AMCIS 2022 Proceedings SIG HIC - Human Computer Interaction

Aug 10th, 12:00 AM

Making Filter Bubbles Understandable

Stefan Hirschmeier
University of Cologne, hirschmeier@wim.uni-koeln.de

Tim Alvaro Ockenga
Cologne Institute for Information Systems (CIIS), ockenga@wim.uni-koeln.de

Making Filter Bubbles Understandable

Completed Research

Stefan Hirschmeier
University of Cologne
hirschmeier@wim.uni-koeln.de

Tim Alvaro Ockenga
University of Cologne
ockenga@wim.uni-koeln.de

Abstract

Recommender systems tend to create filter bubbles and, as a consequence, lower diversity exposure, often with the user not being aware of it. The biased preselection of content by recommender systems has called for approaches to deal with exposure diversity, such as giving users control over their filter bubble. We analyze how to make filter bubbles understandable and controllable by using interactive word clouds, following the idea of building trust in the system. On the basis of several prototypes, we performed explorative research on how to design word clouds for the controllability of filter bubbles. Our findings can inform designers of interactive filter bubbles in personalized offers of broadcasters, publishers, and media houses.

Keywords

Filter bubble, word cloud, recommender systems, understandability, trust.

Introduction

A plethora of algorithms increasingly permeates our daily lives and their importance for society is widely accepted (Just and Latzer, 2016). Many application areas like search engines, news aggregators, or social networks use recommender systems and increasingly personalize their content streams (Agichtein, Brill and Dumais, 2006; Das, Datar, Garg and Rajaram, 2007; Hannak et al., 2013). Recommender systems help to lower search costs (Ricci, 2011), which is a benefit for both the provider and the consumer, but they also tend to produce filter bubbles and lower exposure diversity, sometimes without the user being aware of it. As such, recommender systems are seen critical for the process of building a free and individual opinion and have damaging potential to society (Resnick et al., 2013; Bozdag and van den Hoven, 2015; Pöchhacker, Burkhardt, Geipel and Passoth, 2017).

Pariser, who coined the term filter bubble, described it as a state where non-transparent algorithms create a “unique universe of information” (Pariser, 2011, p. 10) and isolate users from a diversity of viewpoints and content. As a consequence, “the epistemic quality of information and diversity of perspectives will suffer and the civic discourse will be eroded” (Bozdag and van den Hoven, 2015, p. 249).

Yet, for the success of ecosystems for content consumption, personalization is a key factor in the satisfaction of users of content services (Liang, Lai and Ku, 2006). This means, that on the one hand, people need recommender systems to manage information overload; on the other hand, they can lead to people not being exposed to a variety of content and having a biased perception of reality. Thus, it is important to develop innovative solutions that both help users to identify relevant content in an infinite universe of information and avoid the risks associated with filter bubbles.

One approach to deal with filter bubbles is to try to avoid them by putting algorithms in place that take care of diversity in the background (Hirschmeier and Beule, 2018). Another approach is more permissive by allowing filter bubbles but giving the users control over them (e.g., Munson and Resnick, 2013; Resnick et al., 2013; Nagulendra and Vassileva, 2014, 2016; Bozdag and van den Hoven, 2015). For the latter, it is important to understand how visualizations should be designed to allow users to interact with their filter bubble[1].

When we allow users to control their filter bubbles, and, as a consequence, their personal diversity exposure, still unbalanced consumption may take place – but in the hands of the user. User-controlled diversity exposure makes a difference to biased exposure without control and awareness. Giving users control follows the idea to keep the preselection of diversity in line with the self-selection of diversity again and build trust. Users rarely go for completely unbiased consumption and have ever since skipped articles in newspapers, had their favorite magazine, or switched radio channels when they were not interested. Nevertheless, it is important to give back control so that algorithmic pre-selection does not unintentionally limit the user's self-selection. Consequently, to address diversity issues in designing digital artifacts, we want to focus on systems that are understandable so that users may become aware of their filter bubble, and, as a consequence, can influence the system's output. More control over a recommender system has shown a better user experience (Knijnenburg et al., 2012) and increased confidence and trust (Tintarev and Masthoff, 2012; di Sciascio, Sabol and Veas, 2016).

Choosing from the universe of all possible visualizations, we focus on word clouds, as several studies propose word clouds as an effective implementation for explaining recommendations (De Nart and Tasso, 2014; Tsai and Brusilovsky, 2017, 2019). However, little is known about the design of interactive word clouds concerning the optimization of their understandability. Hence, we pose the following research question:

How to design interactive word clouds to increase the understandability of filter bubbles?

The remainder of this work is structured as follows. In section two, we present related work. Then we explain the methodological approach for our study in section three. In section four, we present the results of our study and discuss them in section five. Finally, we summarize our work, explain the limitations of the study, and give suggestions for future research.

Related Work

In this section, we first give a theoretical background on recommender systems and filter bubbles. Then we specify the role of controllability in solving the problem of lacking diversity. Finally, we introduce existing studies that are closely related to this work.

