in the discussion. The general idea behind this is the same as seen in the *Dialogue Game Execution Platform* (DGEP) and the *Dialogue Game Description Language* (DGDL) [37] from ARG-tech[1]. In both approaches, the core is accessible via an interface and the description language specifies which parameters are needed for the next step in the discussion.

Therefore, it is possible to authenticate and participate in the discussion by solely using the API. This raises the argumentation logic of D-BAS to another level and adds another layer of abstraction for new applications.

This abstraction of the core argumentation functionality of D-BAS can then be used to enable dialog-based discussions not only in this standalone-solution but in an arbitrary context. An exemplary implementation can be seen in [36],

[1] http://arg.tech/

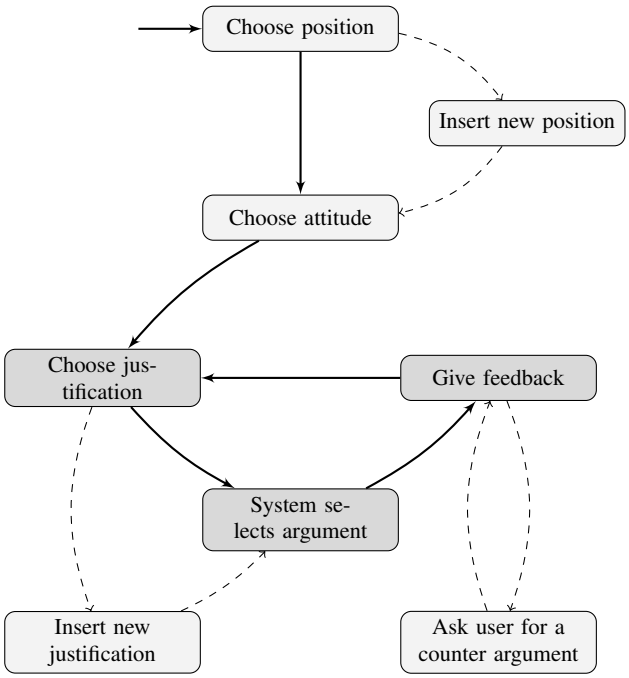


Fig. 11: A discussion in D-BAS step by step.

where D-BAS' backend has been used to implement the web application *discuss*, which embeds dialog-based discussions in any desired website in a similar matter as classical comment sections, for example below an article in online-news-media. More use-cases will be evaluated and implemented by our research group.

9 Conclusion and Future Work

In this paper we have presented the idea of dialog-based online argumentation as a time-shifted dialog between the individual users participating in an online argumentation. We have identified the three main challenges that need to be solved in order to realize this idea: providing feedback on existing arguments, selecting the next argument that should be presented to the user and incorporating user input. For each challenge we have provided an initial solution and we have developed a first prototype implementing them. We also expect that novel ways to embed dialog-based online argumentation in regular web-content such as blogs or online-news-media are a part of this area. To support this we already have implemented an API and *discuss*.

While the work presented in this paper is sufficient to provide a first glimpse at dialog-based online argumentation, there is a multitude of further research questions that have not yet been addressed. We believe that in particular the selection strategies for the next argument provides a lot of research opportunities. New solutions and inspirations in this area might be derived, e.g., from argumentation theory, the

studies on bounded rationality and fallacies of group deliberation, "wisdom of the crowd approaches" or the research area of recommender systems. Furthermore, participants of D-BAS should be able to maintain the content of discussion, so that D-BAS does not need the presence of moderators and is therefore self-maintained and self-controlled by the community. To this end a statement review system would be a large improvement.

Finally, and possibly most importantly, we plan to to perform a large scale empirical evaluation of dialog-based online argumentation in real-life settings, i.e. participatory budgeting.

Acknowledgements This work was done in the context of the PhD programme "Online Participation[2]", funded by the Ministry of Innovation, Science and Research in North Rhine Westphalia, Germany.

References

1. T. Krauthoff, M. Baurmann, G. Betz, and M. Mauve, "Dialog-Based Online Argumentation," in *Computational Models of Argument: Proceedings of COMMA*, 2016.
2. M. Klein, "Using Metrics to Enable Large-Scale Deliberation," in *Collective intelligence in organizations: A workshop of the ACM Group 2010 Conference*, 2010, pp. 103–233.
3. P. Spada, M. Klein, R. Calabretta, L. Iandoli, and I. Quinto, "A First Step toward Scaling-up Deliberation: Optimizing Large Group E-Deliberation using Argument Maps," 2014.
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*, vol. 6, 2006, pp. 5–13. [Online]. Available: http://cgi.csc.liv.ac.uk/~floriana/CMNA6/CMNA06Gordon.pdf
5. T. F. Gordon, "Carneades - tools for argument (re)construction, evaluation, mapping and interchange," http://carneades.github.io/, 2015, [Online, Last access 2017-01-17].
6. M. Klein and L. Iandoli, "Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium," 2008.
7. D. C. Schneider, C. Voigt, and G. Betz, "Argunet – A software tool for collaborative argumentation analysis and research," 2006. [Online]. Available: http://cmna.csc.liv.ac.uk/CMNA7/papers/Schneider.pdf
8. 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*. ACM Press, 2012, pp. 265–274. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2145204.2145249
9. H. W. J. Kunz, Werner; Rittel, "Issues as Elements of Information Systems," 1970. [Online]. Available: http://www.technologyforge.net/enma/6020/6020Lectures/IBISAnalysis/support/KunzRittel.pdf
10. J. Conklin and M. L. Begeman, "gibis: A hypertext tool for exploratory policy discussion," *ACM Transactions on Information Systems (TOIS)*, vol. 6, no. 4, pp. 303–331, 1988.
11. S. Wells, "Supporting Argumentation Schemes in Argumentative Dialogue Games," *Studies in Logic, Grammar and Rhetoric*, vol. 36, no. 1, pp. 171–191, 2014.
12. T. Yuan, D. Moore, C. Reed, A. Ravenscroft, and N. Maudet, "Informal logic dialogue games in human-computer dialogue," *The Knowledge Engineering Review*, vol. 26, no. 02, pp. 159–174, 2011.
13. 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*, vol. 8980, pp. 1–29, 2015. [Online]. Available: http://link.springer.com/10.1007/978-3-662-46485-4
14. T. F. Gordon, "Structured Consultation with Argument Graphs," *From Knowledge Representation to Argumentation in AI. A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday*, pp. 115–133, 2013. [Online]. Available: http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-2856073.pdf
15. F. Bex, J. Lawrence, M. Snaith, and C. Reed, "Implementing the Argument Web." *Communications of the ACM*, vol. 56, no. 10, pp. 66–73, 2013.
16. S. Wells and C. Reed, "MAgtALO: an Agent-Based system for persuasive online interaction," in *AISB 2008 Convention Communication, Interaction and Social Intelligence*, vol. 1. Citeseer, 2008, p. 29.
17. F. Bex, J. Lawrence, and C. Reed, "Generalising argument dialogue with the Dialogue Game Execution Platform," in *Computational Models of Argument: Proceedings of COMMA*, 2014, pp. 141–152.
18. G. Betz, *Theorie dialektischer Strukturen*. Klostermann, 2010.
19. B. Skyrms, *Choice and Chance: An Introduction to Inductive Logic*. Belmont, CA: Wadsworth/Thomson Learning, 2000.
20. U. Hahn and J. Hornikx, "A normative framework for argument quality: argumentation schemes with a bayesian foundation," *Synthese*, vol. 193, no. 6, pp. 1833–1873, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11229-015-0815-0
21. B. Verheij, "Arguments and their strength: Revisiting pollock's anti-probabilistic starting points," pp. 433–444, 2014.
22. P. M. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games," *Artificial Intelligence*, vol. 77, no. 2, pp. 321–357, 1995.
23. J. L. Pollock, "Defeasible reasoning," *Cognitive Science*, vol. 11, no. 4, pp. 481–518, 1987.
24. L. Baccaro, A. Bächtiger, and M. Deville, "Small differences that matter: the impact of discussion modalities on deliberative outcomes," *British Journal of Political Science*, pp. 1–16, 2014.
25. V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, 1966, pp. 707–710.
26. C. McDonough, "Pylons Project: Home," http://docs.pylonsproject.org/en/latest/, 2011, [Online, Last access 2017-01-17].
27. NGINX Inc., "NGINX — High Perfomance Load Balancer, Web Server and Reverse Proxy," https://www.nginx.com/, 2015, [Online, Last access 2016-11-09].
28. Unbit, "The uWSGI project," https://uwsgi-docs.readthedocs.io, 2016, [Online, Last access 2016-11-23].
29. M. Borch and Repoze Community, "Chameleon – Chameleon 2.10 Documentation," http://chameleon.readthedocs.org/, 2011, [Online, Last access 2017-01-17].
30. R. Dahl, "Node.js," https://nodejs.org/en/, 2009, [Online, Last access 2017-01-18].
31. G. Rauch, "Bootstrap - The world's most popular mobile-first and responsive front-end framework." http://socket.io/, 2011, [Online, Last access 2017-01-18].
32. M. Bayer, "SQLAlchemy - The Database Toolkit for Python," http://www.sqlalchemy.org/, 2015, [Online, Last access 2017-01-17].
33. jQuery Foundation, "jQuery," https://jquery.com/, 2015, [Online, Last access 2017-01-17].

[2] `http://fortschrittskolleg.de/en`

34. M. Otto, J. Thornton, C. Rebert, J. Thilo, H. XhmikosR, Fenkart, and P. H. Lauke, “Bootstrap - The world’s most popular mobile-first and responsive front-end framework.” http://getbootstrap.com/, 2015, [Online, Last access 2017-01-17].
35. H. Catlin, N. Weizenbaum, and C. Eppstein, “Bootstrap - The world’s most popular mobile-first and responsive front-end framework.” http://sass-lang.com/, 2015, [Online, Last access 2016-11-02].
36. C. Meter, T. Krauthoff, and M. Mauve, “discuss: Embedding dialog-based Discussions into Websites,” in *Human Computer Interaction International: Proceedings of HCI*, 2017, In Press.
37. S. Wells and C. Reed, “A domain specific language for describing diverse systems of dialogue,” *Journal of Applied Logic*, vol. 10, no. 4, pp. 309–329, 2012.

Chapter 4.

Dialog-Based Online Argumentation: Findings from a Field Experiment

This chapter summarizes the contributions and gives a verbatim copy of our paper [Krauthoff et al., 2017b]:

Krauthoff, T., Meter, C., and Mauve, M. (2017b). Dialog-Based Online Argumentation: Findings from a Field Experiment. In *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence*, volume 2012, pages 85–99. CEUR-WS. Acceptance Rate: $\sim 65\%$

In Chapter 1.1 we stressed out, what we expect from tools for online argumentation. Inter alia with these aspects in mind, we developed D-BAS, which was used for a field experiment. The findings of this experiment are summarized in this paper. We invited all computer science students to participants in a discussion on how to improve the computer science course of studies and how the problems caused by the large number of students can be solved, because the number of students has more than doubled over the past ten years. We had 318 distinct visitors during the experiment period from May, 9^{th} until May, 28^{th} of 2017. Although computer science students are very tech-savy, they are not skilled in argumentation theory and the dialog-based approach is new to them, too.

We started the discussion with a default of two positions and two pro- and contra arguments for each position. 47 of the 318 unique visitors contributed 233 additional statements, which form 235 arguments in total. The final results are available to download as a dump of *PostgreSQL*

database[1] and are licensed under the Creative Commons License CC BY-NC-SA[2]. The downloadable archive contains three versions of the discussion: the original dataset of the discussion in German as well a translated version in English and an English dataset which includes some minor corrections.

D-BAS offers a decentralized moderation system for the users, where any user is able to flag a statement or argument due to the reasons of: (1) *Deletion*: These statements may be off topic, irrelevant, harmful or abusive. If positive collective consensus is reached, statements of this type will be deleted; (2) *Editing*: Edits are proposals where users have submitted and revised a version of an existing statement. If positive collective consensus is reached, the old statement will be replaced by the new one; (3) *Duplicate*: If two separate statements have the same meaning, they may be duplicates. So if collective consensus is reached, the duplicate will be deleted and the original statement will replace it; (4) *Optimization*: Some statements are flagged because they need to be revised. Other users can provide an alternative version of a statement, which will be edited afterwards.

During the experiment users have flagged 47 statements, whereby most of them were flagged correctly, because there were no votes against the flag. We were not only interested in the consensus of votes, but in the behavior of users, too. It can be observed that for each group: contributing statements, flagging statements and reviewing them, there are different power users as well as distinct users enjoying different aspects of contributions to the discussion.

Next to the decentralized moderation system we took a closer look at the quality of argumentation. Users added 22 new positions to the argumentation, whereby each position led to an average argument chain of length three or four. During the field test users have selected 200 undermines, 137 undercuts, 56 rebuts, 44 supports; 19 times they wanted to see another attacking argument and 104 times they wanted to go one step back. The last option will be available, if there is no other counter argument. Otherwise, the user has the option to take another counter argument. This behaviour seems to be very natural, because D-BAS always confronts the user with another attacking argument, so most of the users denied this argument.

After the period of two weeks, we invited all participants to take part in a questionnaire about D-BAS and the discussion. The translated questionnaire can be looked up in Appendix B. Summarized, the evaluation is very positive – even though the benign setting – and the participants report only minor parts, where D-BAS can be improved. Upon completion of the field experiment, we introduced two new queues for our moderation system, where statements could be flagged that consist out of more than one statement, thus statements, which should

[1] `https://dbas.cs.uni-duesseldorf.de/static/data/fieldtest_05_2017.tar.bz2`
[2] `https://creativecommons.org/licenses/by-nc-sa/3.0/`

be split into distinct statements. As a consequence a flag due to a merge of several statements is possible, too.

