FULL PAPER

Making sense of user comments
Identifying journalists' requirements for
a comment analysis framework

Sinn ziehen aus Nutzerkommentaren
Anforderungen von Journalisten an ein
Software-Framework zur Kommentaranalyse

*Wiebke Loosen, Marlo Häring, Zijad Kurtanović, Lisa Merten, Julius Reimer,
Lies van Roessel & Walid Maalej*

Wiebke Loosen, Hans-Bredow-Institute for Media Research, Rothenbaumchausse 36, 20148 Hamburg, Germany; Contact: w.loosen(at)hans-bredow-institut.de

Marlo Häring, Department of Informatics, University of Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany; Contact: haering(at)informatik.uni-hamburg.de

Zijad Kurtanović, Department of Informatics, University of Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany; Contact: kurtanovic(at)informatik.uni-hamburg.de

Lisa Merten, Hans-Bredow-Institute for Media Research, Rothenbaumchausse 36, 20148 Hamburg, Germany; Contact: l.merten(at)hans-bredow-institut.de

Julius Reimer, Hans-Bredow-Institute for Media Research, Rothenbaumchausse 36, 20148 Hamburg, Germany; Contact: j.reimer(at)hans-bredow-institut.de

Lies van Roessel, Hans-Bredow-Institute for Media Research, Rothenbaumchausse 36, 20148 Hamburg, Germany; Contact: l.vanroessel(at)hans-bredow-institut.de

Walid Maalej, Department of Informatics, University of Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany; Contact: maalej(at)informatik.uni-hamburg.de

FULL PAPER

Making sense of user comments
Identifying journalists' requirements for a comment analysis framework

Sinn ziehen aus Nutzerkommentaren
Anforderungen von Journalisten an ein Software-Framework zur Kommentaranalyse

Wiebke Loosen, Marlo Häring, Zijad Kurtanović, Lisa Merten, Julius Reimer,
Lies van Roessel & Walid Maalej

Abstract: Newsrooms are still searching for ways to manage user comments because of both a desire for professional distance from their audiences and a lack of analytical tools. This paper presents findings from our exploratory, interdisciplinary study in journalism research and computer science that focuses on the algorithmic classification and clustering of user comments. In contrast to endeavours that aim at filtering out hate speech or spam, we take a more constructive approach and focus on detecting particularly useful or high-quality user contributions that can be leveraged for journalistic purposes. On the basis of a literature review and our own preliminary research on audience participation and user review analytics, we developed a mock-up of a software framework to help journalists systematically analyze user comments to this end. We then surveyed its effectiveness through two group discussions – one with comment moderators and another with editors from different editorial departments of a large German online newsroom. Features that journalists and comment moderators considered useful include the categorization of user comments in pro- and contra-arguments towards a certain topic, the automated assessment of comments' quality as well as the identification of surprising or exceptional comments and those that present new questions, arguments or viewpoints.

Keywords: User comments, journalism, automated content analysis, software requirements

Zusammenfassung: Redaktionen suchen noch immer nach Wegen im Umgang mit Nutzerkommentaren – Gründe hierfür sind sowohl der Wunsch nach professioneller Distanz zum Publikum als auch fehlende Tools, welche die Analyse von Kommentaren unterstützen können. Der vorliegende Beitrag präsentiert Befunde einer explorativen, interdisziplinären Studie aus den Bereichen Journalismusforschung und Informatik, die sich mit der algorithmischen Klassifizierung und dem Clustern von Nutzerkommentaren beschäftigt. Im Gegensatz zu ähnlichen Vorhaben, die darauf abzielen, Spam- und Hasskommentare herauszufiltern, verfolgt das Projekt einen konstruktiveren Ansatz und fokussiert darauf, Kommentare von besonderem Nutzen oder hoher Qualität zu identifizieren, die für journalistische Zwecke genutzt werden können. Auf der Basis einer Literaturstudie sowie eigener empirischer Vorarbeiten zur Publikumsbeteiligung wurde ein Mock-up für ein Software-Frame-

work entwickelt, das Journalisten bei der systematischen Analyse von Kommentaren unterstützen soll. Das Mock-up wurde in zwei Gruppendiskussionen mit Kommentarmoderatoren und Redakteuren aus verschiedenen Ressorts evaluiert. Funktionen, die als besonders nützlich erachtet wurden, umfassen u. a. die Klassifizierung von Nutzerkommentaren in Pro- und Contra-Argumente zu einem Artikelthema, die automatisierte Analyse ihrer Qualität sowie die Identifikation überraschender bzw. außergewöhnlicher Kommentare und solcher, die neue Fragen, Meinungen und Perspektiven enthalten.

Schlagwörter: Nutzerkommentare, Journalismus, automatisierte Inhaltsanalyse, Software-Anforderungen

1. Introduction

As a consequence of the larger, more general transformation of public communication in the digital age, news organizations are faced with an increasing amount of audience feedback in forums, comment sections, and social media. This trend is fundamentally changing how today's journalists and their audiences perceive, use, and manage feedback in general (Bergström & Wadbring, 2015; Heise, Loosen, Reimer, & Schmidt, 2014a; Loosen & Schmidt, 2012; Loosen & Schmidt, 2017). Even if commenting is just one of many forms in which users engage with news, it is often at the center of academic discourse about participation (e.g., Almgren & Olsson, 2015; McCluskey & Hmielowski, 2012), deliberation (e.g., Rowe, 2015; Ruiz et al., 2011), public opinion (e.g., Pantti, 2016), and community management (e.g., Binns, 2012; Braun & Gillespie, 2011). Recently, media effects research has taken an interest in user comments, looking into the relationship between them and perceptions of journalistic quality and journalistic issues (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Prochazka, Weber, & Schweiger, 2016).

We are, however, currently observing a shift in the understanding of comment sections from being "a space for a new 'deliberative democratic potential'" (Collins & Nerlich, 2015) to being a necessary evil that news media leverage to attract users (Heise et al., 2014a), or even as a threat to deliberation (Ksiazek, Peer, & Zivic, 2015).[1] It is not merely an old adage among journalists that 'one should never read user comments,' as a number of newsrooms have completely shut down the comment sections of their websites.[2] Having journalists and comment moderators manage and summarize the sheer volume of comments seems time consuming if feasible at all (Sood, Churchill, & Antin, 2012) while an automated analysis is expensive and error-prone (Scharkow, 2013). However, most (online) newsrooms still consider comment sections and other audience participation features essential (e.g., Heise et al., 2014b; Loosen, Schmidt, Heise, & Reimer, 2013a; Loosen, Schmidt, Heise, Reimer, & Scheler, 2013b; Reimer et al., 2015; Reimer, Loosen,

---

1   Also the journalistic field is increasingly covering and analyzing the phenomenon. For instance, the British *Guardian* does so in its data-driven dosssier "The dark side of Guardian comments," published in 2016: https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments

2   For examples see the list of news organizations that shut down onsite comments collected by the "Coral Project": https://community.coralproject.net/t/shutting-down-onsite-comments-a-comprehensive-list-of-all-news-organisations/347

Heise, & Schmidt, 2014). Therefore, news organizations largely restrict themselves to stemming the flow of hateful and off-topic responses (Sood et al., 2012) while lacking the analytics and tools that allow them to better manage and, even more importantly, *make sense of* user comments for both journalists and users.

Both the heterogeneity and large volume of user comments raise a number of challenges. First, it increases the moderation overhead for newsrooms. Filtering, for example, off-topic user contributions and contributions that infringe the law or code of conduct requires a great deal of effort. Second, the overwhelming amount of information makes it hard to oversee the current state of a news discussion. This makes it increasingly difficult for journalists to draw on comments for journalistic purposes (e.g., to select quality comments or comments that add new arguments, personal perspectives, or relevant new information) and to get a sense of the overall picture of opinions in a discussion thread. Therefore, the development of tools that help journalists and moderators analyze and filter user comments is considered a major challenge for news organizations (Diplaris et al., 2012). The journalistic field has recently started making major inroads by developing tools for improving commentary practices and audience engagement, such as the "Coral Project," a collaboration between the Mozilla Foundation, The New York Times, and The Washington Post (coralproject.net) or "Perspective," a collaboration between Alphabet incubator Jigsaw, The New York Times and Wikipedia (jigsaw.google.com/projects/#perspective).

This article brings together the outcomes of an exploratory study, as part of a broader interdisciplinary research project situated in the fields of journalism research and computer science. The project's goal is to design and develop need-driven tools for the automated classification, clustering and assessment of user comments in news discussions.[3] Rather than concentrating on identifying hate comments or spam (de-la-Peña-Sordo, Pastor-López, Santos, & Bringas, 2015; Ksiazek, 2016; Sood et al., 2012) we follow a more *constructive* approach that seeks to better reflect the voices of users, reduce analysis workloads, and help journalists make (journalistic) sense of user comments. In addition, we focus on journalists (instead of readers or end-users) as the primary target group for the comment analysis. Even though users and journalists may sometimes have overlapping preferences for the ways comments should be handled (we briefly elaborate on this in Section 4.1), the framework discussed in this paper should first serve the journalists' work rather than the (re)design of the front-end comments section.

The remainder of the paper introduces the foundations behind and the design of our study (Section 2 and 3), presents the study's outcomes (Section 4) and discusses the limitations and our next steps (Section 5). We conducted a case study with an iterative study design. We first developed a mock-up of a software framework for the analysis of user comments based on a literature review and our own preliminary research on audience participation in journalism and user review analysis. We then conducted two group discussions within a large German online newsroom, in which we surveyed the practices around user comments within dai-

---

3   For further information see also our project website: https://scan.informatik.uni-hamburg.de

ly working routines, discussed the mock-up, and identified additional requirements members of the newsroom may have for user comment analysis tools. Finally, we consolidated the requirements for such a framework and reiterated a mock-up that visualizes those requirements.