4.1. Related Work

Only few works consider field experiments with tools for online discussions or focusing on the analysis of distributed moderation processes. According to [Lampe and Resnick, 2004] a form of moderation for online discussions is a necessity rather than a luxury. Their paper analyze the moderation process of the online discussion tool *SlashDot*[3], which is a news and commentary site, mainly for technology issues and open source software. During their study period from May 31^{th}, 2003 through July 30^{th}, 2003 28% of comments received at least one moderation. Additionally they outline that comments, which are not at the top-level or were not rated high, did not get the same attention as other comments did. The question: "If a system of distributed moderation quickly and consistently separate high and low quality comments in an online conversation?" is answered with a qualified yes by the authors of [Lampe and Resnick, 2004].

The authors of [Lampe and Resnick, 2004] continue their research about the website SlashDot in [Lampe et al., 2014], where they observed the same period as well as the period from July 10^{th}, 2004 through October 18^{th}, 2004. They explored the extent to which moderation approaches can reduce an overload of information and foster a more civilized conversations, where the discussion is often labeled as "flaming" or "trolling". Both datasets combined have $867,333$ moderations made on $730,689$ comments. As a result, the authors emphasize that SlashDot's moderation system assigns karma, e.g. reputation scores, for users, when their contributions are in a more appropriate manner. This encourages users to participate in reasoned online conversations and provide good content.

In [Spada et al., 2014] a comparative study between the Deliberatorium [Klein and Iandoli, 2008] (see Chapter 1.4.3) and a forum tool is described. Two groups of 160 participants for both software tools discussed the three questions: "What electoral law should Italy adopt?", "What other important questions should be asked about the electoral system?" and "What topics should be deliberated about in the next intra-party deliberative referendum?" in the period from April 1, 2012 to April 21, 2012. Their experiment shows that the results of argumentation software suffer from the well-known problems of polarization, low signal-to-noise ratio and scattered content. Also, the Deliberatorium induces better analysis of each idea (+120% arguments per idea) and it is also significantly cheaper to manage (42 vs 170

[3] `https://slashdot.org/`

man-hours). All three works show that distributed moderation, reputation points as well as argument mapping itself can increase the quality of argumentation by a multiple.

4.2. Contributions

In this paper, we present the results of a field experiment with D-BAS, where more than 300 participants used D-BAS to discuss how the computer science course of studies can be improved. The experiment lasted for two weeks and was a real-world experiment and not an artificial lab setting. This lead to 255 user-submitted statements, which form 235 arguments together. The experiment lead to the first argumentation map, which was created by users during time-shifted dialogs based on recycled arguments given by other users. In addition all participants were invited to fill out a questionnaire about the discussion D-BAS itself after the two weeks. Results of the questionnaire are presented, too. The experiment indicates that dialog-based online argumentation can result in a high-quality argumentation map without the need of anyone involved being an expert on formal argumentation.

4.3. Personal Contribution

Tobias Krauthoff, who is the author of this thesis, organized and wrote major parts of the paper. He planed and carried out the experiment as well as major parts of the questionnaire, in addition he implemented the review system. Ideas for the review system as well as the argument graph and an opinion barometer are based on discussions with Christian Meter. Both together observed the work of two student assistants, who implemented parts of the graph and the opinion barometer.

4.4. Importance and Impact on Thesis

This paper answers the last two research questions of Chapter 1.2: “How can we display the WoR in an intuitive way and how can we present the information based on the gathered feedback of the participants?” and “How can the community review data from our WoR?”. The first question is solved with the interactive argument graph shown in Figure 14 and specific user page for each participant. This page contains the statistic on the count of created statements, reviewed elements and so on, whereby we only show the nicknames of the participants. In

addition the behaviour of other participants can be seen in an opinion barometer in each discussion step. The barometer offers a bar chart (see Figure 9) as well as a donut chart (see Figure 10). The height of a bar or part of the donut chart is the ratio of the count of users who have seen respectively chosen the statement or feedback option, whereby the width of a part of the donut chart is the count of users who have chosen this option in relation to the total count of users. Second question can be answered by our review system, where distinct participants vote for the given review of another participant. We identify that with this work D-BAS' distributed moderation in the context of online argumentation can improve the contributions by untrained participants and can result in a high-quality argumentation, too.

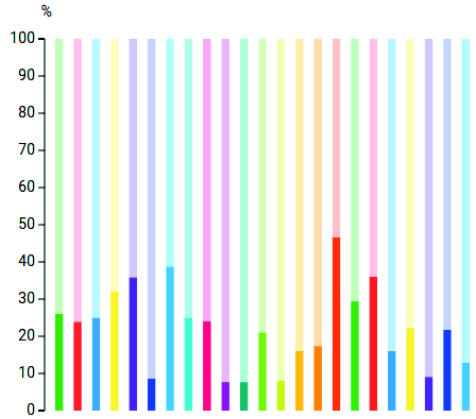


Figure 9.: Opinion barometer as bar chart.

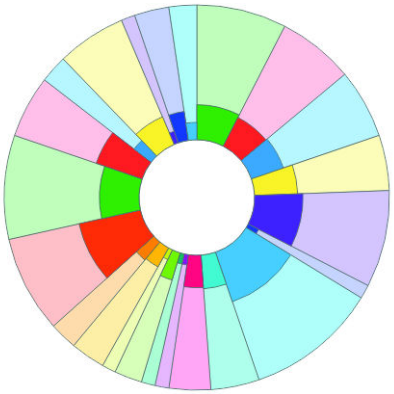


Figure 10.: Opinion barometer as donut chart.

4.5. Additional Findings

In addition to the findings described in this paper, we made several observations during the field experiment that could lead to further improvements of dialog-based online argumentation or to interesting future work. Due to a page limitation we had to drop the Chapters 4.5.1 till 4.5.7. Hence they will be mentioned as additional findings in this thesis.

4.5.1. Visitor Numbers

The development of the user count and page visits can be seen in Figure 11. This data was collected by the open-source analytics platform *Piwik*, now known as Matomo[4]. In this graph an *unique visitor* is an user who visits the page for the first time on a given day. *Visitors* additionally includes visits by users who have not visited the page for more than 30 minutes before returning. As expected, there was a large number of visits right after the invitations were sent out. This stabilized around 50 to 100 visitors per day on weekdays with quite a significant drop during weekends (day 4/5, 11/12 and 18/19). *Unique returners* are users who are returning for the first time after an earlier visit on the same day. *Returners* additionally include every return by any user beyond the first return. Those numbers show that every day there where many users who would monitor the discussion to see how it develops over time.

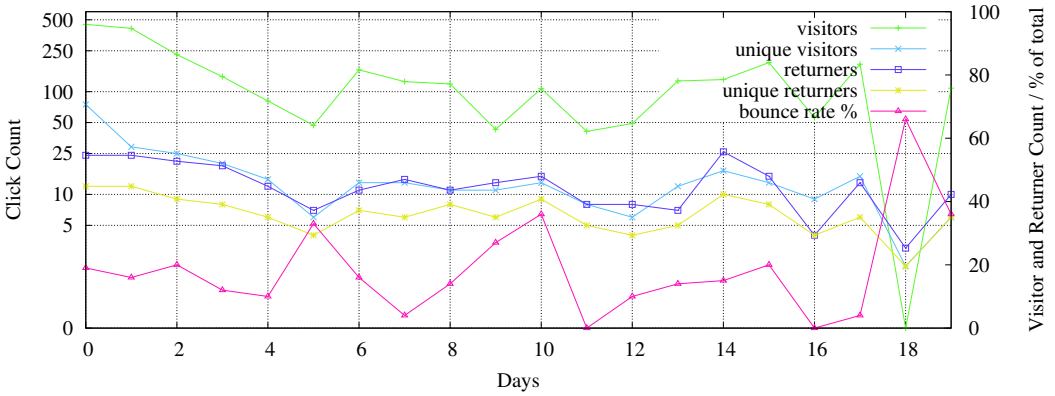


Figure 11.: Growth of visitors and returners over time.

4.5.2. Count of Textversions

The development of the count of statements and related information is plotted in Figure 12. One could assume that users may be lazy, so they do not click through the discussion. This would imply that many positions would be created, but the growth in positions is very slow and constant in contrast to their branch size. It can be seen that the users are interested in reviewing and moderating already given statements, so that almost 50 reviews are given, whereby nine of them regarding the already mentioned duplicates. Hence there are more textversions as statements, because we keep the history of statements.

[4] `https://matomo.org/`

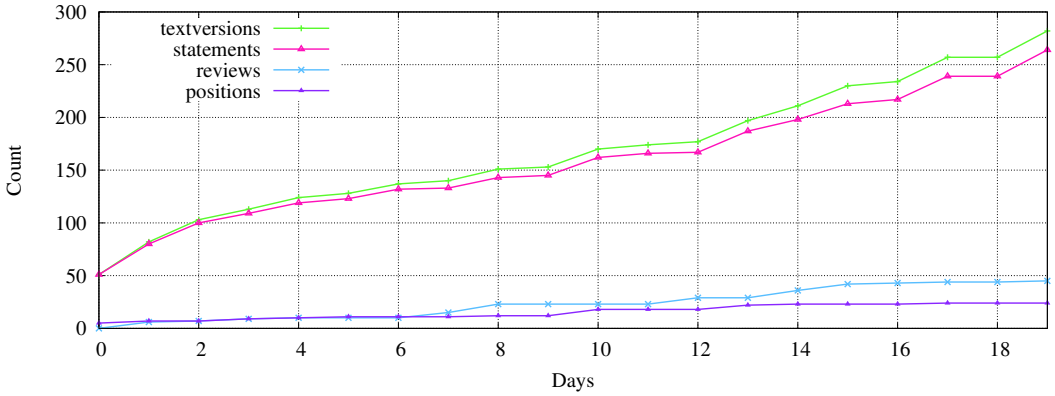


Figure 12.: Growth in number of textversions over time.

4.5.3. Bouncerate

We were interested in the initial reaction of the users when they encountered for the first time. We therefore measured the bounce rate, i.e. the percentage of total users who left the D-BAS on the same page where they started. For the first page of the discussion (this was the page the invitation E-Mail directed the participants to) D-BAS had an average bounce rate of 30%. Consequently, the disproportion between invited students and participants in D-BAS does not rely on the bounce rate, but rather on a lack of interest for the invitation. While there is no commonly agreed-to number that is considered to be a "good" bounce rate, according to blogs [Anders Analytics Limited, 2010, Smarty, 2017, Bez, 2016, Peyton, 2014] an average bounce rate should be below 50% and a good bounce rate should be below 40%. Therefore the bounce rate of D-BAS during our experiment is more than good.

4.5.4. Navigation

We provided an interactive graph-based visualization and navigation of the argumentation as a tool for more advanced users. This feature – displayed in Figure 14 – was somewhat hidden in a sidebar menu. As it turned out, this was a very popular feature, even for regular users. They frequently used it to figure out a starting point where they would like to enter the argumentation. Furthermore, the graph visualization contained a timeline which was often used to identify new statements.

What we learned from the field experiment is that users strongly want more orientation and control over their participation. We will certainly move the graph-based visualization to a

more prominent position. However, we might also reconsider how the functionality that the users want could be realized via other means.

4.5.5. Dealing with Large Numbers of Positions

In the field experiment the discussion had 24 positions. This was much more than we anticipated and it revealed an unexpected problem. At some point the user has to choose which position he is interested in discussing. At the moment this is simply presented as a random list of individual items. This is not optimal, since users typically do not read through a long list positions. At the same time we really do want to present the user with all options since proposing just those that are most popular or that might be of particular interest to the specific participant might lead to a biased-discussion and the formation of filter-bubbles.

A possible solution might be to group positions into categories and then make the selection of the position a two step process: select the category and then select the position from all the positions belonging to that category. For the management of categories it might either be possible to use natural language processing or to use crowd-sourcing, similar to how moderation is done. Both ideas are suggested by Christian Meter.

4.5.6. Length of Statements

We were curious how the distribution of length of statements provided by users looks like. During the discussion 255 additional statements were created with an average length of 124.14 characters. This is less than the length of current Twitter message with 280 characters and even less then Twitter's old limit of 140 characters. The 25%, 50% and 75% percentiles are at 71, 108 and 160 characters. The complete cumulative distribution function is plotted in Figure 13. It can be seen that only 5% of the statements are longer than a Tweet with 280 characters. From this data we gather the fact that the users are able and willing to provide short and precise statements.

4.5.7. Structure of the Argumentation Graph

When looking at the the graph of the argumentation in Figure 14, it is easy to notice a distinct pattern. At the first level there are statements (blue dots). Those blue dots are the target of

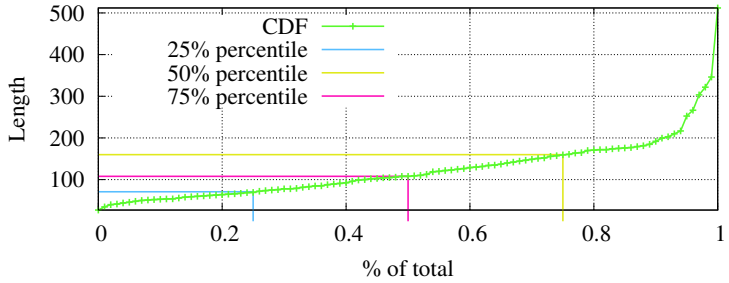


Figure 13.: CDF of the length of statements.

both attacks (red arrows) and supports (green arrows). All other statements are almost always targeted by attacks, only.

This pattern is certainly a result of the way D-BAS conducts the dialog and how it allows the user to react to arguments. When starting the dialog the user can either support or attack a position. Afterwards the user is confronted with an attacking argument regarding her own opinion. At this point the user is then presented with the feedback options presented in [Krauthoff et al., 2016]. Four of the five options attack either the premise or the inference rule of the attacking argument, which clearly favors attacks. However, we experimented with other settings as well. In those settings the statements of users would sometimes not be attacked. Instead, the system would show another statement that was supporting the statement of the user. We also added a feedback option "I share his/her point of view an I do have an additional reason for that." which allowed an user to directly add a supporting statement. However, both changes "felt" wrong. They did not seem to fit the normal pattern of an argumentation.