2. Related research: How journalism and other fields handle user comments

User comments on news websites integrate more traditional forms of media feedback such as letters-to-the-editor or conversations about the news and can count as the quintessential characteristic of the increased communicative options between journalism and (its) audiences (Bergström & Wadbring, 2015; Loosen & Schmidt, 2012). Like no other form of user activity, user comments also pose the challenge of finding a middle ground between engagement and distance representing a "tension between professional control and open participation" (Lewis, 2012) in the "complicated, even paradoxical" nature of the journalism-audience relationship (Loosen & Schmidt, 2012, p. 868). On the one hand, media organizations and individual journalists see a strategic, economically motivated and – to a certain extent – deliberative need to engage audiences by offering opportunities to participate. On the other hand, they simultaneously try to maintain a professional distance between the (participating) audience and themselves (Kramp & Loosen, 2018). Research into journalists' attitudes towards and their management of user comments reflects this tension as well as journalists' varying degrees of willingness to connect with readers by reading or responding to comments. Some journalists consider user comments "vehicles for accomplishing deliberative ideals" (Reich, 2011, p. 102) or the long-awaited opportunity to have meaningful interactions with their audience (Braun & Gillespie, 2011; Graham & Wright, 2015) that can enable a new kind of "reciprocal journalism" (Lewis, Holton, & Coddington, 2014), encouraging multiple, mutually beneficial forms of participation. For others, engaging with user comments threatens their traditional role of the expert who aims to convey news in an objective manner while maintaining a certain distance from the audience (Braun & Gillespie, 2011; Robinson, 2010). Nowadays, many journalists are engaging with user comments in one way or another, even if they do so reluctantly (Chen & Pain, 2017; Diakopoulos & Naaman, 2011; Graham & Wright, 2015; Loke, 2012; Reich, 2011). According to the Social Journalism Study, a survey of more than 3,000 journalists in eleven countries, between 27% (Germany) and 57% (Australia) of journalists monitor discussions of their work on social media and 13% (Finland) to 42% (U. S.) even engage in those discussions (Cision & Canterbury Christ Church University, 2015, p. 13).

Online commenting on the news has grown significantly in recent years (Chen & Pain, 2017). Handling user comments may, therefore, demand a considerable amount of journalists' and editors' resources (Diakopoulos & Naaman, 2011; Graham & Wright, 2015; Loosen et al., 2013a). The volume of these comments can be overwhelming (Braun & Gillespie, 2011) and many may have a toxic or uncivil tone (Reich, 2011; Chen & Pain, 2017; Loke, 2012). Therefore, it can be a tedious job at best to identify 'response-worthy' comments (Braun & Gillespie, 2011; Chen & Pain, 2017; Loke, 2012; Reich, 2011; Heise et al., 2014b; Loosen et al., 2013a;

Loosen et al., 2013b; Reimer et al., 2015). This often leaves journalists and editors with an increased workload and even the absence of any genuine journalist-reader debate – simply due to a lack of time (Graham and Wright, 2015; Reimer et al., 2015). This development is further fueled by the multiplication and differentiation of media channels (like websites, social media, blogs, apps, etc.) through which journalists are increasingly producing and distributing their content (Cision & Canterbury Christ Church, 2015; Neuberger, Langenohl, & Nuernbergk, 2014; Heise et al., 2014a; Reimer et al., 2015; Schmidt, Loosen, Heise, & Reimer, 2013). Each of these channels comes with a new audience segment (e.g., website users, Facebook and Twitter followers), a certain *commenting culture*, and varying degrees of participation. Journalists need to take these factors into account, while encouraging an increased connectivity with their audiences, resulting in an omnipresence of user feedback and other audience contributions (Kramp & Loosen, 2018).

In addition to journalism, user comments are also popular in other domains such as Amazon (Kurtanovic & Maalej, 2017) or the App Store, and coping with them is similarly challenging (Maalej, Kurtanovic, Nabil, & Stanik, 2016a). For instance, analyzing and making sense of user reviews in app stores to guide the app development process has become a popular research topic in the software engineering community (Maalej et al., 2016a). Several exploratory studies on the usefulness of reviews, automated approaches for review analytics, and even commercial tools are available that can inspire journalism even if they are related to a different domain. General exploratory studies have assessed, for example, the relationships between the customer, business, and technical characteristics of products, apps in particular (Finkelstein et al., 2014; Hoon, Vasa, Schneider, & Grundy, 2013; Pagano & Maalej, 2013). Finkelstein et al. (2014) found a solid correlation between customer rating and an app's popularity. Hoon et al. (2013) and Pagano & Maalej (2013) quantified review quality and topics discussed (such as bug reports, feature requests, or helpful app documentation). Researchers have also mined app features mentioned in the reviews and averaged out the opinions expressed about them to help app vendors prioritize their work for the next software release or update (Guzman & Maalej, 2014; Harman, Jia, & Zhang, 2012; Li, Zhang, Zhang, & Shen, 2010). In other studies tools have been developed that filter and summarize reviews based on how informative they are (Carreño & Windblath, 2013; Chen, Lin, Hoi, Xiao, & Zhang, 2014; Maalej et al., 2016a). Researchers have recently begun mining rationales and arguments from user reviews (Alkadhi, Laţa, Guzman, & Bruegge, 2017; Kurtanovic & Maalej, 2017; Xiao, Stromer-Galley, & Sándor, 2017), for instance why a certain product is preferred over others or what the criteria for adopting that product may be.

In general, review analytics tools are entering the mainstream. Prominent examples are Google Analytics and App Annie[4] that combine several mining techniques for identifying app features in the reviews and summarizing user sentiments. These commercial tools usually link the review analysis to other metrics such as user demographics, number of downloads, or sales figures.

---

4   https://www.appannie.com/

3. Research design: Identifying journalists' requirements for a software framework

As outlined above, previous studies into journalists' handling of user comments depict a need to reduce the overall workload so that journalists can make sense of comments and use them to their advantage. We suggest that automatically analyzing user comments, for example by means of machine learning techniques, could contribute to solving this problem. The aim of this study, as part of a larger project, was to identify and validate requirements for such a software framework by developing, refining, and discussing a catalogue of possible features with journalists. To this end, we developed a mock-up (i.e., an initial visual model for a software framework and its potential features), based on features identified through a literature review, that could filter and highlight user comments and offer constructive potential for journalists. We then qualitatively explored the ways journalists currently navigate the plethora of user comments and what they consider to be good or useful user contributions within a concrete newsroom case study. Finally, and most importantly, we discussed the mock-up with journalists and comment moderators to refine requirements and the mock-up itself. Figure 1 depicts the research design and process in its chronological order.

Figure 1. Overview of the research design and process
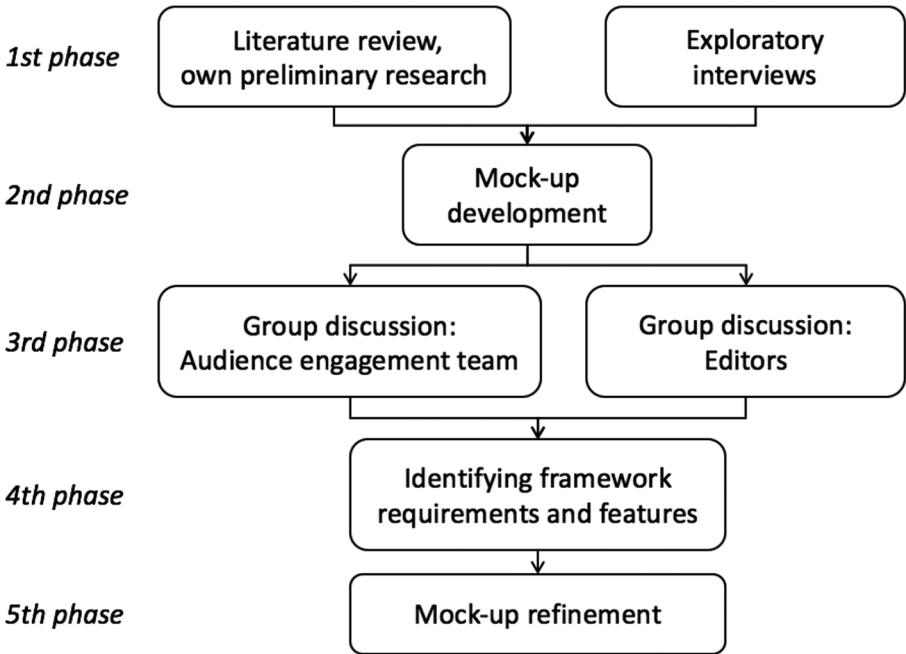
In the *first phase*, we conducted a literature study to identify journalists' preferences for a software framework and developed the initial mock-up. In the literature study we focused on previous research that addresses the ways journalists

deal with user comments in their daily practice and, in line with our aim to develop a tool that leverages user comments' constructive potential, we paid particular attention to what they consider good or useful user contributions. We also used our own preliminary work from a project on audience participation in journalism that featured interviews with journalists (n = 33) and audience members (n = 27) about their attitudes towards and experiences with user comments (cf. for the methodological approach Loosen & Schmidt, 2016; Schmidt & Loosen, 2015[5]).

In parallel with the literature study we conducted two face-to-face, exploratory, semi-structured interviews: one with the managing editor and a user forum specialist from a major German newspaper (in September 2015) and the other with the producer/editor in chief and a comment moderator of a German video and discussion platform that places particular emphasis on user engagement (in November 2015). Each interview lasted for approximately ninety minutes and had two main goals: a) to brainstorm and discuss initial ideas for the (automatic) analysis and clustering of user comments with practitioners in the field, and b) to sound out the possibility of field access to a newsroom as a case study. Both sets of interviewees were enthusiastic about the idea of developing a software system for the automatic analysis of user comments, confirmed the relevance of such an undertaking for their practice, and particularly appreciated the idea of developing a software system capable of collecting user comments across different platforms and channels. During the interviews participants highlighted the need for a means of summarizing and visualizing users' debates as well as filtering comments that can particularly inspire their journalistic work in the future.

The preliminary work and the interviews served as the backdrop for the *second phase*, i.e., the development of the mock-up, which was completed over several project meetings between our interdisciplinary team of journalism and software engineering researchers. During these meetings, we gradually formulated a list of potential features and discussed various visualization strategies. We also designed guidelines for the follow-up group discussions and defined criteria for an ideal case study, i.e., an established online news outlet with a) high popularity, audience reach and loyalty, b) broad thematic coverage, and c) a vibrant comment section. We then established contact with the deputy editor-in-chief of a suitable online newsroom, who helped with the recruitment of group discussion participants within the organization.

The *third phase* consisted of two face-to-face group discussions that each lasted approximately 120 minutes. In the first group discussion, conducted in February 2016, members of the audience engagement team, responsible for onsite comments and the newsroom's Facebook page, and the deputy editor-in-chief met with four researchers representing both disciplines from our project team. In the second group discussion, held in April 2016, we spoke to editors from five different editorial departments (politics, sports, health, human interest, technology/digital life) to address the potential differences in their experiences with user com-

---

5   In this preliminary work we investigated journalists and users of four different established media outlets; for each outlet we recruited heavy users as well as occasional contributors and lurkers to the comment sections on the website and the medium's social media profiles, respectively.

ments, differences that may arise between different topics as well as their needs and any ideas they may have had concerning certain features for their analysis. Table 1 provides an overview of the participants.

Table 1. Participants of the group discussions

First group discussion: Audience engagement team	Second group discussion: Editors
1. Head of comment section/forum 2. Social media editor 3. Managing editor with focus on user comments 4. Deputy editor-in-chief	1. Political editor 2. Science & health editor 3. Sports editor 4. Editor for technology and digital life 5. Editor human interest

Both group discussions were guided by the following three topics, aimed at further developing and refining possible features for the mock-up:

(1) *Existing practices and frameworks*: Initially we explored the current practice of handling (moderating, filtering, and reacting to) user comments within the newsroom. This helped us understand preexisting tools, guidelines and practices related to user engagement, spam, and hate speech. We also discussed the various channels available for user feedback including the homepage, email, and social media as well as a possible systematic bundling and comparison of these comments.

(2) *Challenges*: While discussing current newsroom practices of handling user comments, we were already transitioning into the second phase of the discussion that focused on the most pressing problems practitioners face when dealing with user comments. We presented the mock-up to further explore how a (semi-) automated analysis of user comments could be improved. Furthermore, we discussed features for the concrete analysis of comments such as the identification of discussion topics, arguments, and addressees.

(3) *Quality*: We also asked the participants to express their thoughts on what makes comments especially valuable or helpful, for example, in aiding journalists in their search for new topic ideas or sources. This then led to questions about types of comments and commenters. While reflecting on quality indicators and the value of user comments, we also discussed whether certain topics elicit a particularly high number of high quality comments, and if their respective comment sections could benefit, for example, from tailored features such as a barometer of public opinion or a crowdsourcing application that would allow users to rate comments.

In the *fourth phase*, we analyzed and summarized the results of the group discussions. As we were not allowed to record them, three researchers took notes during the sessions. In order to create a comprehensive transcript, these notes were combined into one document for each discussion, which was structured according to the discussion guidelines. The resulting documents were then analyzed by three project team members using a joint categorization following the three di-

mensions of our guidelines. After that, the outcome was compared to the mock-up features. One result was an overview of a) specific characteristics that could be analyzed (semi-) automatically by a software framework, b) comment aspects that were appreciated or criticized with regard to features of the mock-up, and c) illustrative quotes. In the *fifth and final phase*, we reiterated the mock-up to include the results from the group discussions. The screenshots in the appendices show the current version of the mock-up.

4. Results: Journalists' requirements for comment analysis framework

Our research led to the following results: a visual mock-up for a software framework condensing knowledge based on (our own) previous research, insights from two group discussions in which the mock-up was used as a stimulus to discuss and better understand journalists' and comment moderators' requirements, and, resulting from these, an overall list of features for a software framework for user comment analysis (see Table 2). In the following section, we structure the results along these different outcomes. First, we present the main outcomes of the literature study aimed at identifying analytical dimensions and potential features of the framework (4.1). We then introduce our findings from the group discussions in terms of the main functionalities and requirements (4.2). Part of this section is also a feature list that represents our overall findings, these are the features identified in the mock-up for a software framework for user comment analysis as well as those that came up during group discussions.

4.1 Literature study and preliminary work

Previous research (including our own) offers an abundance of material that sheds light on journalists' rationales for engaging with user comments, both for reading user contributions and for actively responding to or otherwise interacting with them, and on what they consider useful feedback or high quality comments. Thus, this body of research is a valuable resource that we tapped into during the first phase to identify potential analytical dimensions and features for a software framework adapted to journalistic requirements.

Previous studies including our own found that journalists engage with comments for the following reasons:

■ to feel closer to the user base, i.e., fostering mutually beneficial connections with their audience (Chen & Pain, 2017) allowing them to "gauge their (readers') reactions, get closer to them" (Loke, 2012, p. 244), and build relationships (Heise et al., 2014b; Reimer et al., 2015);

■ to better understand their audience's preferences and views, that is, using comments as an additional source for coming to and assessing editorial decisions (Reich, 2011), which often results in journalists and moderators developing certain everyday theories on the typologies of commenters and images about particularly active users that they recognize individually (Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Reimer et al., 2015);

343

- to keep the tone civil and increase the quality of news discussions (Chen & Pain, 2017), as comments are regarded as a form of content that requires editorial control (Diakopoulos & Naaman, 2011; Schmidt et al., 2013), with the added purpose of minimizing the potential negative effects on users' perceptions of an article's quality and the media brand (Houston, Hansen, & Nisbett, 2011; Prochazka et al., 2016; von Sikorski, 2016; von Sikorski & Hänelt, 2016);
- to maintain their gatekeeping function by steering the discussion and giving additional information and explanations by adopting the role of 'experts' (Heise et al., 2014b);
- to meet the expectations users have of journalists to engage in discussions and to build audience loyalty (Heise et al., 2014b; Reimer et al., 2015);
- to find sources and other materials, gather new story ideas and use the expertise of the audience in a crowdsourcing manner (Graham & Wright 2015; Reich, 2011; Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Reimer et al., 2015);
- to receive feedback on and criticism of their own work and use it to reflect on their writing (Diakopoulos & Naaman 2011; Graham & Wright, 2015; Reimer et al., 2015);
- to identify error reports or questions that are directed to them personally or to the newsroom in general (Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Reimer et al., 2015).

Looking at how other fields – in particular online stores – make sense of user comments, it appears that extracting product features mentioned in the comments and summarizing users' respective opinions or sentiments both help product developers reflect on and improve their products (Maalej et al., 2016a; Pagano & Maalej, 2013). Those features might correspond to topics, sections or aspects discussed in a journalistic article. Clustering similar comments, for example, concerning the informativeness of and the actors addressed in a comment (addressees such as developer, quality manager, manager, users, sales etc.) is also helpful (Guzman & Maalej, 2014; Kurtanovic & Maalej, 2017).

Moreover, during the exploratory interviews, participants repeatedly highlighted two aspects we came across in previous research. First, different platforms and social media like Facebook and Twitter stimulate different kinds of user feedback as they tend to attract different audiences (Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Reimer et al., 2015; Schmidt et al., 2013). Second, an automated comment analysis tool should reflect the diversity of arguments and the entire spectrum of a debate among users.

In addition, previous studies have shown that journalists share a common sense of what they professionally consider to be useful audience feedback or high quality comments. Appreciated comments are those that (Diakopoulos 2015a; Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; McElroy 2013; Reich 2011; Reimer et al., 2015):

- add additional information to or a new perspective, argument, or opinion on the story being commented on,

- describe personal experiences,
- help to identify potential interview partners for a certain topic,
- include links or other references leading to further information on the story's topic,
- offer ideas for further stories,
- give hints towards corruption or other illegal or dubious practices,
- report errors or contain critique addressed to the quality of an article or its author.

Interestingly, with regard to users' needs and their motivations for reading comments, research suggests that these are to a great extent similar to those of journalists. For instance, Diakopoulos and Naaman (2011) found that readers' main motivations for looking through the comments section include gaining more information, finding additional reporting on a story, and seeing the "perspectives or views from the community, see people's true feelings on a topic, gauge political response or agenda, and take the pulse of the community" (p. 137). Similarly, Ziegele (2016) found that the main cognitive objectives for readers of user comments are to gain additional information on a topic, to broaden their knowledge on a topic, and to evaluate the general attitude towards a topic (see also Heise et al., 2014a; Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Reimer et al., 2015). Although the initial aim of our project was to identify journalists' requirements for a software framework for user comment analysis, the parallels between both groups already indicate that some analytics could serve the needs of journalists and users alike. However, visualization techniques and front-end design for such a software system in particular would have to be adapted to the requirements of each group. The case for this argument has been made in research that has shown that particular features in the front-end of commentary sections can influence commenting practices (e.g., Davies & Chandler, 2012; Peacock, Scacco, & Stroud, 2017).

Even though these aspects indicate the notion that journalists (and users) have certain ideas about how to make sense of user comments and what they appreciate about them, one of the most recurring problems mentioned by journalists is that finding particularly good, useful or high quality comments is like finding a needle in a haystack: "More than one source expressed the difficulty they found in sifting particularly good or useful comments from the constant flow of user-generated content" (Braun & Gillespie, 2011, p. 387; see also Heise et al., 2014b; Reimer et al., 2015). Consequently, as Park, Sachar, Diakopoulos, and Elmqvist (2016, p. 2) point out, we are witnessing "a growing body of research in the area of computational journalism, which includes tools that are tailor-designed to suit journalistic tasks and workflows, and which take into account the professional norms and use-cases of journalists." They suggest a system that aims to help moderators identify high quality comments – utilizing the New York Times' "Picks," where user comments are selected by dedicated content moderators to show a broad range of viewpoints on a certain topic (Park et al., 2016; see also Diakopoulos, 2015b).

345

To overcome these obstacles, a software framework could help to facilitate workflow and reduce the workload both for journalists who are merely reading the comments and for those who actively engage in a discussion. Against the backdrop of the findings discussed above, it could, for instance, be useful for organizing and displaying information about which topics and actors are actually mentioned and discussed, the division of opinions represented in a comments section, or by pointing at response- or note-worthy comments such as those that directly address a journalist with a question or an error report (Kramp & Loosen, 2018).