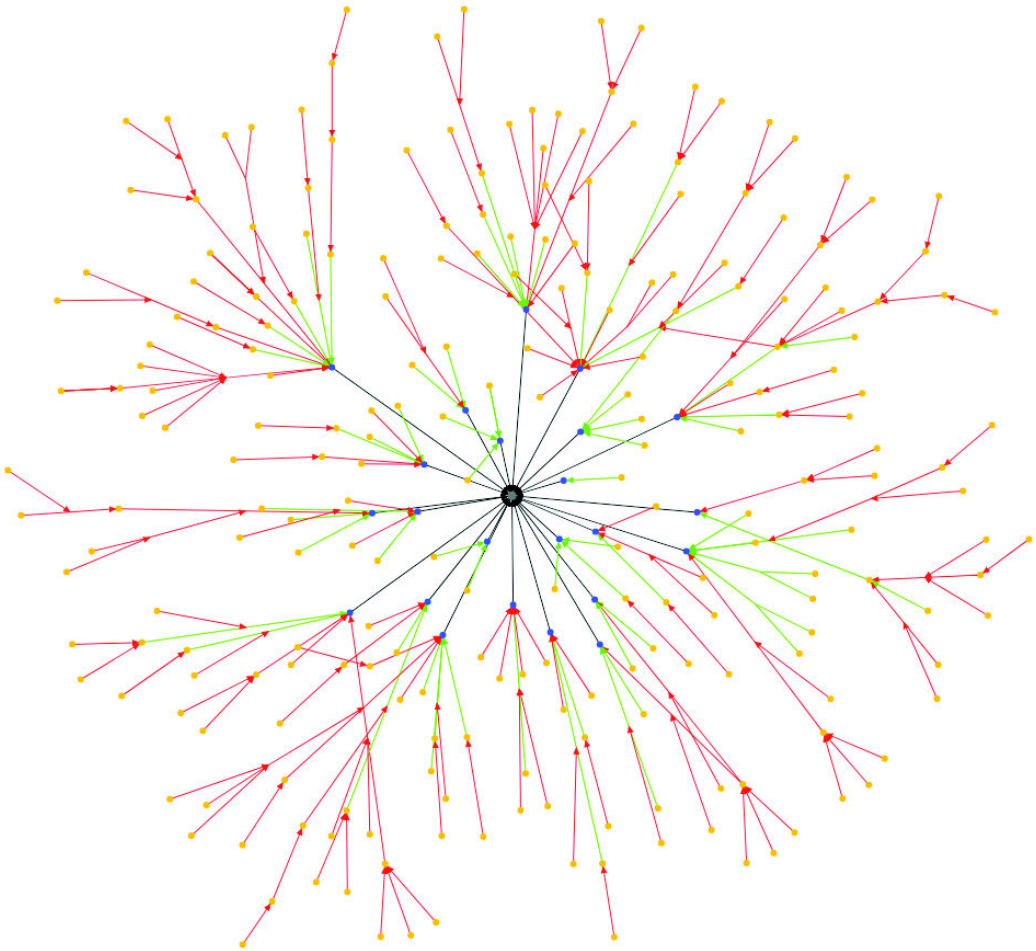


Figure 14.: Argumentation graph created by participants in D-BAS. The grey dot is the root of the discussion, the blue dots are positions and the yellow dots are statements that are not positions. Green arrows denote supporting arguments and red arrows denote attacking arguments

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)[1], which is available on *GitHub* as

[1] https://dbas.cs.uni-duesseldorf.de/

open source software[2]. 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[3] 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

[2] https://github.com/hhucn/dbas
[3] 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 and to gather text contributions regarding specific topics as well as their relations among themselves. 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.

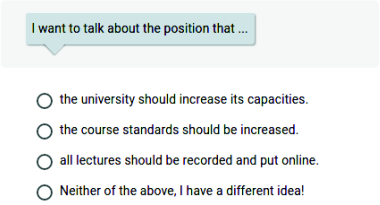


Fig. 1. Choosing an initial position.

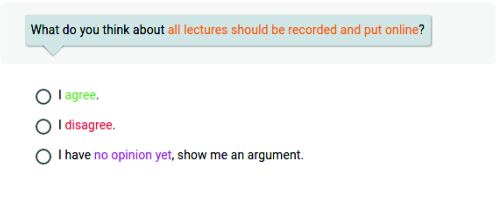


Fig. 2. Choosing attitude towards a position.

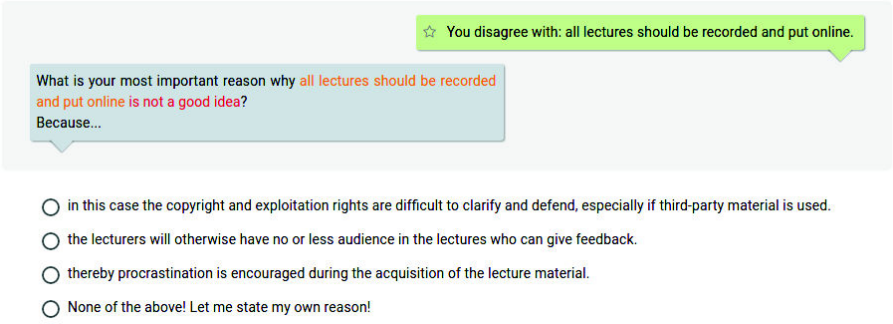


Fig. 3. Selecting a premise for the initial argument.

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.

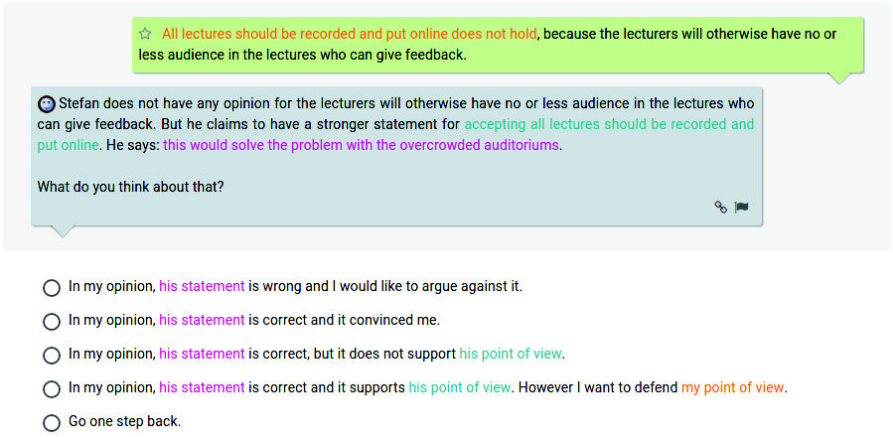


Fig. 4. Challenging the user's argument and getting feedback from the user.

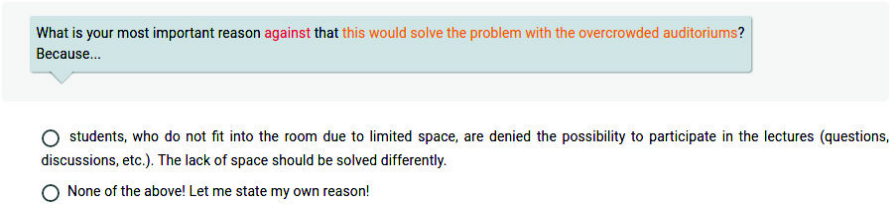


Fig. 5. Justification of the opinion in D-BAS.

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

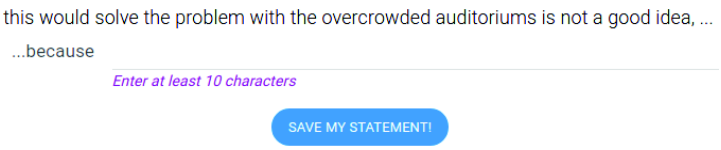


Fig. 6. User interface for entering a new statement.

increased number of students. The number of students has more than doubled in 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^{th} 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^{th} 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[4].

In order to allow others to analyze the discussion, it is available for download[5] as a dump of a *PostgreSQL* database and is licensed under the Creative Commons License *CC BY-NC-SA*[6]. 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

[4] https://dbas.cs.uni-duesseldorf.de/discuss/improve-the-course-of-computer-science-studies#graph
[5] https://dbas.cs.uni-duesseldorf.de/static/data/fieldtest_05_2017.tar.bz2
[6] https://creativecommons.org/licenses/by-nc-sa/3.0/

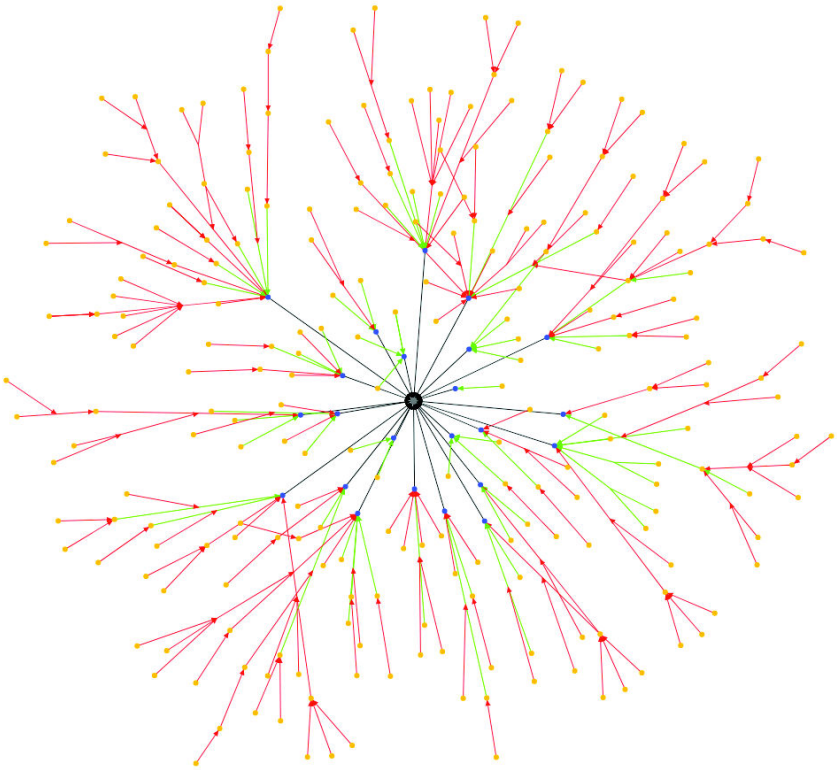


Fig. 7. Argumentation graph created by participants in D-BAS. The grey dot is the root of the discussion, the blue dots are positions and the yellow dots are statements that are not positions. Green arrows denote supporting arguments and red arrows denote attacking arguments

"Lectures should be recorded and released on a streaming platform because ...". 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 chose 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 for-

mal argumentation structure without anyone involved knowing anything about formal argumentation.

The decentralized moderation system implemented in D-BAS has been inspired by *Stack Overflow*[7] 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 loose 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

[7] https://stackoverflow.com/review

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 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.

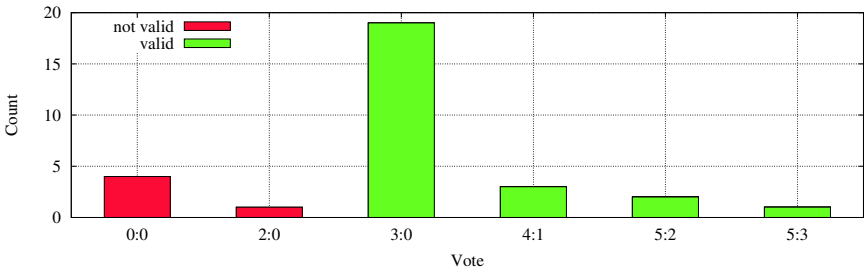


Fig. 8. Overview of voting in the D-BAS review system.

In the discussion, positive consensus was reached in every singe 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 where 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

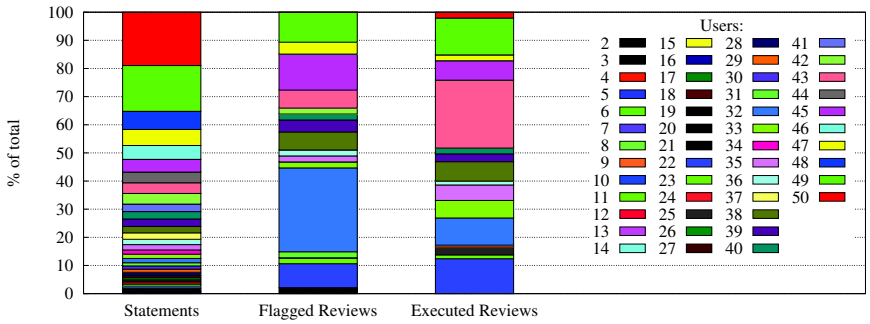


Fig. 9. Distribution of the users activity in D-BAS.

statements. This would come in handy if an inexperienced user includes both premise and conclusion or multiple distinct premises in a single text contribution. 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[8].

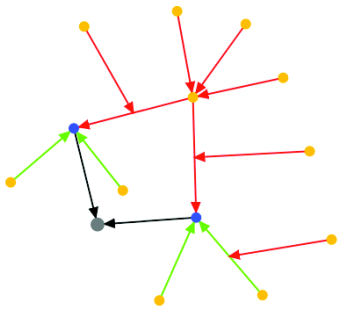


Fig. 10. Connected subgraph during a discussion.

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[9], 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

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

[9] Piwik is an open-source analytics platform: https://piwik.org/.

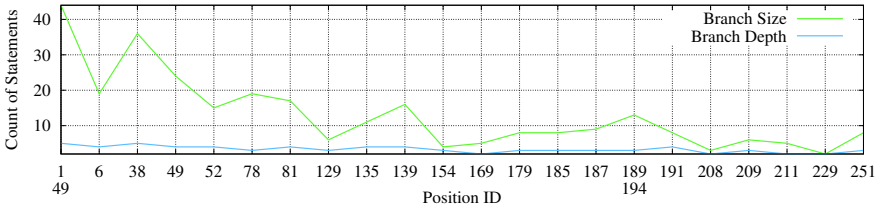


Fig. 11. Size of sub-graphs and length of argument chains for each position.

in relation to the argument it was a reaction to. This holds true for every single 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 were 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*[10].

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

[10] http://www.unipark.com/en/

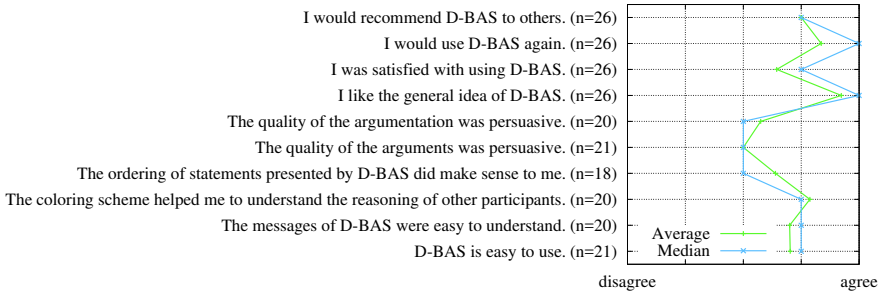


Fig. 12. Users evaluation of usability questions, based on *SUMI* [6].

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

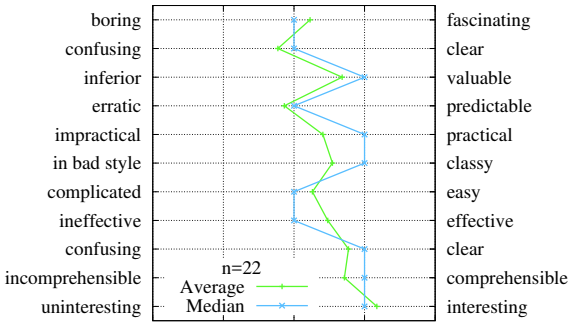


Fig. 13. Users evaluation of bipolar word pairs, based on *AttracDiff* [5].

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.

Chapter 5.

discuss: Embedding Dialog-Based Discussions into Websites

This chapter summarizes the contributions and gives a verbatim copy of our paper [Meter et al., 2017]:

> Meter, C., Krauthoff, T., and Mauve, M. (2017). discuss: Embedding Dialog-Based Discussions into Websites. *International Conference on Learning and Collaboration Technologies*, pages 449–460. Acceptance Rate: 28.3%

Many websites, news media or blogs provide their readers an easy way to discuss or comment on their content. It is precisely this ability to participate in discussions about the content, why many users prefer online content over offline articles. These comment sections are usually below the online content and often vertical-oriented or rather just a list like the one presented in Chapter 1.4.2. These styles are well known for leaving user comments, but have many shortcomings when it comes to online participation and online discussions, like the lack of structure or presence of irrelevant posts [Klein, 2010, Spada et al., 2014].

To solve these problems, *discuss* was implemented to embed the interface of a structured online argumentation tool into the current websites and replaces the previous comment section. By way of example for an online argumentation tool, D-BAS' interface for dialog-based online argumentation is used. D-BAS is a novel approach for online argumentation, whereby the user performs a time-shifted dialog with those users that already participated. Thereby only one argument at a time is presented and the user can give feedback from a list of possibles answers towards this argument or provide new statements. This system has been implemented in [Krauthoff et al., 2016]. With the help of *discuss*, users are able to mark parts in the article

and map these parts with arguments from the structured argumentation tool. Thus, other users can jump directly into the discussion where they are interested in. The different options for this jump are an extended version of the already well-known feedback options of D-BAS. The same options can be used by navigating via the visualization of the argumentation graph in D-BAS, whereby we rearranged the order and modified the options in a way that we have a logically consistent set of answers to respond to a new argument. In contrast to the navigation of the WoR of D-BAS, the user may not have necessarily seen any part of the argument yet. So we propose the following options:

1. accept the conclusion and the premise,
2. accept the conclusion but add a new, more credible premise,
3. accept the conclusion but disagree that the premise leads to accepting the conclusion,
4. reject the conclusion and
5. reject the premise.

5.1. Related Work

The most popular tools to supply the readers with a place to discuss are simple comment sections with a linear interface that has the already mentioned shortcomings. Nevertheless we want to discuss three tools in more detail. At first the system *Disqus* [Disqus, Inc., 2016] enables discussion on arbitrary websites, so that users can discuss on different websites about the same topic. *Disqus* does not tackle current problems of comment sections, however, it enables global discussion via multiple websites, which is an interesting approach. According to this approach, a more heterogeneous peer group can be reached, which results in discussions with higher quality [Hoffman and Maier, 1961]. At second, *rbutr* [Greenup and O'Shannessy, 2017] is a browser extension to link multiple websites sharing the same topic. With this technique websites could be used as source or even proof for arguments in discussion on *rbutr*. The general idea behind this is to use content from websites to support the user's opinion, but the user has to visit another site for discussions and cannot stay on her website containing the current article. At third, *ArguBlogging* [Bex et al., 2014] can be installed as bookmarklet, which is JavaScript code that is stored in the web browsers cache to enable new functionality on websites. The main functionality is to mark and post current text passages from a website to a supported blogging page like *tumblr* or *Blogger*. Important is, that the posts have a link to their source, so that the text passages can be traced back. Nevertheless each user has to manually download and install this bookmarklet for the additional functionality to talk about the article.

5.2. Contribution

The contribution in this paper is a web application, which uses text passages from arbitrary websites as evidence for statement in the same discussion. This enables structured discussion not only on the operators websites, but it is accessible from multiple websites.

At first a possibility to integrate an interface for well-structured online discussion, here D-BAS, into the regular web-content, without any actions by the end users, is provided. Currently only the operator of the website has to include a small JavaScript-file to embed *discuss*. Second, *discuss* allows references between text passages of the original content and the arguments given in the online-discussion tool. Thus any desired statement can be proven with the website. Third, we introduce a way to navigate the argumentation with the help of links and already marked text passages, especially of the argumentation of D-BAS. And fourth, we provide a tool to use the same discussion, hosted on another server, across multiple websites on the Internet.

5.3. Personal Contribution

Tobias Krauthoff, the author of this thesis, contributed to this paper by giving considerations for the interface to access a discussion from outside. Before *discuss* was introduced, the proposed feedback options of D-BAS always based on previous decisions of the user [Krauthoff et al., 2016], but with *discuss* in mind, a new interface had to be written. The unskilled user should be able to give feedback for an argument, which is based on at least one premise, an inference rule and a conclusion, without any knowledge about the argument itself or rather in argumentation theory. The author of this wrote the introduction as well as explanations about dialog-based online argumentation and D-BAS.

5.4. Importance and Impact on Thesis

This paper is an answer to the question on how to link discussions in structured online discussion tools, like D-BAS is, with arbitrary websites. The developed software *discuss* is a web application, which follows basic principles of the dialog-based approach and extends discussions by enabling references, global discussions on multiple sites as well as a flexible inclusion into any website.

discuss: Embedding Dialog-Based Discussions into Websites

Christian Meter, Tobias Krauthoff, and Martin Mauve

Department of Computer Science
Heinrich-Heine-University Düsseldorf, Germany
{meter,krauthoff,mauve}@cs.uni-duesseldorf.de
http://cn.hhu.de

Abstract. In this paper we present the web application *discuss*, which provides a novel approach to embed structured discussions into any website. These discussions employ a formal argumentation system in their backend and can be used in addition to or replace existing comment sections. By interacting with the content of the website, we allow to include this content in the discussion. Furthermore, the same discussions can be accessed from multiple websites to bring their audiences together and create a single large discussion. To form a combined audience, it is necessary to use a common backend and we present an exemplary implementation of this scenario.

Keywords: online argumentation, collaborative work, discussions, dialog-based approach, web technologies, computer science

1 Introduction

Many websites and online news media provide their readers with the opportunity to comment and discuss their content. In fact, the ability to participate in such a discussion or to read what others think about an article is a major reason to prefer online content over offline media. While current solutions are quite suitable to provide simple feedback, they do a rather poor job at fostering meaningful discussions among the readers. This is especially true in those cases where this would be most needed: for articles that receive a lot of reader-feedback due to their popularity or controversial nature.

Commonly, comment sections are located below online articles. They provide a vertical-oriented discussion, where one comment follows the other, often combined with the possibility to directly reply to an individual comment. This is the same design used, for example, by Facebook or Twitter or, in fact, in most forum systems. It is well known, that this design has significant flaws when used for discussions and argumentation rather than simple feedback [1,2], for example redundant comments, lack of structure or simply missing scalability when large numbers of users try to express their opinions. Some online editorials, e.g. The Guardian, are really interested in the comments from the users to enrich the journalism, but often they are abusive, violate their community standards and

the journalists are confronted with huge numbers of comments, which they have to moderate [3]. In general, online editorials show keen interest in the discussions in the comments and are interested in the user's opinions.

To solve these problems and allow for meaningful online argumentation regarding issues raised in an online news media article, we propose to integrate *dialog-based online argumentation* in the website hosting the article. In dialog-based online argumentation the user performs a time-shifted dialog with those users who previously participated in the discussion. The new user can then react to statements from those other users and provide her own statements. This dialog is performed in natural language and the user does not need any specific skill other than being able to read and write. This concept has been implemented in the argumentation system *D-BAS* [4], which is a public accessible web-application. The system also provides an application programming interface to use its backend to remotely perform steps in the argumentation.

In this paper, we present *discuss*, which uses the interface of D-BAS to embed structured discussions in arbitrary websites. discuss is a JavaScript-based extension, which can seamless integrate dialog-based discussions into websites. This tool can be used to enhance or replace existing comment sections whenever a discussion is intended to be held with or among the readers. It gives users that participate in the discussion the option to add references to parts of the online article to their statements. Those parts are then marked in the article, so that other readers can jump right into the ongoing discussion. Furthermore, it is possible to browse and search for those arguments in the discussion that reference the current website.

Our main contributions in this paper are: (1) integrating the interface for dialog-based online argumentation into regular web-content, (2) allowing for references between the argumentation and the content of the website, (3) navigating the argumentation by means of links and search requests and (4) providing a way to use the same discussion across multiple websites.

The paper is structured as follows: Section 2 contains the related work to compare our approach to existing established technologies in the Web. Section 3 is about the prototype D-BAS and the concept of dialog-based argumentation. Section 4 describes the functionality of discuss, while Section 5 focuses on our implementation. The last Sections 6 and 7 conclude the paper and give an outlook to future work.

2 Related Work

The most popular tools to provide reader feedback are simple comment sections in form of a linear list of user statements or the use of forum-based systems. Both display all the negative aspects mentioned above. There are three specific systems that we want to discuss in more detail:

The first system is *Disqus*, which enables discussions on arbitrary websites [5]. In fact, Disqus is a JavaScript application, which needs to be installed by webmasters and brings a hosted alternative to self-hosted comment sections. One

unique characteristic is that instances from different websites can discuss about a global topic. Disqus does not introduce new techniques to enhance discussions and, in general, provides the same functionality as normal comment sections, i.e. add, reply to and vote on comments. This tool is popular for its simplicity and is therefore used quite frequently. It does not address the common problems of comment sections, though. Enabling a global discussion, however, is quite interesting and will also be used in our application.

rbutr [6] is a browser extension which gives the users the ability to link several websites sharing a common topic. These links can then be combined with arguments to introduce information from website B, which might support or rebut the article presented on website A. When a user then visits website A, she is presented a small popup showing that B provides arguments against the contents of A. Therefore, rbutr can be used to link contents from different websites to adjust false information presented on another website. The general idea of using contents from the Internet to support one's own statement is also used in discuss.

ArguBlogging from ARG-tech [7] can be installed as a bookmarklet[1], which needs no further configuration and can directly be used by interested users. The main concept of this tool is to select arbitrary text passages from websites and post them with a reference to the original source on one of the supported blogging sites, currently *tumblr* and *Blogger*. ArguBlogging then creates a post on the user's personal blog and gives her the ability to discuss about this text passage. A popup is presented to other users, who use ArguBlogging, when they arrive on a website, where another user already has selected some text and discussed it on her blog. These other users can then react to this statement and join the discussion. The idea behind the text-selection feature from ArguBlogging is also used in discuss, but in our case it will be directly integrated into a dialog-based discussion.

3 Dialog-Based Online Argumention

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 avoids, or at the very least reduces, 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

[1] `http://www.bookmarklets.com/about/`

demands that a certain action can be taken. Further on 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 [4]. 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 then the user provides feedback on those arguments.

A 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. After she has chosen or provided her position, she is asked to select or provide a statement explaining her choice. This statement is used as the premise, whereas the position forms the conclusion.

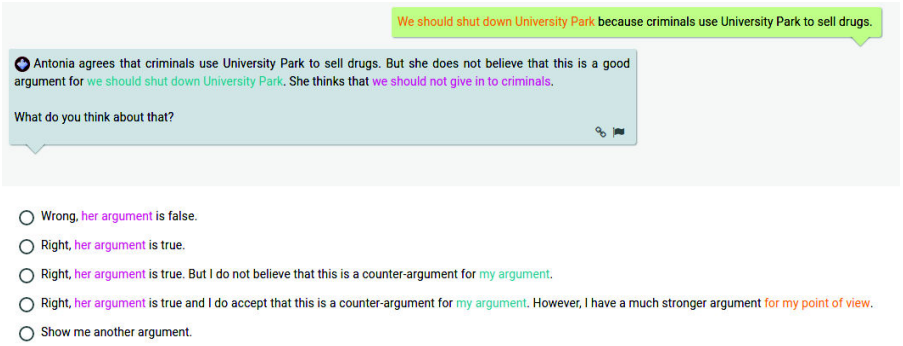


Fig. 1: Gathering feedback during a confrontation in D-BAS.

Once a user is confronted with an argument (see Fig. 1), she can provide feedback on the argument, as shown in Figures 1 and 2. Based on the feedback the system then selects the next argument that is shown to the user. A first

prototype implementing this idea is the dialog-based argumentation system (D-BAS) which is available for testing[2].