4.2 Group discussion findings and overall feature list

The insights gained in the first step of our study were used to build a mock-up, which was then evaluated and discussed in two group discussions. We identified three main content-related dimensions for the analysis of user comments adapted to journalistic needs. These are features that allow the organization and display of information on (1) topics, mentioned actors, and those directly addressed (addressees), (2) the division of opinions and arguments, and, (3) different indicators that could help to assess the quality or note- and response-worthiness of comments. As studies have found that journalists develop certain typologies of commenters, we thought of features that would allow the system to identify different commenters based, for example, on their commenting practices.

The first mock-up version of the comment analysis framework embodied the findings of the first study phase while simultaneously functioning as a stimulus for the group discussions during which we talked about its basic functions and features to refine its functionality. To avoid redundancies, and since several mock-up features were confirmed by the group discussions as both useful and desirable, we will combine the mock-up description with the findings from the group discussions. Table 2 presents both the main outcomes of the mock-up development and the group discussions together as a list of features and analytical variables for a potential software framework that aims to help journalists analyze and filter user comments with a constructive potential and to get a sense of the discussion. The list is divided into seven main categories. The elements and features stem either from the first and second phase (simplified as mock-up development) or from the group discussions and analysis of the minutes (group discussions in the table).

Table 2. List of identified features for a software framework for user comment analysis

Category	Feature	Source
Article & Channel selection (see appendix Figure 1)	Multichannel selection: Select the articles from which the comments are analyzed (one article or multiple articles) Select sources of comments such as commentary sections on the homepage Facebook, Twitter and email (filter available in all tabs)	Mock-up development
	Show the number of comments in time/progress of a discussion Show the number of comments per commenter of a discussion Recommend whether commenting should be enabled for an article	Group discussions
Topics & Addressees (see appendix Figure 2)	Show an overview of the topics mentioned in comments Show an overview of the addressees mentioned in comments Show a set of exemplary comments that refer to mentioned topics and addressees Identify and display who is addressed in comments: e.g., the author/journalist/newsroom, a person mentioned in the article, other actors, other users Identify and display level of reference (topic-related or related to an aspect of journalistic preparation, e.g., style of writing)	Mock-up development

Discussion & Argumentation (see appendix Figure 3)	Identify and display pro- and contra-arguments in comments
	    towards the article's stance on a topic
	    towards the topics mentioned in the article
	Show an overview of pro- and contra-arguments over time
	Show an exemplary set of pro- and contra-arguments
	Identify hate speech (and highlight indicative words, phrases, or sentences if possible – to also help a moderator/journalist to better understand the system's decisions)
	Show most rated and most discussed comments
	Show top rated/most frequent arguments extracted from comments
		Mock-up development

	Identify outliers, such as non-typical comments and commenters to highlight exceptional/surprising comments Alert feature for journalists: a push notification for interesting comments or discussed topics Alert feature for community managers: a push notification for when a discussion escalates (e.g., hate speech)	Group discussions

Category	Feature	Source

Quality Indicators (see appendix Figure 4)	Show high quality comments based on different quality indicators such as:
	   Length
	   Readability
	   Information density
	   Compliance with netiquette
	   Sentiment
	   Entertainment value
	   Article reference (on-/off-topic or related to a meta aspect of the journalistic product e.g., style of writing)
	   References to other comments
	   Quality of arguments/internal coherence (e.g., do they make a coherent argument?)
	Additional sources:
	   Identify and show comments that contain a URL
	   Extract and show all URLs reported in the comments
		Mock-up development

	Estimate the originality of a comment: is a new view/perspective/aspect raised? Identify comments that serve as 'bug reports,' e.g., typos, factual errors, readers as proofreaders	Group discussions
Comparison (see appendix Figure 5)	Show a comparison of comments based on their metadata, e.g., compare different authors, sections, topics Show a comparison of comments based on their different variables, e.g., the abovementioned: information density, level of reference, etc. Show a comparison of comments based on sociodemographic data of commenters, e.g., gender or age	Mock-up development
Sociodemographics & Commenter Typologies (see appendix Figure 6)	Estimate the commenter's gender, age and level of education	Mock-up development
	Identify commenter types, e.g., expert, affected person, bot, lobbyist, troll, extremist, spammer Invite expert commenters to contribute Identify users misusing the commentary section Help to deal with misconduct (e.g., recognize spammers with multiple accounts)	Group discussions

Further Features for Commenters	Commenter features
	   Notify the commenter about the acceptance or rejection of her/his comment, including the underlying reasons
	   Recommend readers to other readers
	   Recommend articles to a reader based on his/her comments and read articles
		Group discussions

In general, the group discussions supported what we already learned from previous research. For instance, the interviewees confirmed that journalists usually feel that they should offer their readers participation options, that they are more-or-less willing to read comments and to engage with commenters, and that they think good or useful user comments could be leveraged in a journalistic way. However, making sense of user comments in this way was perceived as coming with a workload that the interviewees felt was barely manageable. As such, the

prospect that a software system could provide support was welcomed enthusiastically but accompanied by a certain incredulity towards what is technically achievable. At this stage of the discussion, however, we encouraged the participants to not restrict themselves with considerations of (technical) feasibility.

One of our project team's first ideas was to conceptualize the software framework as a '*multichannel framework*.' Given that journalistic content is produced, used, and distributed via multiple platforms including social media, and that each of these channels generate different kinds of user feedback (Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Maleej, Nayebi, Johann, & Ruhe, 2016b; Reimer et al., 2015; Schmidt et al., 2013), the software framework should be able to collect and combine user comments from different channels (Maalej & Pagano, 2011). The 'channel filter' would allow the user to sort an article's comments according to the channel in which the comments appeared (e.g., on the newsroom's own website, Facebook, Twitter, or sent by email) (see appendix Figure 1). The channel filter is available in every tab of the mock-up, so that every analysis can be done comparatively (see appendix Figure 5) or just for selected channels. During group discussions, this 'bundling function' of the mock-up was evaluated as highly useful and journalists as well as comment moderators stressed the fact that audiences and their comments vary across different channels. For instance, journalists indicated that the most valuable feedback is sent via email. However, this function also raised concerns about the extent to which a potential software framework needs to be compatible with the existing IT-infrastructure within the newsroom.

'Topics & Addressees' (see appendix Figure 2) addresses journalists' need to get an overview of what is discussed and who is mentioned and directly addressed (e.g., the newsroom or the author of an article) at a glance (Maalej et al., 2016a; Pagano & Maalej, 2013). Included in all categories of the mock-up is a feature that will highlight occurrences in comments that point to certain classification protocols (e.g., pro-/contra-argument) rendering the system's classification routines transparent enough to enable the identification of certain tropes when summarizing news discussions. An additional advantage is that, should a comment be misclassified, the user can correct the classification to improve the overall functionality of the system. With respect to the identification of topics discussed in the comments, journalists have observed differences between user comments to news articles with different themes: Comments below articles about sports or hobbies such as DIY or cookery were perceived as more civil, containing mainly positive sentiments, while health-related articles (e.g., vaccinations) or political issues (e.g., immigration) were seen as generating (a) a higher number of comments and (b) more heated debates and incivility. In addition, 'Addressees' (see Table 2) refers to a specification of the actor(s) addressed in a comment to acknowledge the fact that commenters not only mention actors that are the objects of the news article, but will sometimes address the journalists (as authors of the article), the particular newsroom, the media in general, or other commenters – often at a level beyond the topic covered and with a critical tone towards the media. Here, a software system could assist, for example, by pointing to 'response-worthy' comments such as comments that directly address the journalist with a question.

'Discussion & Argumentation' (see appendix Figure 3) refers to a feature that allows the classification of comments in terms of pro- and contra-arguments towards a certain question or topic (Guzman & Maalej, 2014). The number of overlapping comments, i.e. those that express a similar viewpoint, is indicated at the top of each comment box (boxes sorted in descending order). On the diagram, the development of the pro/contra comments' share over time is displayed. During group discussions, the 'Discussion & Argumentation' feature stimulated a number of additional ideas. For instance, participants in both group discussions remarked that opinion pieces and commentaries regularly receive more user comments than other journalistic formats. Consequently, the ability to classify user comments as pro- or contra-arguments towards a certain topic or opinion was one of the various features that were confirmed as useful by the interviewees. Furthermore, it was considered helpful for newsrooms and users alike to have tools for analyzing and illustrating the (chronological) dynamics of news discussions, that is, to show how discussions develop over time, for example in terms of their saturation, or to offer an 'alert function' that signals to moderators that 'something is escalating.' Another striking idea that came up during the group discussion with the editors was the identification of 'outliers,' i.e., non-typical comments (or commenters) that highlight something exceptional or surprising. This draws our attention to the fact that journalists are not only looking for a general overview on news discussions, but also for comments that stand for a certain unexpectedness, comments that stand out from the crowd.

The category *'Quality Indicators'* (see appendix Figure 4) represents a collection of different (quality) metrics that is meant to offer a condensed overview (e.g., readability, information density). Here, interviewees appreciated the on-/off-topic feature and also came up with the recurring issue of redundant comments, which is considered as particularly annoying for moderators. Participants from both group discussions felt that users often restate previously mentioned arguments without reading other comments or even without reading the article itself, which was considered as a hindrance to productive discussion. One of the editors suggested a feature for users: to display a box with the most common arguments directly below the article. Commenters would then only be "allowed" to write a comment if they had read through these arguments and could add something new to the discussion. Editors stated that a software framework could estimate the originality of a comment and whether or not it raises a new viewpoint or adds to the debate. To these ends, a software framework could help to identify redundant comments and provide a quick overview of what has already been said. However, these ideas also raise awareness for the following two aspects: First, they illustrate that our interviewees also thought of the mock-up as something that includes features that could be used to offer audiences an additional service by analyzing and providing an overview of news discussions, and, second, that such an extension of a software framework to the user side would have to acknowledge the effects that certain features may have on hindering or stimulating commenting practices.