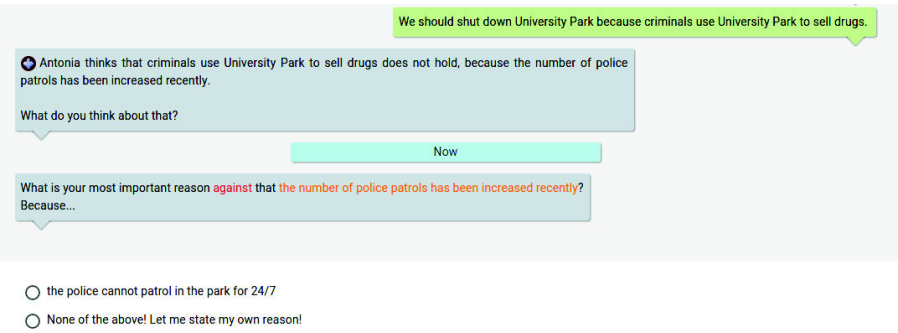


Fig. 2: Justification of the opinion in D-BAS.

4 Functionality of discuss

The idea of discuss is to embed dialog-based online argumentation into regular website content. To describe our implementation in more detail, we use an example where a city wants to reduce its spending and asks the citizen to propose some actions (positions) and to discuss them in detail. A user provided the position "We should shut down University Park" and other users started to discuss this position. This is the current state and we will show through this example how discuss works.

4.1 Embedding discuss into Online Articles

Imagine we have a discuss-powered website and have an article about the situation of the University Park. This article contains facts about the future of the University Park, which other users have proposed to close to cut spending of the city. As an example we assume that the article contains information about an investor, who is going to bear the costs of the park for the next years. We also assume, that our exemplary reader already has knowledge about the ongoing discussion and therefore knows some arguments in it. This is not absolutely necessary, but simplifies the explanation of our contribution.

The user starts reading this article. On her way through, she finds an interesting fact, which she wants to integrate in the discussion about closing University Park. To this end, she selects the appropriate text from the article, e.g. "*But apparently there is an anonymous investor ensuring to pay the running costs for*

[2] `https://dbas.cs.uni-duesseldorf.de`

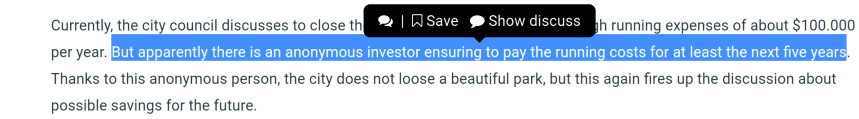


Fig. 3: Tooltip pops up when the user selects a text passage in the article.

at least the next five years". Selecting the text provides her with a *tooltip* (see Fig. 3). Possible options are "*Save*" and "*Show discuss*", where the first option stores the current selection in a *clipboard* for subsequent assembly of an argument for the discussion. The second option toggles the interface to discuss, so that she can directly participate in the discussion. To be flexible and not limited to specific websites, the interface is bound to a *sidebar*, which slides in from the right side, when the second option has been selected. In this sidebar all relevant elements are located which are necessary to participate in the discussion, see Fig. 4.

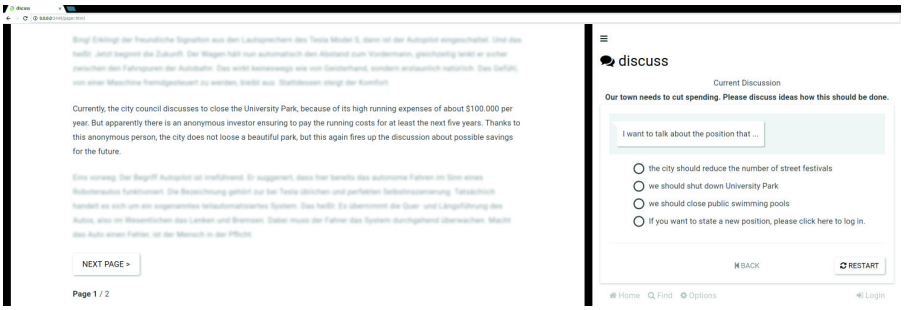


Fig. 4: Side-by-side integration of discuss into an online article.

Now, she can use the selected text and the interface of discuss to participate in the discussion and to create a direct citation of the text passage to her choice. We call these citations *text references* and the user can connect them with any statement in the discussion. With the knowledge the reader gained with this article, she is able to form a counter argument against closing the park and add a suitable reference to her statement. In this case her selection from above, pictured in Fig. 3, seems to be best-fitting, because it describes the future of the University Park in one sentence. These new facts are relevant and can stop the discussion about closing the University Park (if the sources of this article are trustworthy and the contents are true).

As a last step, the reader needs to add her argument to the correct location in the discussion. Since we are assuming, that she already has knowledge about the discussion, she can use the *search engine* for navigation. When the user now wants to add the fact that the investor is going to bear the costs, she needs

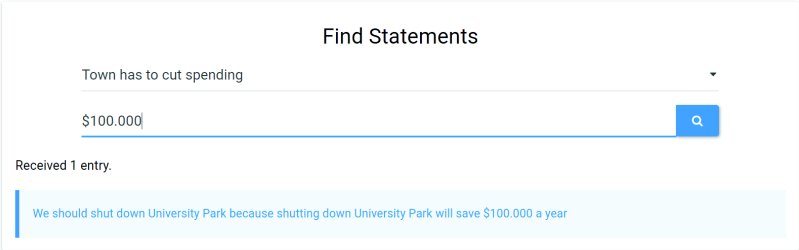


Fig. 5: Find position in the discussion, where the high costs of University Park is discussed.

to find the correct argument from the other user, e.g. "*We should shut down University Park, because shutting down University Park will save $100.000 a year*". Adding the exemplary input "$100.000" in the search engine (see Fig. 5) provides the statement we are looking for and we can now formulate our own argument against it supported by the reference from this article as it can be seen in Fig. 6. This completes the interaction with discuss and the user can close the sidebar to continue reading the article.

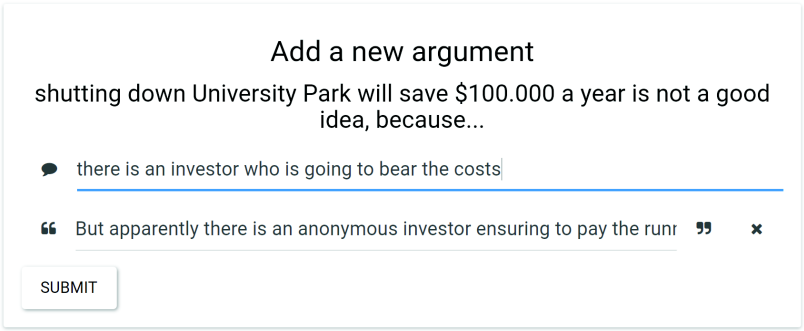


Fig. 6: Constructing a new argument with a text reference.

Arguments, references and their relations are stored in a common backend. All references from this article, which have been used in the discussion, are then highlighted in green color and appear in the text (see Fig. 7). Returning users or new readers of this article can easily see, that these text passages have been used in the discussion, and can interact with them by clicking on a reference. This click again toggles the sidebar and offers a simple interface with all linked locations in the discussion, where this reference has been used (see Fig. 8). Multiple locations are possible, since many users could use the same reference in their arguments.

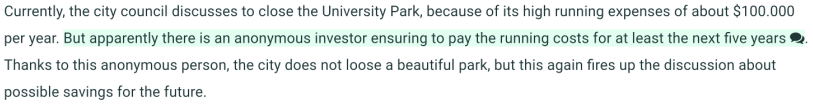


Fig. 7: Highlighted text reference which was previously used by a user.

Text references provide the easiest way to jump to a relevant position in the discussion and to directly start to discuss, because through a reference, our application presents the context of the related argument and asks the user how she wants to react to the argument, see Fig. 9.

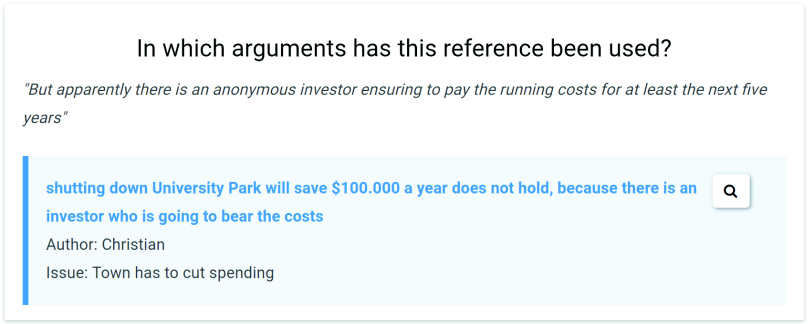


Fig. 8: Jump locations – shows where the references have been used.

4.2 Global Discussion

Common online news media websites, which provide a self-hosted comment section, only allow a local discussion. There is no possibility to leave the borders of this website to interact with users from other news media websites. Disqus [5] provides a feature for inter-website discussions, which we also included in discuss. To realize global discussions, we use one D-BAS instance as a common backend for websites that integrate discuss.

With these global discussions, a more heterogeneous peer group can be reached. Studies showed, that heterogeneous groups have a positive impact on the outcome of a discussion, i.e. solutions emerging from these discussions have a significantly higher quality and those solutions from homogeneous groups were never better compared to the heterogeneous group [8]. Therefore, enabling discussions among users from different online news media, with various levels of education and contrasting opinions, mutually support the discussion. Online news media are often known to have different audiences or specific political orientations and it could be very interesting to analyze discussions between those divergent peer groups, but this leaves the scope of this publication.

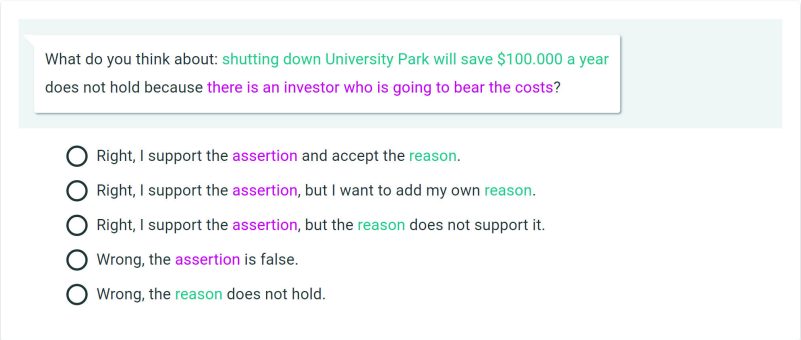


Fig. 9: Jump options – giving the user multiple options how she wants to react to the related argument.

5 Implementation

While implementing discuss we encountered a number of challenges that we outline in the following sections.

5.1 Technical Foundation

To create an application, which does not slow down existing websites and can pick any desired position in the *document object model* (DOM) of the website, we need to have powerful programming techniques and languages fitting our needs. The first prototype was implemented in pure JavaScript, but after few weeks the application became too complex and it was clear that we needed a framework to keep clean code and to reduce complexity. We were also unsatisfied with state-handling and the general language design of JavaScript, which is why we switched to the functional programming language *ClojureScript*[3] and re-implemented the functionality of the first prototype with just a few lines. ClojureScript compiles down to optimized JavaScript code with the *Google Closure Compiler*[4], which results in much faster code than we could manually develop. Using this compiler collection produces also much smaller production files thanks to advanced optimizations and dead code elimination. For dynamic user-interface handling, we chose Facebook's *React.js*[5].

These components allowed us to implement a stable and small web-application without disturbing or conflicting the website it has been embedded into. Since discuss adds many features and DOM manipulations as seen in the previous section, it is very important to choose the best-fitting components, because otherwise it would result in a slow or crowded application.

[3] `https://clojurescript.org/`
[4] `https://developers.google.com/closure/compiler/`
[5] `https://facebook.github.io/react/`

5.2 Including discuss in an Arbitrary Website

Website operators only need to include a single compiled and compressed JavaScript file to enable the features described in this paper for their websites. discuss searches in the DOM for a suitable entry point to enable dialog-based discussions. Selecting the text according to Subsection 4.1 is automatically available and the sidebar invisibly includes itself until the toggle in the tooltip is pressed. If an optional `div` is available in the DOM, an additional interface will be displayed on the website.

Enabling the discussion directly when the user reads the text is a difficult problem: the integration should not disturb the user, but should encourage her to participate in the discussion. In our first approach we put the discussion system directly between the lines of the article and split the text when the user toggled discuss with a switch. But this slide effect was very confusing and is possibly not usable in most kinds of websites. We then experimented with including the interface below the article. This also proved to be a bad choice since the reader then has to jump to the bottom of the article to participate in a discussion triggered by a statement in the article. In our final version, we used the sidebar to interact with discuss. Optionally, the webmaster can include a second interface by simply adding a `div` with a specific ID.

Using a tooltip can be seen on several websites, like Medium [9]. We added listeners to the article to activate the tooltip, when a text passage has been selected. This provides an unobtrusive method to interact with our application.

The clipboard temporarily stores the user's text selections for later usage. This has been implemented to provide the possibility to read the text, store interesting passages and keep on reading, see Fig. 10. In the end, the user can pick her favorite selection to add it to her argument via drag and drop.

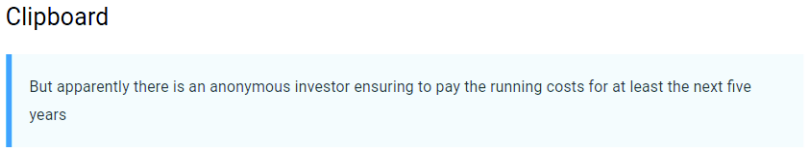


Fig. 10: Using a clipboard to locally store text references.

It is not possible to directly modify the contents of a reference. Our idea is that it should be a direct quotation of the article which is also technically required to find the same text passage in the article. Otherwise, new users will not be able to see the colored reference in the text. We are aware that it is currently still possible to modify the DOM to add a reference of your own desires or to use the browser console for modifications. This would create an untruthful reference, which could lead to false information and false trust in an argument. A server-side verification that the provided string can exactly be found in the article is thinkable, but is currently not implemented.

5.3 Execution Platform

For first testing purposes, we set up a D-BAS instance at our university. As the default configuration uses discuss this backend to directly demonstrate a fully functional application with global discussions enabled. It is possible to use its own backend, which is conform to our application. Therefore, it is not necessarily needed, that the backend is a D-BAS instance – it just needs to provide a suitable interface so that discuss can interact with it.

We are following common best-practices in web development and implemented a RESTful API in D-BAS to expose an interface for external applications, who want to use this dialog-based backend for their applications, whilst discuss is the first project using this interface. This approach for discussion software has already been described in [10] and it presents the general approach how to achieve reusable components in software development, which is why we are also following this structure. Furthermore, [10] proposes the idea to encapsulate the core argumentation logic into an own platform called *Dialog Game Execution Platform* to develop a reusable argumentation core and make it accessible for other applications. In our examples from this paper are we using D-BAS as our default execution platform.

6 Conclusion

Asking the readers to leave a comment below an online news media article is common practice on most websites. But with state-of-the-art comment sections, crowded masses of comments are a typical result. discuss helps to structure discussions and to conduct more productive discourses.

In this paper we used techniques from dialog-based online argumentation to enable our idea of more structured discussions in arbitrary contexts. To achieve this, we implemented discuss as a web application, which follows basic principles of our dialog-based approach and extends discussions by enabling references, global discussions and flexible inclusions into websites.

Feel free to test discuss under `http://cn.hhu.de/discuss` and you are welcome to provide us your feedback.

7 Future Work

We are currently working on more use cases of dialog-based discussions and are evaluating, where our approach could enhance the discourse experience on the Internet. Next, we will extend discuss to support more functions from our backend, e.g. premise groups. In addition, we will evaluate our application in real-world applications and try to cooperate with well-known online news media providers.

Since many people are actively participating in discussions in social networks like Facebook, we will investigate how we can integrate structured discussions into this context. Conceivable are solutions as social bots, which interact with the users based on text messages.

References

[1] Klein, M.: Using metrics to enable large-scale deliberation. In: Collective intelligence in organizations: A workshop of the ACM Group 2010 Conference, pp. 103–233 (2010)

[2] Spada, P., Klein, M., Calabretta, R., Iandoli, L., Quinto, I.: A First Step toward Scaling-up Deliberation: Optimizing Large Group E-Deliberation using Argument Maps (2014). DOI 10.13140/RG.2.1.3863.5688

[3] The Guardian: The dark side of guardian comments (2016). URL `https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments`. Accessed on: 2016-10-14

[4] Krauthoff, T., Baurmann, M., Betz, G., Mauve, M.: Dialog-Based Online Argumentation. Proceedings of the 2016 conference on Computational Models of Argument (COMMA 2016) (2016). DOI 10.3233/978-1-61499-686-6-33

[5] Disqus, Inc.: Disqus - the no.1 way to build an audience on your website (2016). URL `https://disqus.com`. Accessed on: 2016-10-14

[6] rbutr: rbutls, debunkings and counter arguments to misinformation on the internet - think again (2016). URL `https://rbutr.com`. Accessed on: 2016-10-14

[7] Bex, F., Snaith, M., Lawrence, J., Reed, C.: ArguBlogging: An application for the Argument Web. Web Semantics: Science, Services and Agents on the World Wide Web **25**, 9–15 (2014). DOI 10.1016/j.websem.2014.02.002. URL `http://linkinghub.elsevier.com/retrieve/pii/S1570826814000079`

[8] Hoffman, L.R., Maier, N.R.: Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. The Journal of Abnormal and Social Psychology **62**(2), 401 (1961). DOI 10.1037/h0044025

[9] A Medium Corporation: Medium (2016). URL `https://medium.com`. Accessed on: 2016-10-16

[10] Bex, F., Lawrence, J., Reed, C.: Generalising argument dialogue with the dialogue game execution platform. Proceedings of the 2014 conference on Computational Models of Argument (COMMA 2014) pp. 141–152 (2014). DOI 10.3233/978-1-61499-436-7-141

Chapter 6.

Conclusions and Future Work

In the last years, several decision processes (mainly at the local level) took place in Germany. Although these processes were executed according to the order of representative democracy, the election of democratic representatives might not always be sufficient to gain endorsement for political decisions by their citizens. This might constitute the presumption, that citizens want to actively participate on several decisions, which affect them. With the rise of the Internet, people are now able to discuss online – independent of space and time. A central part of online discussions is the exchange of opinions and arguments among the participants. These exchanges can be very heated, if the topics or opinions are controversial at all. Hence many tools for online argumentation have been proposed, but current online discussion tools do not fit for large-scale online discussions and provide many common pitfalls.

The motivation of this thesis is to enable online argumentation to any user without prior knowledge in argumentation theory to participate actively in a large-scale online argumentation with the help of the dialog-based approach. In addition we want to develop a tool, which avoids the mentioned problems of the unstructured presentation of contributions, the need of skilled participants and the scalability as well as the polarization during online discussions. To verify our software D-BAS, or rather to verify the solutions to the common problems, our approach has been tested in a field experiment with unskilled participants and a heated issue. In the next steps we want to embed the techniques of a structured discussion in different contexts on the Internet, for example in online news media, to enhance the discussion-culture in comment sections.

In the following, we conclude our findings about the proposed dialog-based approach and summarize the findings from our implementation and the field experiment. Afterwards we introduce some future research topics, we came in touch with, and finally this thesis is completed by some closing words.

6.1. Conclusion

The majority of online argumentation tools suffer many problems like the lack of structure, the usability by unskilled participants or the well-known problems of scalability and polarization. Therefore, we introduce dialog-based online argumentation, which does not require any prior knowledge or training from the user and avoids the shortcomings of other systems while still allowing complex argumentation. We released the achievements of D-BAS in four publications which had the purpose to implement, test, evaluate and spread our tool for large-scale online discussions.

In conclusion, we focused on introducing and motivating D-BAS as a sample implementation for dialog-based online argumentation in the Chapters 2 and 3. Hence we outlined the three main challenges of the new kind on argumentation: 1) bootstrapping and navigating the WoR; 2) providing feedback options for already given arguments and 3) accepting new statements by users and connect them to the existing WoR. In addition, we proposed solutions to these challenges. Furthermore we provided more information about the data structure as well as the possibilities how our model can be interpreted. Next, we outlined the finding from a first laboratory experiment about the completeness of our proposed feedback options as well as the related mechanisms of color-coded statements and substitutions.

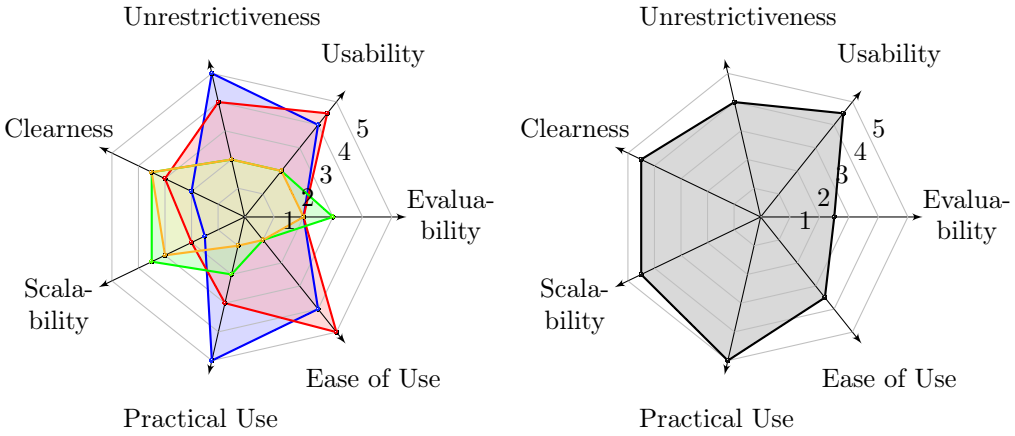
One goal of D-BAS is to be suitability for everyday use, therefore we presented the findings from our first field experiment in Chapter 4. We have invited all computer science students at our university to a discussion about how to improve the computer science course of studies and how the problems caused by the large number of students can be reduced. We had 318 distinct visitors during the experiment period from May, 9^{th} until May, 28^{th} of 2017. About 15% of these visitors contributed 233 statements, which lead to 235 arguments in total. In addition, we tested our mechanisms of a decentralized moderation system, where each user can flag statements due to reasons of deletion, edits, duplicates or optimizations. To avoid abuses, only other participants can take a vote whether they accept or decline the edits or deletions. After the discussion, we invited all students to take part in a questionnaire about the discussion itself as well as about the design of D-BAS. In summary, we found out that D-BAS is a very promising approach and it is accepted by the satisfied participants as well as working well for large-scale online discussions. Upon completion of the field experiment, we implemented two new queues for “splitting” and “merging”, to our moderation system, because three inputs were malformed and we want a system, that can be fully controlled by the participants. The first queue splits statements into new ones, while the latter queue concatenates several statements into a new one.

In the fourth part of our research, we were interested to replace common comment sections of websites like newspapers and/or blogs with the dialog-based approach. The results are presented in Chapter 5. Hence we expanded the application programming interface of D-BAS, so that every discussion step is available for external software. Thus we are able to map any text piece on other websites with given statements in such way that users can bootstrap a discussion in D-BAS via external text pieces on arbitrary websites.

To sum up, this thesis presents a novel approach for large-scale online discussions, which avoids common pitfalls. We have circumvented or rather solved the problems of Chapter 1.1 and we provided solutions on several question from Chapter 1.2, e.g. how to present dialog-based online argumentation, how the provision of feedback options should look like or how to build a WoR during a dialog with untrained participants.

Our results are implemented as microservices and are available as open-source software on GitHub[1]. In addition, we host D-BAS as live system, where everyone can start and share discussions on their own[2]. Logins are also possible via the open authentication protocol, whereat we currently use Facebook, GitHub and Google as providers. We outlined in Chapter 3 that D-BAS' datastructure can be interpreted in quite different ways. Although it was developed by considering real world argumentation and therefore independent of any argumentation model, discussions can be interpreted as defeasible knowledge bases and exported into formats of three well-established argumentation models [Neugebauer, 2017]: the abstract Argumentation Frameworks [Dung, 1995b], the ASPIC Framework [Caminada and Amgoud, 2007], and the Abstract Dialectical Frameworks [Brewka and Woltran, 2010]. As presented in Chapter 1.4, D-BAS is comparable with other software solutions presented in Figure 15. We state that D-BAS combines many of the previous mentioned advantages, but avoids their pitfalls. Thus, participants are able to state their own contributions but unfortunately it is not as unrestrictive as a forum, which is achieved with the structured approach of dialog-based argumentation. Nevertheless the use of D-BAS is more open compared to argumentation maps, because the frontend forgoes on complex representation. Furthermore, the dialog itself is very clear for the user and D-BAS assists the users with an auto completion during the entry of new statements. As base for the auto completion D-BAS takes already given statements, so that duplicates are reduced and the system will scale very well. If any duplicates occur, they can be flagged and reviewed. In addition, D-BAS is very easy to use, but obviously not very well-known in practical environments yet. Therefore the degree of ease of use slightly exceeds the average. And last of all, the evaluability is slightly better than the values of the argumentation maps. The contributions still have to be evaluated manually, but the relation between contributions are already given by the WoR and the nodes of our WoR only contain one statement each.

[1] `https://github.com/hhucn/dbas`
[2] `https://dbas.cs.uni-duesseldorf.de/`



(a) Current online discussion tools

(b) *D-BAS*

Key: Forum, Pro- and contra, Argumentation map, Matrix- and container, D-BAS

Scale: 1 ≡ very weak, 2 ≡ weak, 3 ≡ average, 4 ≡ strong, 5 ≡ very strong

Figure 15.: Kiviat diagrams for the different properties of online argumentation tools compared to *D-BAS*.

6.2. Future Research Plan

During the work on dialog-based argumentation we came in touch with many interesting research topics. As we are not able to cover all of them with one thesis, our work can be extended in future through several ways.

6.2.1. D-BAS for Social Networks

At present, we mentioned that D-BAS is available as open-source software on GitHub, free to use for everyone as online service and offers open authentication. In order to increase the scope and usage of dialog-based online argumentation, D-BAS should be able to interact with popular platforms like Amazon Alexa, Google Assistant or Facebook Messenger. In recent years, the research fields of artificial intelligence, machine learning, and cognitive computing got more important than ever before and have influences for everyone's everyday life. One common example is Google Assistant, which provides helpful information, before the user requests them.

Therefore we could use the RESTful *Application Programming Interface* (API) of D-BAS

and an artificial intelligence agent to connect D-BAS with other platforms. This agent can be understood as some module to understand natural language. Next to this agent we can define a mapping between what action should be taken by D-BAS and what users are saying respectively what the agent thinks users are saying. With this service, we are able to bring the isolated discussions of D-BAS to the places where most of the people are: Facebook and other social networks.

We think that this idea is very important for the next steps of dialog-based argumentation to increase the acceptance of novel techniques for online argumentation as well as to increase the scope of D-BAS and its potential for the rational exchange of arguments. In addition we are able to learn – with the help of machine learning – how users are really discussing online without nudging them into a "proper way of argumentation" like D-BAS does (and other online argumentation software of Chapter 1.4).

6.2.2. Recommender System for Arguments

At present, D-BAS only selects attacking arguments to present to the user. Although we have tested several ways to present supporting arguments towards the user – i.e. arguments with the same conclusion as the selected one – we had the feeling that this does not occur in everyday (offline) argumentation. Our findings from the field experiment have confirmed that the most chosen attack was an *undermine*, while a *support* was the weakest option. Furthermore D-BAS has a very naive way to choose an attacking argument. First, an attack type (undermine, undercut or rebut) is chosen, whereby the attacks itself are evenly distributed among each user. Second, an attacking argument of the specified type is chosen randomly, where D-BAS avoids to choose an argument twice when another argument was not yet chosen. Even though this approach was very successful until now, we think that the selection of the attacking argument could have great influence on the participants motivation during the discussion.