'Quality Indicators' also includes features to identify, extract and display comments with related links that may contain additional sources or information on a

topic that may be useful for developing stories, or, to identify sources that users repeatedly refer to (see appendix Figure 4).

The *'Comparison'* tab allows users to directly compare the quality indicators of user comments between two different channels to indicate the differences between comments appearing on, for instance, a newsroom's own website and their Facebook page (see appendix Figure 5).

For *'Sociodemographics & Commenter Typologies'* the perspective switched from the comments themselves to their authors. As shown above, journalists and moderators regularly develop certain presumptions about particularly active users whom they, on the one hand, recognize individually, and, on the other, use as 'model users' to develop certain typologies of commenters (Heise et al., 2014b; Loosen et al., 2013a; Loosen et al., 2013b; Reimer et al., 2015). The feature, as shown in appendix Figure 6, was designed to learn more about those who are commenting, for instance in terms of gender, age, or political orientation, and to identify certain commenter types, such as the 'know-it-all' or the 'troll.' It triggered both a lot of enthusiasm, for example to expand its functionalities, and reticence, as it was considered a "nightmare for privacy protection" by one of the editors. In addition, in order to react to comments that contained questions, a feature was developed during the group discussion with the editors that invites experts to comment on these questions. Another feature should enable moderators to provide feedback to a commenter whenever a comment was rejected or their feedback was incorporated.

Generally, one difference between the groups surveyed that was quite clear from the beginning and framed both discussions as a whole, is that comment moderators are mainly concerned with the exclusion of what they deem low-quality, off-topic or even hate comments, whereas editors tend to grapple more with ways to make (journalistic) sense of user feedback in general and of user comments (on their own pieces) in particular. As a consequence, for the editors, it was much more obvious to think of user comments in terms of potentially constructive feedback that may be leveraged for journalistic purposes than it was for moderators. It seems, therefore, that it is a matter of a lack of resources rather than a lack of ideas that constrains newsrooms in their making sense of user comments.

5. Discussion

In this section, we summarize our findings, go into the limitations of our study, and elaborate on the feasibility of the framework.

5.1 Summary of findings

The literature on audience participation in journalism and particularly on journalists' attitudes and practices towards user comments highlights that most newsrooms are overwhelmed by the sheer quantity of audience feedback and are still searching for ways to manage it. Research has also shown, however, that journalists have a sense for what they deem useful contributions and hold various ideas on how they could exploit user comments for journalistic purposes. Such 'sense-making' comes with a heavy workload and requires resources that most news

organizations lack. This points to a need for software tools that will assist journalists and moderators with this task.

This observation was the starting point for our project, which is situated at the intersection of journalism research and computer science, and which aims to identify suitable machine learning algorithms, classification features, parameter configurations, and visualizations for the classification and clustering of user comments in news discussions. In this paper, we presented the findings of an initial exploratory study. Its aim was twofold: first, to develop a mock-up of a software framework that aims to systematically analyze and assess user comments, and second, to survey its usefulness and to ascertain further functional requirements through group discussions with practicing journalists and moderators in the context of a concrete newsroom case study. Here, the interdisciplinary nature of our project, combining domain-specific expertise in journalism (research) with software and requirement engineering helped us to specify the challenges that user comments pose for newsrooms and to consider solutions that address these through computational means.

Based on a literature review and our own preliminary work, we developed an initial mock-up for comment analysis. Main features include the integration of different platforms on which user comments occur, the clustering of pro- and contra-arguments, the chronological representation of discussion dynamics, and the identification of high-quality, informative user contributions. This mock-up was evaluated in two group discussions in a large German online newsroom with a wide reach and a vibrant comments section. These discussions confirmed that most features would be useful in journalistic practice and led to a list of potential dimensions, features, and requirements formulated by journalists and comment moderators for a comparable software framework. One feature, for example, that was criticized by the participants was the classification of user typologies (see appendix Figure 6). This was, however, more because of data protection concerns and less because of any doubts about its general usefulness. During the group discussion with the audience engagement team, this feature was considered to be rather helpful as a way to mitigate moderation efforts and to even identify certain lobby groups that regularly make a concerted effort to 'flood' comments sections.

Regardless of the fact that our mock-up is meant primarily to represent a software framework for journalists to make sense of user comments, it became clear during the group discussions that journalists have their own ideas about how to process user comments to facilitate audiences' overview of news discussions as well. In fact, journalists repeatedly thought of the software framework as a tool that could make news discussions more appealing and valuable to users, too, and most of their additional ideas originated from the premise of giving user comments added value for the audience. The literature has also shown that, to some degree, both groups have similar attitudes towards (what they deem to be) useful user comments. This suggests that the analytical features of our framework may also be used to process user comments as a service for users and to help build a front-end application for this target group. However, analysis features for audiences would have to look and function differently from what we have conceptualized for journalists. Another important question would be: how would the implementation of such a software framework influence the commenting practices of users?

5.2 Limitations and threats to validity

Our study was designed with a focus on in-depth qualitative understanding and exploration rather than on generalizability. Therefore, we have refrained from claiming that our findings are complete or generalizable. The goal was to identify requirements – which can later serve to form hypotheses that can be tested with a broader and larger sample. However, the general features we built into our mock-up were confirmed by previous research, which related to or derived from what journalists – and, to some extent, audiences – mentioned as useful characteristics and motivations for reading comments, such as getting an overview of viewpoints expressed and issues mentioned in comment sections. Nonetheless, the evaluation of our mock-up is only based on two group discussions with a total of nine participants within one large German general-interest newsroom. We can assume that, from our discussions with those participants and their evaluation of our initial mock-up, we have a reasonable idea of what the challenges and practices concerning user comments might be. However, other newsrooms, for instance, with a focus on specific interests, a smaller scope, or newsrooms in countries with a different 'participatory culture' may require alternative or additional features. Moreover, as with any empirical study conducted directly in the field, this study might be affected by researcher bias, particularly in the group discussions. We tried to mitigate this by keeping the discussions open and only asking questions to clarify the answers or encourage the information flow. During our discussions we first enabled participants to express their needs and preferences and only then did we discuss our mock-up and its associated requirements with them, so that they would not only refer to ideas represented in the mock-up, but would also express their own thoughts on the subject freely.

Designing and conducting a study with researchers from two different fields (journalism and software engineering research) presented additional methodological and alignment challenges. This is, for instance, illustrated by the differing definitions of terms referring to similar methodological concepts such as 'group discussions' and 'requirement workshops' (Pohl, 2010; Wiegers & Beatty, 2013). Moreover, requirements and software are usually developed in an incremental, iterative manner. In our study, we tried to enable an iterative development process while defining the methodological cornerstones and traced what features and changes resulted from which method.

5.3 Feasibility and next steps

There are many researchers, including ourselves, who develop automated or semi-automated approaches to implement the framework features presented in this paper (Maleej et al., 2016b). Yet, there are still a number of technical challenges. As journalism deals with a broad range of topics and user comments address a correspondingly broad thematic range, the first challenge is to develop techniques and tools that function efficiently and reliably on a large volume of heterogeneous natural text that is, above that, typically informal and often of low quality.

Such an approach must be highly flexible (e.g., regarding vocabulary and grammar rules) and provide support across a wide range of issues and topics.

While the feasibility of the requirements in the category "Article Selection" is straightforward, the feasibility of the other categories is rather challenging and might require a combination of various technologies including natural language processing, supervised machine-learning (ML) (which can learn certain commenting behavior or certain types of information in the comments), crowdsourcing (to label, train, or correct automated analysis) and unsupervised, deep learning approaches to deal with new trends, topics and vocabularies. For instance, topic modelling techniques (e.g., LDA) might be suitable for the category "Topics & Addressees." However, as commenters often use different names for the same entity (e.g., for the German Chancellor: Merkel, Angela, Mutti, Murksel, Angie), a native keyword based approach will not be sufficient. Therefore, a semantic and adaptive representation of the keywords to cover all meanings of the commentators' understanding of words is required.

For implementing the category "Discussion & Argumentation," syntactical ML features in addition to lexical ML features and topic modelling techniques might play an important role. For instance, discourse markers might be used to identify argumentative units (Eckle-Kohler, Kluge, & Gurevych, 2015), while topic modeling might be useful to identify significant keywords as indicators for pro- and contra-stances (Mandya, Siddhartan, & Wyner, 2016). Additionally, ML features such as text sentiments, text sophistication, and quality metrics might be useful (Somasundaran & Wiebe, 2010) as similar user arguments might have similar sentiments. While sentiments might be particularly useful to distinguish between pro- and contra-arguments, text sophistication might be used as an indicator of the presence of argumentative text. However, vague language use, implicit knowledge (Boltuzic & Snajder, 2016) as well as community bias (e.g., significantly more pro-commenters than contra-commenters) make identifying arguments a challenging endeavor.

Another prerequisite for the successful research and development of the discussed analytics features is the availability of high-quality datasets and corpora – sometimes of pre-labeled comments. The creation of such corpora is typically laborious and carried out by human coders, however it can be supported by crowdsourcing platforms such as Amazon's Mechanical Turk. Furthermore, some newsrooms already have dumps of 'labeled' comment data for quality assessment or for indicating comments that were, for instance, specifically informative or helpful. If such data is made public, it will not only help fine-tune automated approaches but assist in their evaluation as well. This task is best tackled by both the journalism and computer science research communities and in the spirit of open source and open data. We therefore plan to make the framework open source and we have already rescoped our project website into a community site that shares information on all our research with the public.[6]

---

6   https://scan.informatik.uni-hamburg.de

With more datasets available and with the recent advances in natural language processing and machine learning, we think that most if not all features can be implemented with a high degree of accuracy and performance. The next question that arises, then, is how journalists and their audiences will use and interact with software of this kind and what kind of impact it will have on their behavior and the quality of public discourse.