In order to provide the "best" attacking argument towards the participants, a recommender system could be implemented. A recommender system is some kind of information system that seeks to predict the rating an user would give to an item respectively to an argument [Adomavicius and Tuzhilin, 2005]. There are several ways to collect information for rating items, but the most suitable way for D-BAS would be a collaborative approach where the recommender system builds a model for the users past behaviour regarding other arguments. Based on the given information of several users as well as their histories of reactions during the dialogs, the recommender system may be able to predict the "best" attacking argument. Thereby we have different options to interpret the attribute "best". On the one hand "best" can

be the best item for D-BAS. This would be an argument, where D-BAS only has an insufficient set of information, e.g. a newly created argument. On the other hand, the best item for an user can be an item that keeps the user motivated during the dialog. This would lead to another question: How to keep users motivated during dialog-based online argumentation?

A very interesting direction of this part of future work would be to compare the pattern of argumentation resulting from the use of recommender systems in dialog-based online argumentation with regular, live argumentation. Currently we consider it likely, that they will be quite similar. However, if this turns out to be wrong, then we should adapt the way the next statement is shown to the user and her feedback options accordingly. We think that a recommender system would be a very interesting research field and very interesting not only for D-BAS, but for the entire field of online argumentation as well.

6.2.3. Arguments as a Resource

During our field experiment, outlined in Chapter 4, we noticed that many statements of a discussion were reused as premise for another argument (see Figure 10 in [Krauthoff et al., 2017b]). Our analysis of the field experiment showed that many statements would be usable for the same question, but set up for other courses or even completely different discussion topics. Also during our work for [Meter et al., 2017], we assumed that arguments respectively statements could be used as a data resource in the web to set up an argument network which helps to eradicate redundant and repetitive parts of common discussions. The first challenges for this goal were already presented in [Schneider and Meter, 2017].

Nevertheless, this task offers many more challenges which have to be tackled. One common problem is the context of statements expressed by reference words (see [Schneider and Meter, 2017]). We think that statements and arguments as a reusable resource would help the users to deliberate more efficiently and that new and empty discussion could be seeded by already ongoing arguments. This would improve the discussions quality a lot, because also expert statements could be spread in a very efficient way.

6.2.4. Improvements by Practical Experiments

D-BAS was developed for online argumentation with unskilled participants, e.g. participants who are not familiar with argumentation theory and are unskilled in handling and formulating well-formed arguments or expressing their opinion in a rational and not emotional way. Hence

we ran a first field experiment with the course of studies in computer studies, but it cannot be denied that this was a very benign setting.

In order to improve dialog-based online argumentation for large-scale online discussions and to avoid the pitfalls of other discussion and argumentation tools (see Chapter 1.4 and Figure 15a) D-BAS has to be run in a wider setting. First, the number of students, who actively participate, should be increased many times, but for this we have to deal with large numbers of positions (see Chapter 4.5.5). Second, tests in larger environments, and outside the university, would be very exciting, because dialog-based online argumentation has the potential to increase the quality of discussions not only within the university, but also for controversial decision processes, mainly at the local level as described in Chapter 1.

6.3. Closing Words

Online discussions are in the focus of research since more then three decades now. Many implementations have been provided, but today there is a huge variety of different approaches and services provided. Still, the question remains whether participants will accept newly provided approaches for large-scale online discussion and how they should look like. Even if this thesis provides a new promising approach for dialog-based online argumentation, the question is whether participants at the local level will accept the technology or do they stuck in their old behavioral patterns? In conjunction with these questions, there are many more tasks which derive from the field of acceptance of new technology and the technology acceptance model that models how users come to accept and use a technology. But these questions are far outside the scope of computer science itself.

This work also leads to another interesting questions like how to use arguments and statements as a resource and how to connect the discussions of several websites to one common backend as well as the sharing and edits of arguments given by several content aggregators. Hopefully, this research will lead to discussions with arguments in better quality, reusable arguments so that we do not have to repeat ourselves again and again and discussions with increased quality of user contributed information.

Personal Publications

Articles

Krauthoff, T., Meter, C., Baurmann, M., Betz, G., and Mauve, M. (2017a). Dialog-Based Online Argumentation. *Computer Supported Cooperative Work (CSCW). Note: This publication was rejected after the submission of the dissertation. Nevertheless we have rewritten this paper and it was published as [Krauthoff et al., 2018].*

Reviewed Conference Papers

Krauthoff, T., Meter, C., and Mauve, M. (2017b). Dialog-Based Online Argumentation: Findings from a Field Experiment. In *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence*, volume 2012, pages 85–99. CEUR-WS.

Krauthoff, T., Baurmann, M., Betz, G., and Mauve, M. (2016). Dialog-Based Online Argumentation. In *Proceedings of the 2016 Conference on Computational Models of Argument*. iOS Press. *Note: This publication was rejected after the submission of the dissertation. Nevertheless we are planing a new paper based on the reviewers feedback and we will submit the paper elsewhere again.*

Meter, C., Krauthoff, T., and Mauve, M. (2017). discuss: Embedding Dialog-Based Discussions into Websites. *International Conference on Learning and Collaboration Technologies*, pages 449–460.

Bibliography

[Adomavicius and Tuzhilin, 2005] Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749.

[Anders Analytics Limited, 2010] Anders Analytics Limited (2010). What is an Average Bounce Rate? `http://www.andersanalytics.com/blog/153-what-is-an-average-bounce-rate`. [Online]. Last access on: 2017-11-09.

[Baccaro et al., 2016] Baccaro, L., Bächtiger, A., and Deville, M. (2016). Small differences that matter: the impact of discussion modalities on deliberative outcomes. *British Journal of Political Science*, 46(3):551–566.

[Bell, 1997] Bell, P. (1997). Using Argument Representations to Make Thinking Visible for Individuals and Groups. *Proc. 2nd Int. Conf. Comput. Support Collab. Learn.*, pages 10–19.

[Bench-Capon et al., 2015] Bench-Capon, T., Atkinson, K., and Wyner, A. (2015). Using Argumentation to Structure E-Participation in Policy Making. *Transactions on Large-Scale Data-and Knowledge-Centered Systems XVIII*, 8980:1–29.

[Betz, 2010] Betz, G. (2010). *Theorie dialektischer Strukturen*. Klostermann.

[Betz, 2012] Betz, G. (2012). On degrees of justification. *Erkenntnis*, 77(2):237–272.

[Bex et al., 2013] Bex, F., Lawrence, J., Snaith, M., and Reed, C. (2013). Implementing the Argument Web. *Communications of the ACM*, 56(10):66–73.

[Bex et al., 2014] Bex, F., Snaith, M., Lawrence, J., and Reed, C. (2014). Argublogging: An application for the argument web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 25:9–15.

[Bez, 2016] Bez, R. (2016). Ecommerce Bounce Rates – 2016 Benchmarks. `https://blog.compass.co/ecommerce-bounce-rates-2016-benchmarks/`. [Online]. Last access on: 2017-11-09.

[Bimber, 2003] Bimber, B. (2003). *Information and American democracy: Technology in the evolution of political power*. Cambridge University Press.

[Brewka and Woltran, 2010] Brewka, G. and Woltran, S. (2010). Abstract dialectical frameworks. In *Proceedings of the 12th International Conference on the Principles of Knowledge Representation and Reasoning*. AAAI Press.

[Caminada and Amgoud, 2007] Caminada, M. and Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310.

[Chadwick, 2006] Chadwick, A. (2006). Internet politics: States, citizens, and new communication technologies. *New York, NY*.

[Conklin and Begeman, 1988] Conklin, J. and Begeman, M. L. (1988). gIBIS: A Hypertext Tool for Exploratory Policy Discussion. *ACM Transactions on Information Systems (TOIS)*, 6(4):303–331.

[Disqus, Inc., 2016] Disqus, Inc. (2016). Disqus - the no.1 way to build an audience on your website. [Online]. Last access on: 2017-11-09.

[Dung, 1995a] Dung, P. M. (1995a). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357.

[Dung, 1995b] Dung, P. M. (1995b). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357.

[Friess and Eilders, 2015] Friess, D. and Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3):319–339.

[Gilbert, 1996] Gilbert, M. A. (1996). Goals in Argumentation. In *International Conference on Formal and Applied Practical Reasoning*, pages 223–230. Springer.

[Gordon, 2013] Gordon, T. F. (2013). Structured Consultation with Argument Graphs. *From*

Knowledge Representation to Argumentation in AI. A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday, pages 115–133.

[Gordon, 2017] Gordon, T. F. (2017). Carneades Argumentation System. `http://carneades.github.io/`. [Online]. Last access on: 2017-11-09.

[Gordon and Karacapilidis, 1997] Gordon, T. F. and Karacapilidis, N. (1997). The zeno argumentation framework. In *Proceedings of the 6th international conference on Artificial intelligence and law*, pages 10–18. ACM.

[Gordon and Walton, 2006] Gordon, T. F. and Walton, D. (2006). 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.

[Gould and Lewis, 1985] Gould, J. D. and Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311.

[Greenup and O'Shannessy, 2017] Greenup, S. and O'Shannessy, C. (2017). Find rebuttals and promote critical thinking with rbutr. `http://rbutr.com/`. [Online]. Last access on: 2017-11-09.

[Hahn and Hornikx, 2015] Hahn, U. and Hornikx, J. (2015). A normative framework for argument quality: argumentation schemes with a bayesian foundation. *Synthese*, 193(6):1833–1873.

[Hoffman and Maier, 1961] Hoffman, L. R. and Maier, N. R. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *The Journal of Abnormal and Social Psychology*, 62(2):401.

[Iandoli et al., 2017] Iandoli, L., Quinto, I., Spada, P., Klein, M., and Calabretta, R. (2017). Supporting argumentation in online political debate: Evidence from an experiment of collective deliberation. *New Media & Society*, page 1461444817691509.

[ISO, 1997] ISO, B. (1997). 9241-1: 1997. *Ergonomic requirements for office work with visual display terminals*.

[Karacapilidis and Papadias, 2001] Karacapilidis, N. and Papadias, D. (2001). Computer sup-

ported argumentation and collaborative decision making: the hermes system. *Information systems*, 26(4):259–277.

[Karacapilidis et al., 2009] Karacapilidis, N., Tzagarakis, M., Karousos, N., Gkotsis, G., Kallistros, V., Christodoulou, S., and Mettouris, C. (2009). Tackling cognitively-complex collaboration with cope_it! *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 4(3):22–38.

[Kitchenham and Pfleeger, 1996] Kitchenham, B. and Pfleeger, S. L. (1996). Software quality: the elusive target [special issues section]. *IEEE software*, 13(1):12–21.

[Klein, 2010] Klein, M. (2010). Using Metrics to Enable Large-Scale Deliberation. In *Collective intelligence in organizations: A workshop of the ACM Group 2010 Conference*, pages 103–233.

[Klein, 2012] Klein, M. (2012). How to Harvest Collective Wisdom on Complex Problems: An Introduction to the MIT Deliberatorium. *Center for Collective Intelligence working paper*, page 15.

[Klein and Iandoli, 2008] Klein, M. and Iandoli, L. (2008). Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium.

[Krauthoff et al., 2016] Krauthoff, T., Baurmann, M., Betz, G., and Mauve, M. (2016). Dialog-Based Online Argumentation. In *Proceedings of the 2016 Conference on Computational Models of Argument*. iOS Press.

[Krauthoff et al., 2017a] Krauthoff, T., Meter, C., Baurmann, M., Betz, G., and Mauve, M. (2017a). Dialog-Based Online Argumentation. *Computer Supported Cooperative Work (CSCW)*.

[Krauthoff et al., 2018] Krauthoff, T., Meter, C., Baurmann, M., Betz, G., and Mauve, M. (2018). Dialog-Based Online Argumentation. In *Proceedings of the 2018 Conference on Computational Models of Argument*. iOS Press.

[Krauthoff et al., 2017b] Krauthoff, T., Meter, C., and Mauve, M. (2017b). Dialog-Based Online Argumentation: Findings from a Field Experiment. In *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence*, volume 2012, pages 85–99. CEUR-WS.

[Kriplean et al., 2012] Kriplean, T., Morgan, J., Freelon, D., Borning, A., and Bennett, L. (2012). Supporting Reflective Public Thought with ConsiderIt. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 265–274. ACM Press.

[Kunz, Werner; Rittel, 1970] Kunz, Werner; Rittel, H. W. J. (1970). Issues as Elements of Information Systems.

[Lampe and Resnick, 2004] Lampe, C. and Resnick, P. (2004). Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM.

[Lampe et al., 2014] Lampe, C., Zube, P., Lee, J., Park, C. H., and Johnston, E. (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly*, 31(2):317–326.

[Linn et al., 2003] Linn, M. C., Clark, D., and Slotta, J. D. (2003). WISE Design for Knowledge Integration. *Sci. Educ.*, 87:517–538.

[Mcalister et al., 2004] Mcalister, S., Ravenscroft, A., and Scanlon, E. (2004). Combining interaction and context design to support collaborative argumentation using a tool for synchronous cmc. *Journal of Computer Assisted Learning*, 20(3):194–204.

[Meter et al., 2017] Meter, C., Krauthoff, T., and Mauve, M. (2017). discuss: Embedding Dialog-Based Discussions into Websites. *International Conference on Learning and Collaboration Technologies*, pages 449–460.

[Modgil and Prakken, 2014] Modgil, S. and Prakken, H. (2014). The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5:31–62.

[Mohammed, 2016] Mohammed, D. (2016). Goals in Argumentation: A Proposal for the Analysis and Evaluation of Public Political Arguments. *Argumentation*, 30(3):221–245.

[Neugebauer, 2017] Neugebauer, D. (2017). Generating Defeasible Knowledge Bases from Real-World Argumentations using D-BAS. In *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence*, volume 2012, pages 105–110. CEUR-WS.

[Oikarinen and Reed, 1993] Oikarinen, J. and Reed, D. (1993). Internet relay chat protocol. RFC 1459, RFC Editor. `http://www.rfc-editor.org/rfc/rfc1459.txt`.