References

Alkadhi, R., Laţa, T., Guzman, E., & Bruegge, B. (2017). Rationale in development chat messages: an exploratory study. In *Proceedings of the 2017 IEEE/ACM 14th international conference on mining software repositories (MSR 2017)* (pp. 436–446). Piscataway: IEEE Press. https://doi.org/10.1109/MSR.2017.43

Almgren, S. M., & Olsson, T. (2015). 'Let's get them involved' … to some extent: analyzing online news participation. *Social Media + Society*, *1*(2), 1–11. https://doi.org/10.1177/2056305115621934

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. https://doi.org/https://doi.org/10.1111/jcc4.12009

Bergström, A., & Wadbring, I. (2015). Beneficial yet crappy: journalists and audiences on obstacles and opportunities in reader comments. *European Journal of Communication*, *30*(2), 137–151. https://doi.org/https://doi.org/10.1177/0267323114559378

Binns, A. (2012). Don't feed the trolls! Managing troublemakers in magazines' online communities. *Journalism Practice*, *6*(4), 547–562. https://doi.org/https://doi.org/10.1080/17512786.2011.648988

Boltuzic, F., & Šnajder, J. (2016). Fill the gap! Analyzing implicit premises between claims from online debates. In *Proceedings of the 3rd workshop on argument mining* (pp. 124–133). Stroudsburg: ACL. Retrieved from http://www.aclweb.org/anthology/W/W16/W16-28.pdf

Braun, J., & Gillespie, T. (2011). Hosting the public discourse, hosting the public: When online news and social media converge. *Journalism Practice*, *5*(4), 383–398. https://doi.org/https://doi.org/10.1080/17512786.2011.557560

Carreño, L. V. G., & Winbladh, K. (2013). Analysis of user comments: an approach for software requirements evolution. In Notkin, D., Cheng, B. H. C., Pohl, K. (Eds.), *Proceedings of the 35th international conference on software engineering (ICSE 2013)* (pp. 582–591). Piscataway: IEEE. https://doi.org/10.1109/ICSE.2013.6606604

Chen, G., & Pain, P. (2017). Normalizing online comments. *Journalism Practice*, *11*(7), 876–892. https://doi.org/10.1080/17512786.2016.1205954

Chen, N., Lin, J., Hoi, S. C., Xiao, X., & Zhang, B. (2014). AR-miner: mining informative reviews for developers from mobile app marketplace. In Jalote, P., Briand, L., & van der Hoek, A. (Eds.), *Proceedings of the 36th international conference on software engineering (ICSE 2014)* (pp. 767–778). New York: ACM. https://doi.org/10.1145/2568225.2568263

Cision, & Canterbury Christ Church University. (2015). *2015 global social journalism study*. Retrieved from http://cision-wp-files.s3.amazonaws.com/de/wp-content/uploads/2015/09/Cision-SJS-2014-15-International.pdf

Collins, L., & Nerlich, B. (2015). Examining user comments for deliberative democracy: a corpus-driven analysis of the climate change debate online. *Environmental Communication*, *9*(2), 189–207. https://doi.org/10.1080/17524032.2014.981560

Davies, T., & Chandler, R. (2012). Online deliberation design: choices, criteria, and evidence. In T. Nabatchi, J. Gastil, M. Leighninger, & G. Weiksner (Eds.), *Democracy in motion: evaluating the practice and impact of deliberative civic engagement* (pp. 103–131). Oxford: Oxford University Press.

de-la-Peña-Sordo, J., Pastor-López, I., Santos, I., & Bringas, P. G. (2015). Using compression models for filtering troll comments. In *Proceedings of the 2015 IEEE 10th conference on industrial electronics and applications (ICIEA)* (pp. 655–660). Piscataway: IEEE. https://doi.org/10.1109/ICIEA.2015.7334191

Diakopoulos, N. (2015a). Picking the NYT Picks: editorial criteria and automation in the curation of online news comments. *#ISOJ*, *5*(1), 147–166. Retrieved from https://isoj-journal.wordpress.com/2015/04/15/picking-the-nyt-picks-editorial-criteria-and-automation-in-the-curation-of-online-news-comments/

Diakopoulos, N. A. (2015b). The editor's eye: curation and comment relevance on the new york times. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing (CSCW '15)* (pp. 1153–1157). New York: ACM. https://doi.org/10.1145/2675133.2675160

Diakopoulos, N., & Naaman, M. (2011). Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 conference on computer supported cooperative work (CSCW '11)* (pp. 133–142). New York: ACM. https://doi.org/10.1145/1958824.1958844

Diplaris, S., Papadopoulos, S., Kompatsiaris, I., Heise, N., Spangenberg, J., Newman, N., & Hacid, H. (2012). "Making sense of it all": an attempt to aid journalists in analysing and filtering user generated content. In *WWW 2012 companion. Proceedings of the 21st international conference on world wide web* (pp. 1241–1246). New York: ACM. https://doi.org/10.1145/2187980.2188267

Eckle-Kohler, J., Kluge, R., & Gurevych, I. (2015). On the role of discourse markers for discriminating claims and premises in argumentative discourse. In *Proceedings of the 2015 conference on empirical methods in natural language processing (EMNLP '15)* (pp. 2249–2255). Stroudsburg: ACL.

Finkelstein, A., Harman, M., Jia, Y., Martin, W., Sarro, F., & Zhang, Y. (2014). *App store analysis: mining app stores for relationships between customer, business and technical characteristics*. (Research note RN/14/10). London: University College London, Department of Computer Science. Retrieved from http://www.cs.ucl.ac.uk/fileadmin/UCL-CS/research/RN_14_10.pdf

Graham, T., & Wright, S. (2015). A tale of two stories from 'below the line' comment fields at the guardian. *The International Journal of Press/Politics*, *20*(3), 317–338. https://doi.org/10.1177/1940161215581926

Guzman, E., & Maalej, W. (2014). How do users like this feature? A fine grained sentiment analysis of app reviews. In *Proceedings of the 22nd IEEE international requirements engineering conference (RE 2014)* (pp. 153–162).

Harman, M., Jia, Y., & Zhang, Y. (2012). App store mining and analysis: msr for app stores. In Gorschek, T., & Lutz, R. (Eds.), *Proceedings of the 9th IEEE working confer-*

ence on mining software repositories (MSR 2012) (pp. 108–111). Piscataway: IEEE. https://doi.org/10.1109/MSR.2012.6224306

Heise, N., Loosen, W., Reimer, J., & Schmidt, J.-H. (2014a). Including the audience. Comparing the attitudes and expectations of journalists and users towards participation in German TV news journalism. *Journalism Studies*, *15*(4), 411–430. https://doi.org/ 10.1080/1461670X.2013.831232

Heise, N., Reimer, J., Loosen, W., Schmidt, J.-H., Heller, C., & Quader, A. (2014b). *Publikumsinklusion bei der Süddeutschen Zeitung. Fallstudienbericht aus dem DFG-Projekt "Die (Wieder-)Entdeckung des Publikums"* [Audience inclusion at the Süddeutsche Zeitung. Case study report from the DFG-funded project "(Re-)Discovering the Audience"] (Hans Bredow Institute working papers no. 31). Hamburg: Hans-Bredow-Institut für Medienforschung an der Universität Hamburg. Retrieved from http://www.hans-bredow-institut.de/webfm_send/1050

Hoon, L., Vasa, R., Schneider, J. G., & Grundy, J. (2013) *An analysis of the mobile app review landscape: Trends and implications*. Melbourne: Swinburne University of Technology.

Houston, J. B., Hansen, G. J., & Nisbett, G. S. (2011). Influence of user comments on perceptions of media bias and third-person effect in online news. *Electronic News*, *5*(2), 79–92. https://doi.org/10.1177/1931243111407618

Kramp, L., & Loosen, W. (2018). The transformation of journalism: from changing newsroom cultures to a new communicative orientation? In A. Hepp, A. Breiter, & U. Hasebrink (Eds.), *Communicative figurations: transforming communications in times of deep mediatization* (pp. 205–239). London: Palgrave Macmillan.

Ksiazek, T. B. (2016). Commenting on the news. Explaining the degree and quality of user comments on news websites. *Journalism Studies*, online first. https://doi.org/10.1080/ 1461670X.2016.1209977

Ksiazek, T. B., Peer, L., & Zivic, A. (2015). Discussing the news. Civility and hostility in user comments. *Digital Journalism*, *3*(6), 850–870. https://doi.org/10.1080/21670811. 2014.972079

Kurtanovic, Z., & Maalej, W. (2017). Mining user rationale from software reviews. In *Proceedings of the 25th IEEE international conference on requirements engineering (RE 2017)* (pp. 61–70). Piscataway: IEEE. https://doi.org/10.1109/RE.2017.86

Lewis, S. C. (2012). The tension between professional control and open participation: journalism and its boundaries. *Information, Communication & Society*, *15*(6), 836–866. https://doi.org/10.1080/1369118X.2012.674150

Lewis, S. C., Holton, A. E., & Coddington, M. (2014). Reciprocal journalism: a concept of mutual exchange between journalists and audiences. *Journalism Practice*, *8*(2), 229–241. https://doi.org/10.1080/17512786.2013.859840

Li, H., Zhang, L., Zhang, L., & Shen, J. (2010). A user satisfaction analysis approach for software evolution. In *Proceedings of the IEEE international conference on progress in informatics and computing (PIC 2010)* (pp. 1093–1097). Piscataway: IEEE. https://doi. org/10.1109/PIC.2010.5687999

Loke, J. (2012). Old turf, new neighbors: journalists' perspectives on their new shared space. *Journalism Practice*, *6*(2), 233–249. https://doi.org/10.1080/17512786.2011.61 6649

Loosen, W., & Schmidt, J.-H. (2012). (Re-)Discovering the audience. The relationship between journalism and audience in networked digital media. In: *Information, Communication & Society*, *15*(6), 867–887. https://doi.org/10.1080/1369118X.2012.665467

Loosen, W., & Schmidt, J.-H. (2016). Multi-method approaches. In T. Witschge, C. W. Anderson, D. Domingo, & A. Hermida (Eds.), *The SAGE handbook of digital journalism* (pp. 562–575). London et al.: Sage.

Loosen, W., & Schmidt, J.-H. (2017). Between proximity and distance: including the audience in journalism (research). In B. Franklin & S. Eldridge II (Eds.), *The Routledge companion to digital journalism studies* (pp. 354–363). Abingdon: Routledge.

Loosen, W., Schmidt, J.-H., Heise, N., & Reimer, J. (2013a). *Publikumsinklusion bei einem ARD-Polittalk. Fallstudienbericht aus dem DFG-Projekt "Die (Wieder-)Entdeckung des Publikums"* [Audience inclusion at an ARD political TV talk. Case study report from the DFG-funded project "(Re-)Discovering the Audience"] (Hans Bredow Institute working paper No. 28). Hamburg: Hans-Bredow-Institut für Medienforschung an der Universität Hamburg. Retrieved from http://www.hans-bredow-institut.de/webfm_send/739

Loosen, W., Schmidt, J.-H., Heise, N., Reimer, J., & Scheler, M. (2013b). *Publikumsinklusion bei der Tagesschau. Fallstudienbericht aus dem DFG-Projekt "Die (Wieder-)Entdeckung des Publikums"* [Audience inclusion at the Tagesschau. Case study report from the DFG-funded project "(Re-)Discovering the Audience"] (Hans Bredow Institute working paper No. 26). Hamburg: Hans-Bredow-Institut für Medienforschung an der Universität Hamburg. Retrieved from http://www.hans-bredow-institut.de/webfm_send/709

Maalej, W., Kurtanovic, Z., Nabil, H., & Stanik, C. (2016a). On the automatic classification of app reviews. *Requirements Engineering, 21*(3), 311–331. https://doi.org/10.1007/s00766-016-0251-9

Maalej, W., Nayebi, M., Johann, T., & Ruhe, G. (2016b). Toward data-driven requirements engineering. *IEEE Software, 33*(1), 48–54. https://doi.org/10.1109/MS.2015.153

Maalej, W., & Pagano, D. (2011). On the socialness of software. In *Proceedings of the IEEE ninth international conference on dependable, autonomic and secure computing (DASC 2011)* (pp. 864–871). Piscataway: IEEE. https://doi.org/10.1109/DASC.2011.146

Mandya, A., Siddharthan, A., & Wyner, A. (2016). Scrutable feature sets for stance classification. In *Proceedings of the 3rd workshop on argument mining* (pp. 60–69). Stroudsburg: ACL. Retrieved from Retrieved from http://www.aclweb.org/anthology/W/W16/W16-28.pdf

McCluskey, M., & Hmielowski, J. (2012). Opinion expression during social conflict: comparing online reader comments and letters to the editor. *Journalism*, *13*(3), 303–319. https://doi.org/10.1177/1464884911421696

McElroy, K. (2013). Where old (gatekeepers) meets new (media). Herding reader comments into print. *Journalism Practice*, *7*(6), 755–771. https://doi.org/10.1080/17512786.2013.774117

Neuberger, C., Langenohl, S., & Nuernbergk, C. (2014). *Social Media und Journalismus* [Social media and journalism] (LfM Dokumentationen Band 50). Düsseldorf: Landesanstalt für Medien Nordrhein-Westfalen (LfM).

Pagano, D., & Maalej, W. (2013). User feedback in the app store: an empirical study. In *Proceedings of the 21st IEEE international requirements engineering conference (RE 2013)* (pp. 125–134). Piscataway: IEEE. https://doi.org/10.1109/RE.2013.6636712

Park, D., Sachar, S., Diakopoulos, N., & Elmqvist, N. (2016). Supporting comment moderators in identifying high quality online news comments. In *Proceedings of the 2016 CHI conference on human factors in computing systems (CHI '16)* (pp. 1114–1125). New York: ACM. https://doi.org/10.1145/2858036.2858389

Pantti, M. (2016). "Despicable, disgusting, repulsive!!!" Public emotions and moralities in online discussions about violence towards refugees. *Javnost – The Public*, 23(4), 363–381. https://doi.org/10.1080/13183222.2016.1248099

Peacock, C., Scacco, J. M., & Stroud, N. J. (2017). The deliberative influence of comment section structure. *Journalism.* Advance online publication. https://doi.org/10.1177/1464884917711791

Pohl, K. (2010). *Requirements engineering: fundamentals, principles, and techniques.* Berlin, Heidelberg: Springer.

Prochazka, F., Weber, P., & Schweiger, W. (2016). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism Studies.* Advance online publication. https://doi.org/10.1080/1461670X.2016.1161497

Reich, Z. (2011). User comments: the transformation of participatory space. In J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, Z. Reich, & M. Vujnovic (Eds.), *Participatory journalism. Guarding open gates at online newspapers* (pp. 96–117). Chichester: Wiley-Blackwell.

Reimer, J., Heise, N., Loosen, W., Schmidt, J.-H., Klein, J., Attrodt, A., & Quader, A. (2015). *Publikumsinklusion beim Freitag. Fallstudienbericht aus dem DFG-Projekt "Die (Wieder-)Entdeckung des Publikums"* [Audience inclusion at Der Freitag. Case study report from the DFG-funded project "(Re-)Discovering the Audience"] (Hans Bredow Institute working papers no. 36). Hamburg: Hans-Bredow-Institut für Medienforschung an der Universität Hamburg. Retrieved from http://www.hans-bredow-institut.de/webfm_send/1115

Reimer, J., Loosen, W., Heise, N., & Schmidt, J.-H. (2014). Erwartungen und Erwartungserwartungen an Publikumsbeteiligung bei der Tagesschau [Expectations and expected expectations towards audience participation at the Tagesschau]. In B. Stark, O. Quiring, & N. Jackob (Eds.), *Von der Gutenberg-Galaxis zur Google-Galaxis. Alte und neue Grenzvermessungen nach 50 Jahren DGPuK* [From the Gutenberg galaxy to the Google galaxy. Old and new boundary surveys after 50 years of DGPuK] (pp. 135–150). Konstanz: UVK.

Robinson, S. (2010). Traditionalists vs. convergers. Textual privilege, boundary work, and the journalist-audience relationship in the commenting policies of online news sites. *Convergence*, 16(1), 125–143. https://doi.org/10.1177/1354856509347719

Rowe, I. (2015). Deliberation 2.0: comparing the deliberative quality of online news user comments across platforms. *Journal of Broadcasting & Electronic Media*, 59(4), 539–555. https://doi.org/10.1080/08838151.2015.1093482

Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics*, 16(4), 463–487. https://doi.org/10.1177/1940161211415849

Scharkow, M. (2013). Thematic content analysis using supervised machine learning: an empirical evaluation using German online news. *Quality & Quantity*, 47(2), 761–773. https://doi.org/10.1007/s11135-011-9545-7

Schmidt, J.-H., Loosen, W., Heise, N., & Reimer, J. (2013). Journalism and participatory practices – blurring or reinforcement of boundaries between journalism and audiences? *Recherches en Communication*, *39*, 91–109.

Schmidt, J.-H., & Loosen, W. (2015). Both sides of the story. Assessing audience participation in journalism through the concept of inclusion distance. *Digital Journalism*, *3*(2), 259–278. https://doi.org/10.1080/21670811.2014.930243

Somasundaran, S., & Wiebe, J. (2010). Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (CAAGET '10)* (pp. 116–124). Stroudsburg: ACL.

Sood, S. O., Churchill, E. F., & Antin, J. (2012) Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, *63*, 270–285. https://doi.org/10.1002/asi.21690

von Sikorski, C. (2016). The effects of reader comments on the perception of personalized scandals: exploring the roles of comment valence and commenters' social status. *International Journal of Communication, 10*, 4480–4501. Retrieved from http://ijoc.org/index.php/ijoc/article/view/5748

von Sikorski, C., & Hänelt, M. (2016). Scandal 2.0. How valenced reader comments affect recipients' perception of scandalized individuals and the journalistic quality of online news. *Journalism & Mass Communication Quarterly*. Advance online publication. https://doi.org/10.1177/1077699016628822

Wiegers, K., & Beatty, J. (2013). *Software requirements*. New York: Pearson Education.

Xiao, L., Stromer-Galley, J., & Sándor, Á. (2017). Toward the automated detection of individuals' rationales in large-scale online open participative activities: a conceptual framework. *Group Decision and Negotiation*, *26*(5), 891–910. https://doi.org/10.1007/s10726-016-9516-4

Ziegele, M. (2016). *Nutzerkommentare als Anschlusskommunikation: Theorie und qualitative Analyse des Diskussionswerts von Online-Nachrichten* [User comments as follow-up communication: theory and quantitative analysis of the discussion value of online news]. Wiesbaden: Springer VS.

Appendix

The following figures display screenshots of the mock-up. All text has been translated from German to English.

**Figure 1.** Article and Channel Selection
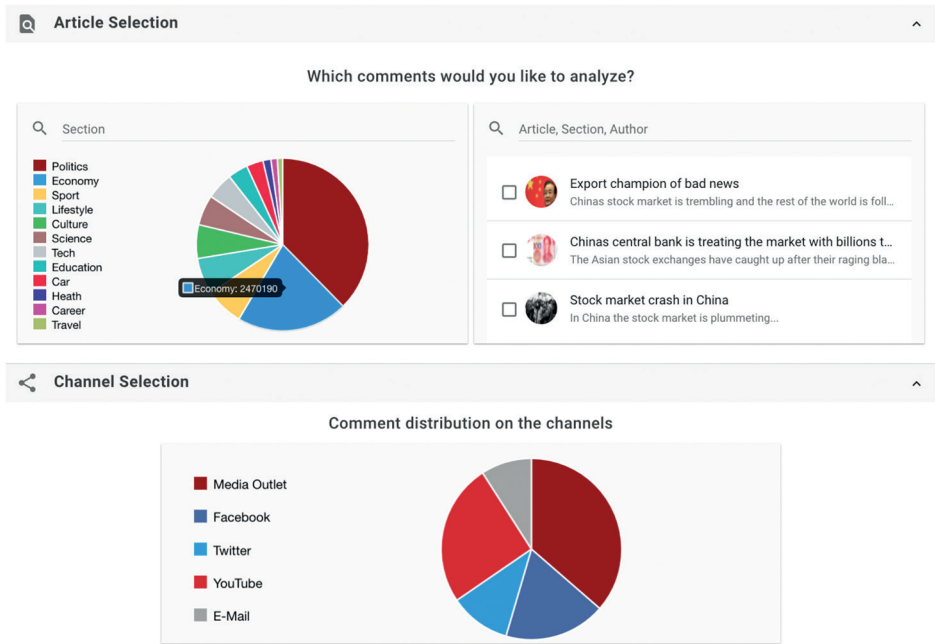
Note. The editor/journalist can select the section/articles (top) and channels (bottom) of which s/he wants to analyze the comments. In the channel selection area, the articles' comments can be selected based on the channel in which the comments were written (e.g., on the outlet's own website, Facebook, Twitter, or sent by email).