[Peyton, 2014] Peyton, J. (2014). Good, Bad, Ugly, and Average Bounce Rates | The Rocket Blog. `http://www.gorocketfuel.com/the-rocket-blog/whats-the-average-bounce-rate-in-google-analytics/`. [Online]. Last access on: 2017-11-09.

[Pollock, 1987] Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science*, 11(4):481–518.

[Reed et al., 2008] Reed, C., Devereux, J., Wells, S., and Rowe, G. (2008). AIF + : Dialogue in the Argument Interchange Format. *2*, pages 311–323.

[Reed and Rowe, 2004] Reed, C. and Rowe, G. (2004). Araucaria: software for argument analysis, diagramming and representation.

[Scheuer et al., 2010] Scheuer, O., Loll, F., Pinkwart, N., and McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *Int. J. Comput. Collab. Learn.*, 5:43–102.

[Schneider and Meter, 2017] Schneider, A. and Meter, C. (2017). Reusable Statements in Dialog-Based Argumentation Systems. In *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence*, volume 2012, pages 100–104. CEUR-WS.

[Skyrms, 2000] Skyrms, B. (2000). *Choice and Chance: An Introduction to Inductive Logic*. Wadsworth/Thomson Learning, Belmont, CA.

[Smarty, 2017] Smarty, A. (2017). What is Normal Bounce Rate? `https://www.internetmarketingninjas.com/blog/search-engine-optimization/normal-bounce-rate/`. [Online]. Last access on: 2017-11-09.

[Spada et al., 2014] Spada, P., Klein, M., Calabretta, R., Iandoli, L., and Quinto, I. (2014). A first step toward scaling-up deliberation: Optimizing large group e-deliberation using argument maps. *Under Review*, 36.

[Stephen, 1958] Stephen, T. (1958). *The Uses of Argument*. Cambridge University Press Cambridge.

[Sunstein, 2002] Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, 10(2):175–195.

[Sunstein, 2009] Sunstein, C. R. (2009). *Republic. com 2.0*. Princeton University Press.

[Suthers, 2001] Suthers, D. D. (2001). Architectures for computer supported collaborative learning. *Proc. IEEE Int. Conf. Adv. Learn. Technol.*, pages 25–28.

[Suthers, 2003] Suthers, D. D. (2003). Representational guidance for collaborative inquiry. *Arguing to Learn Confronting Cogn. Comput. Collab. Learn. Environ.*, pages 27–46.

[Suthers et al., 1995] Suthers, D. D., Weiner, A., Connelly, J., and Paolucci, M. (1995). Belvedere : Engaging Students in Critical Discussion of Science and Public. *Artif. Intell.*

[Tannen, 2012] Tannen, D. (2012). *The argument culture: Moving from debate to dialogue*. Ballantine Books.

[Twohill, 2008] Twohill, L. (2008). Offical Googel Blog: Project 10^100. `https://googleblog.blogspot.de/2008/09/project-10100.html`. [Online]. Last access on: 2017-11-09.

[Twohill, 2010] Twohill, L. (2010). $10 million for project 10^100 winners. `https://googleblog.blogspot.de/2010/09/10-million-for-project-10100-winners.html`. [Online]. Last access on: 2017-11-09.

[Van Alstyne et al., 1996] Van Alstyne, M., Brynjolfsson, E., et al. (1996). Electronic communities: global village or cyberbalkans? In *Proceedings of the 17th International Conference on Information Systems. New York: Wiley*.

[van Gelder, 2003] van Gelder, T. (2003). Enhancing Deliberation Through Computer Supported Argument Mapping. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, 2:97–115.

[Verheij, 2014] Verheij, B. (2014). Arguments and their strength: Revisiting pollock's anti-probabilistic starting points.

[Wyss and Beste, 2017] Wyss, D. and Beste, S. (2017). Artificial facilitation: Promoting collective reasoning within asynchronous discussions. *Journal of Information Technology & Politics*, 14(3):214–231.

Appendix A.

Discussion Metrics

In Chapter 1.4.1 we suggested to compare different representation styles as well as their effects on the discussion and its quality. This Appendix contains everything mentioned dimension, whereby Appendix A.1 contains the dimension about the content of the discussion. Appendix A.2 deals with the deals with the usability of the software and Appendix A.3 is about the effect and impact of the software and discussion on the participants. But first, we will sketch out ideas for different kind of experiments, which could be used to measure the metrics.

Experiments

We usually differentiate between two kind of experiments *T1* and *T2*. First one indicates field experiments with discussions like: "Do you prefer cats or dogs?", which every participant can do at home. A possible time limit could be set to two weeks. Latter indicates laboratory experiments, where the participants behaviour can be monitored, like the one executed in Chapter 3. Both experiments can be combined with either a pre- and post-questionnaire, where the changes of several valuation can be explored (*Q1*) or a post-questionnaire only (*Q2*).

A.1. Content

Completeness

- **Criteria:** A good discussions should cover a large set of feasible statements. Therefore one criteria for a good discussion is the number of statements of the current discussion.
- **Metrics:** Total number of arguments.
- **Experiment:** T1

Quality

- **Criteria:** Every discussion consists out of posts, which varies in their quality. Therefore the quality has to be determined. This could be done with several codes based on the work of [Friess and Eilders, 2015].
- **Metrics:** Manual analysis of the posts with:
  - Relevance of the issue of the current post: 1 if the relevance is given, 0 otherwise.
  - Expression of opinion: 1 if the expression is given, 0 otherwise.
  - Request of information: 1 if a requests is given, 0 otherwise.
  - Indication of information: 1 if the indication is given, 0 otherwise.
- **Experiment:** T1

Connections

- **Criteria:** Any post in a discussion should have at least one connections to another post, because a discussion cannot consists out of incoherent monologues. Thus a good post will be mentioned in at least one other post. Therefore good discussions have many relation between the posts.
- **Metrics:** Number of relations of this post.
- **Experiment:** T1

Replication

- **Criteria:** Every post of a discussion should have at least one unique statement, so that posts wont be replicates. If statements are repeating, the argumentation could no longer be free of circles. Therefore the replication-count has to be low, or rather zero.

- **Metrics:** 1 if the post is a replication of a further post in the current section/branch/part, 0 otherwise. Therefore we need a manual analysis of the posts.
- **Experiment:** T1

Relevance

- **Criteria:** Posts, which have a high relevance, contain at least one keyword of the discussion. Therefore this keyword has to be identified.
- **Metrics:** One could use the term frequency–inverse document frequency based on the extracted main keywords of the discussion. The frequency is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.
- **Experiment:** T1

Well-justified premises

- **Criteria:** The linkage between a premise and a conclusion should be very detailed, so that participants are able to understand the rationale behind an argument.
- **Metrics:** A possible rating could be 0 for no linkage 1 for a simple linkage and 2 for a full linkage.
- **Experiment:** T1

Reference

- **Criteria:** An essential element of deliberation is the kind of reference to other posts. These references can be critically or agreeing.
- **Metrics:** A possible rating could be 1, if the post is agreeing referring to another post. It is 0, if the post is substantially referring to another post, but this reference is neither critically nor agreeing or -1, if the post is agreeing critically to another post.
- **Experiment:** T1

A.2. Usability

Effectiveness

- **Criteria:** According to [ISO, 1997] this part is about the degree to which a software can be used by specified users to achieve goals with effectiveness (task completion by users) in an especial context of use.
- **Metrics:** The metrics is based on time and / or clicks until a task is finished and operating errors: Observe the participant and identify whether the error has influences for the discussion. A possible rating could be: (0) error has no influence for the discussion, (1) error has slight influences on the discussion, (2) error has large influence on the discussion.
- **Experiment:** T2. Additionally, tasks could be done, where the needed time can be measured. like "Please write a post for the given topic.".

Efficiency

- **Criteria:** According to [ISO, 1997] this part is about the degree to which a software can be used by specified users to achieve goals with efficiency (task in time) in an especial context of use.
- **Metrics:** Like effectiveness.
- **Experiment:** Like effectiveness.

Satisfaction

- **Criteria:** According to [ISO, 1997] this part is about the degree to which a software can be used by specified users to achieve goals with satisfaction (responded by user in term of experience) in an especial context of use.
- **Metrics:** Usage of a x-point Likert scale in a questionnaire, whether the participant would use the software again.
- **Experiment:** T1/T2 + Q2

User Experience

- **Criteria:** According to [ISO, 1997] this "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service". So this is a combination out of look & feel of the software as well as the usability.
- **Metrics:** Usage of a x-point Likert scale in a questionnaire, whether the participant would use the software again or the participants exceptions were met.
- **Experiment:** T1/T2 + Q2

Attractiveness

- **Criteria:** The first impressions of something is always important.
- **Metrics:** Usage of a x-point Likert scale in a questionnaire, whether the participant finds this software attractive.
- **Experiment:** T1/T2 + Q2

A.3. Effect / Impact

Understanding

- **Criteria:** Did participants understand the positions and arguments of the other participants? Can they recall them?
- **Metrics:** Just count how many arguments and statements do the participants remember after the discussion is over (this is an indirect metric - the assumption is that participants can remember things better, if they understand them).
- **Experiment:** T1/T2 + Q2

Learning Progress / Knowledge Acquisition

- **Criteria:** A discussion will be successful, if a participant learned something new or has a subjective feeling of knowledge acquisition. Otherwise, if the participants thinks, that every discussed argument was old stuff, she will be disappointed. Therefore it is important, that participants learned something new during the discussion.
- **Metrics:** Knowledge acquisition can be checked with specific questions about the content of the discussion.

- **Experiment:** T1/T2 + Q2

Opinion change

- **Criteria:** Not only the in- and output of a discussion is highly important, but also the process of opinion making. Therefore it is very interesting, why the participant has made is opinion [Baccaro et al., 2016].
- **Metrics:** The participant should be asked about her position before and after the discussion (just one question after the experiment is possible, too). She should be asked, whether the she changed/made her opinion, because other trustworthy participants hold different beliefs or her beliefs changed in view of the arguments that have been advanced [Betz, 2012].
- **Experiments:** T1/T2 + Q2

Fun Factor

- **Criteria:** With the use of "good software", the participation at a discussion should make fun or at least should not harry the participants. We hope that the participation rate is higher with software, which is not harrying.
- **Metrics:** Just a question about the individual sense of fun.
- **Experiment:** T2

Appendix B.

Questionnaire

In the following we will just list the questions from our questionnaire used in Chapter 4. As you can clearly see, many of these question are inspired by the dimensions and metrics in Appendix A.

Did you participated on the platform? Participants had the answer options of "yes" or "no". If the participants choose "no", they were asked, why they have not contributed to the discussion? Every statement had the answer options of a five point Likert-scale with "totally agree" and "totally disagree", whereby every statement but the first had a sixth option with "I did not participated."

- I was not not interested in participating.
- I did not understand the discussion.
- I did not understand D-BAS.
- I had problems with the orientation in D-BAS.
- Neither the question nor the statements of the system were enunciated.
- The idea of D-BAS did not convinced me.

Question about the tutorial: These question were available, if the participants answered "yes" in the previous filter question, whether they read the tutorial in the beginning. If they did so, they had a five point Likert-scale with "totally agree" and "totally disagree" again, whereby the statements were:

- The tutorial helped me to get along with D-BAS.
- The tutorial prepared me well for the discussion.

- The tutorial was unnecessary.
- The tutorial had to be more extensive.
- The tutorial was essential for the understanding.
- During the discussion, I noticed new elements, which were not introduced during the tutorial.

General Questions: Participants had the answer options of a five point Likert-scale with "totally agree" and "totally disagree".

- I would used D-BAS again.
- I would recommend D-BAS.
- D-BAS made fun.
- All in all I was satisfied.
- In general, I like the idea of D-BAS.

Word Pairs: Participants had the answer options of a five point Likert-scale with bipolar words on each side. The main question was: "How did you liked to talk to others in D-BAS?".

- uninteresting & interesting
- incomprehensible & comprehensible
- confusing & clear
- ineffective & effective
- complicated & easy
- in bad style & classy
- impractical & practical
- erratic & predictable
- inferior & valuable
- confusing & clear
- boring & fascinating

Quality of the discussion: Participants had the answer options of a five point Likert-scale with "totally agree" and "totally disagree". The task was to evaluate the quality of the discussion.

- D-BAS promotes a constructive discussion.
- Through D-BAS I learned more about the topic.
- I could well follow the arguments of the others.

- I was ready to be convinced by other arguments in the course of the discussion.
- In the course of the discussion, I neglected my own arguments, because I found arguments of other participants better.
- The discussion has influenced my opinion on the topic.
- The discussion has led to a higher interest in the subject.
- The contributions of the other participants made it easier for me to find new arguments.
- The quality of the arguments convinced me.
- In general, the quality of the discussion convinced me.
- The discussion with D-BAS was effective.
- It was hard for me to find arguments for my claims.
- The discussion was understandable for me.
- I found the discussion useful within the platform.
- Through the discussion I got a better understanding of the subject.

Behaviour during the discussion: Participants had the answer options of a five point Likert-scale with "totally agree" and "totally disagree". The task was to rate the discussions behavior of the other participants.

- I noticed that a few participants dominated the discussion.
- It was hard for me to defend my own position.
- It has been easy for me to find a common position with others.
- The discussion was complete for me and all the important positions and reasons were named.

How did you perceive the system? Participants had to answer with "very clear " and "very unclear" based on a five point Likert-scale.

System Participants had the answer options of a five point Likert-scale with "totally agree" and "totally disagree" on each side. The main question was: "After the tutorial, you could directly start to react on arguments and enter your own ones. How did you come up with D-BAS?".

- D-BAS was easy to use.
- I could understand D-BAS right away.
- The orientation in D-BAS did not bother me.
- The statements of the system were always clear and clear.

- The color coding of different text excerpts is noticeable to me.
- The coloring of justifications and assertions has me helped to understand the arguments of the others.
- The order of the statements on D-BAS were meaningful.
- In the course of the discussion, it was always easier for me to use the system.
- Through the structuring came a more meaningful
- Discussion as I expected.
- D-BAS responds quickly to input.

Did you used the review system? Participants had the answer options of "yes" or "no".

Demographic Data: Simple question about the gender, age and section of the study.