Figure 2. Topics and Addressees
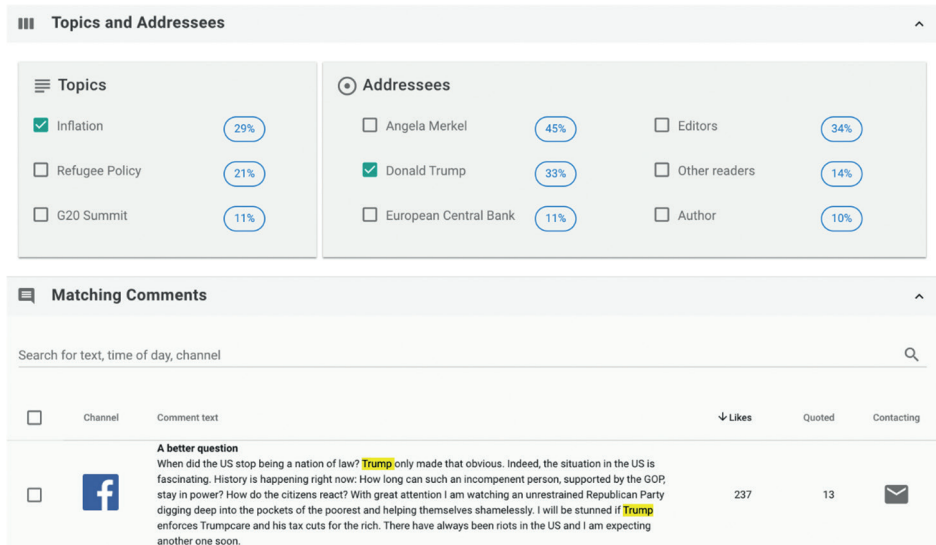
Note. This area (Topics & Addressees) shows popular topics and addressees with the share of comments mentioning them. A topic/addressee combination can be selected to filter a sample of comments matching the selection. The occurrences are highlighted in the comments.

**Figure 3.** Discussion and Argumentation
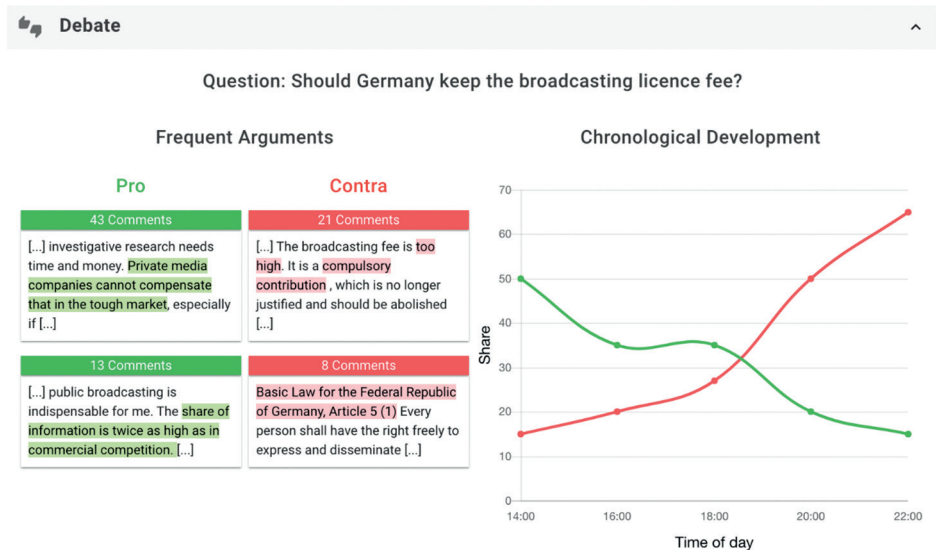
Note. In this tab the comments are classified in terms of pro and contra towards a certain question or discussion. The number of overlapping comments, i.e. those that express a similar point of view, is indicated at the top of each comment box (boxes sorted in descending order). On the right side the development of the pro/contra comments' share over time is shown.

Figure 4. Quality Indicators
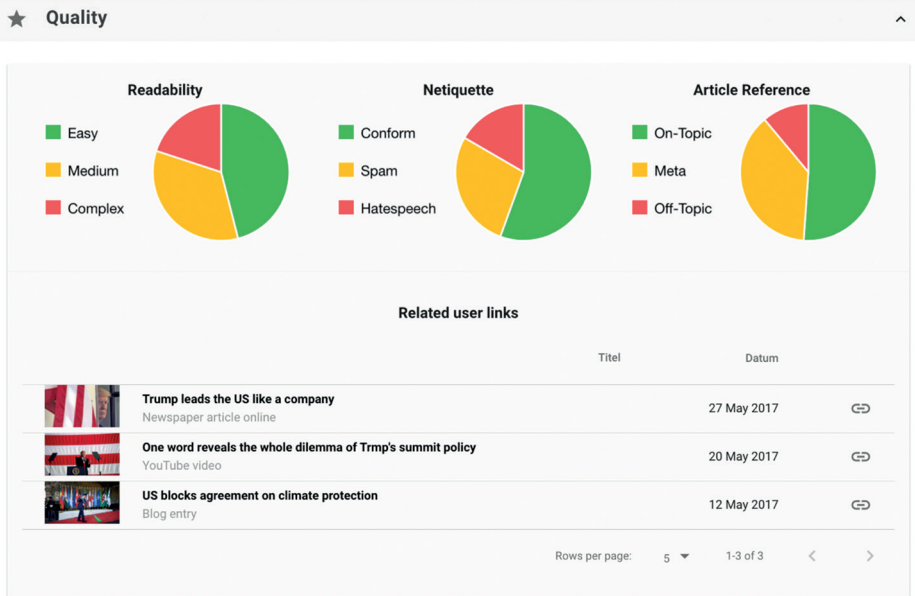
Note. In this tab comments are analyzed according to different quality metrics: readability, netiquette, article reference. The list also displays relevant links.

Figure 5. Channel Comparison
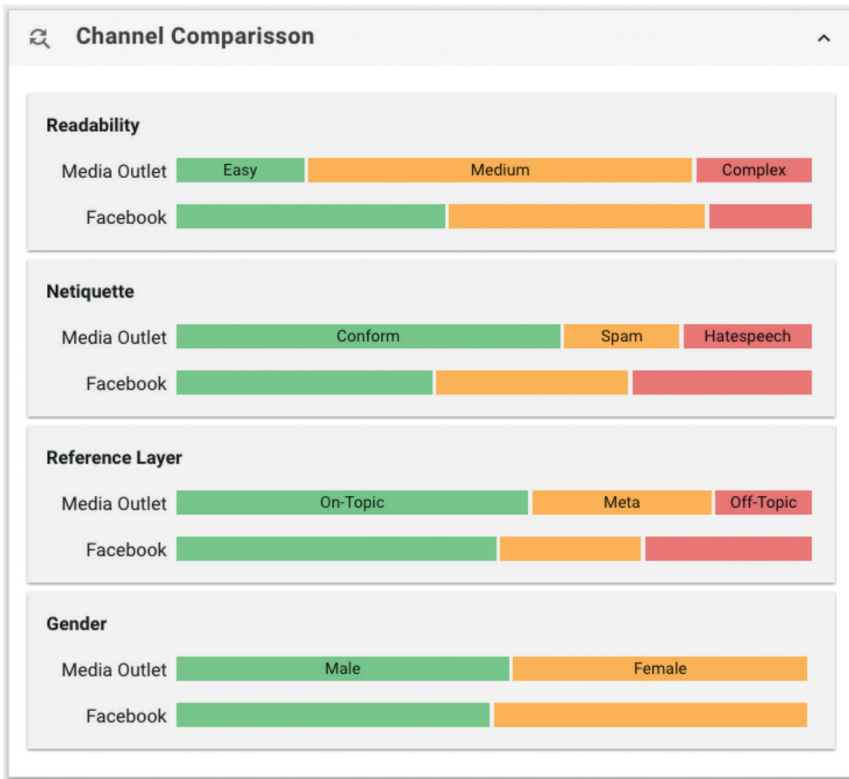
Note. This section compares two different sources of comments in terms of readability, netiquette, reference layer, and gender of commenters.

Figure 6. Sociodemographics and Commenter Typologies
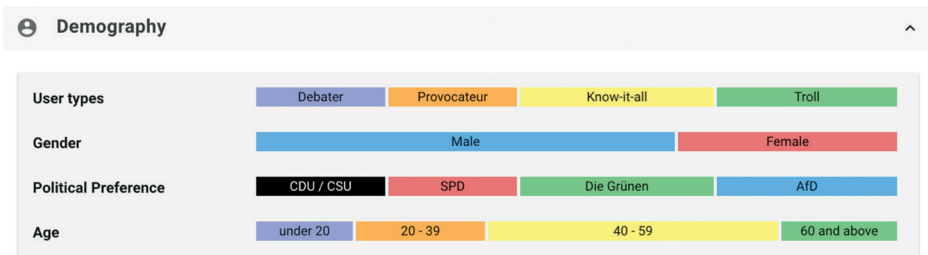
Note. In this tab the share of the demographic metrics, extracted from the comments, are shown. Together with data drawn from the comments analyses, these data may be used to identify certain types of commenters.