Recommender Systems and Filter Bubbles

Recommender systems permeate many areas of our daily lives (Lu et al., 2015), such as music recommendations (e.g., Slaney and White, 2006; Martin, 2013), movie recommendations (e.g., Diao et al., 2014; Chen, Teng and Chang, 2015), and news recommendations (e.g., Turcotte et al., 2015; Bernstein et al., 2020). Burke describes a recommender system as "[...] any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options" (Burke, 2002, p. 1). These systems find their raison d'être in a phenomenon called *information overload*, which refers to a state in which people are overwhelmed because they have too many choices (Borchers, Herlocker, Konstan and Riedl, 1998). Recommender systems support users in overcoming this state by pre-filtering and suggesting a subset of the most relevant options. Depending on how recommender systems filter the content, they are typically assigned to one of three classes. *Content-based filtering (CBF)* recommender systems suggest items similar to the ones the user preferred in the past. *Collaborative filtering (CF)* recommender systems suggest items that other users with similar tastes and preferences. *Hybrid approach* recommender systems combine both collaborative and content-based approaches.

Recommender systems often suffer from a *serendipity problem* (Lops, 2011), which contributes to the phenomenon of filter bubbles and a lack of diversity exposure. This is a problem from at least two

[1] Actually, we do not focus on the filter bubble itself, but on the user profile that leads to the filter bubble. However, the aim is to give the user control over their filter bubble, so we stick to the idea of making filter bubbles understandable, even though technically, the term user profile would be more correct.

perspectives. First, lacking diversity may seriously threaten democracy by obstructing the civic discourse (Bozdag and van den Hoven, 2015). Second, diversity and transparency play a major role in the satisfaction of the users with the recommender system (Aytekin and Karakaya, 2014).

Increasing Diversity through Controllability

Diversity in context with recommender systems has first been described by Bradley and Smyth (2001), who defined it as the opposite of similarity. Today it is one of the leading topics of recommender systems research, both as an approach to solving the over-specialization problem and a way to improve the quality of the user's experience with the recommender system (Kunaver and Požrl, 2017). Tintarev et al. (2013) conducted a study in which they examined the impact of diversity on users with different personality traits. The results show that users who are more open to new experiences prefer a greater level of diversity in their recommendations than users who were less open to new experiences (Tintarev et al., 2013). This means that diversity is at least partly subjective. Kunaver and Požrl (2017) argue that implementing diversity therefore requires the involvement of users to find the right level of diversity for each user.

Studies show that controllability is a means to increase diversity within recommender systems (Nagulendra and Vassileva, 2014, 2016). In another study, users were allowed to adjust the diversity level of their recommendations. The results show that the diversity level increased (Aytekin and Karakaya, 2014). This indicates that users are willing to participate in the process of recommendation diversification.

Interactive Word Clouds for Understandability of Filter Bubbles

Word clouds, also called tag clouds, are a common visualization technique for textual content. The main feature is that the relative size of a word within the word cloud indicates the importance or relevance of that word, usually in terms of its relative frequency. This feature can also be used to increase the transparency of recommendations. Recommender systems can be explained by presenting attributes that contributed to the suggestion of an item. However, this would only explain a single recommendation. To explain the recommender system's assumptions about user preferences (user model) more comprehensively, a system could visualize the most important (relevant) attributes of a user as words forming a word cloud. The word size would indicate how relevant an attribute is from the recommender's perspective.

Several works propose word clouds as effective implementation for explaining recommendations (Jones, Brun, Boyer and Hamad, 2011; Tsai and Brusilovsky, 2017, 2019). Word clouds can also be interactive and are therefore an approach to implement both transparency and controllability. Webster & Vassileva (2007) proposed an interactive visualization approach to explain and modify a recommender system. The system shows a user which other users are most similar in terms of their preferences. Additionally, the user can change the degree of influence that the other users have on the recommendations individually for each user. Nagulendra and Vassileva (2014, 2016) introduced another interactive visualization approach to illustrate and increase the awareness of the filter bubble in Online Social Networks. They designed their visualization based on a bubble metaphor to improve the comprehensibility of the filtering process in a way that is intuitive for the user. Category bubbles within the filter bubble show which categories are relevant from the recommender's perspective. The larger the bubble the greater the relevance of the category. Their system also offers the possibility to show which friends are most similar in terms of their interests. Additionally, users can modify the filter bubble by adding or deleting categories using a drag and drop function.

Methodology and Research Setting

Our research took place in the context of a 2.5-year design-oriented research project on recommender systems in public broadcasting. Public broadcasters have a public-service remit to fulfill, so the management of filter bubbles is of special interest to them.

The term filter bubble usually describes a mixture of both unbalanced topics and unbalanced opinions. Within the scope of this research, we focus on **content filter bubbles**, i.e., we look at filter bubbles from the perspective of content. To the best of our knowledge, there has only been little research on the understandability of filter bubbles from a human-computer interaction perspective. Therefore, we chose to conduct a qualitative survey.

For preparation, we performed a design thinking (Plattner, Meinel and Leifer, 2011) workshop to generate prototypes. In the workshop, we iteratively developed prototypes for the design of word clouds for filter bubbles (Figure 1). The prototypes consist of words representing topics (e.g., "politics" or "musical") and a circle that separates words inside the filter bubble (inside the circle) from words outside of the filter bubble (outside the circle). The variants had different characteristics with regards to text style, the local arrangement of words, and the quantity of information. After developing 7 prototypes, we found that we had a sufficient basis to let interviewees talk about the differences and obtain rich information.

We then conducted a qualitative interview study to see how participants perceive the different prototypes. As common in usability research, we used the think-aloud method. This method encourages participants to verbally describe what they see, do, and think while exploring a prototype (Bruun and Stage, 2015). The goal was to gather rich data on the user perceptions regarding the understandability and controllability of filter bubbles.

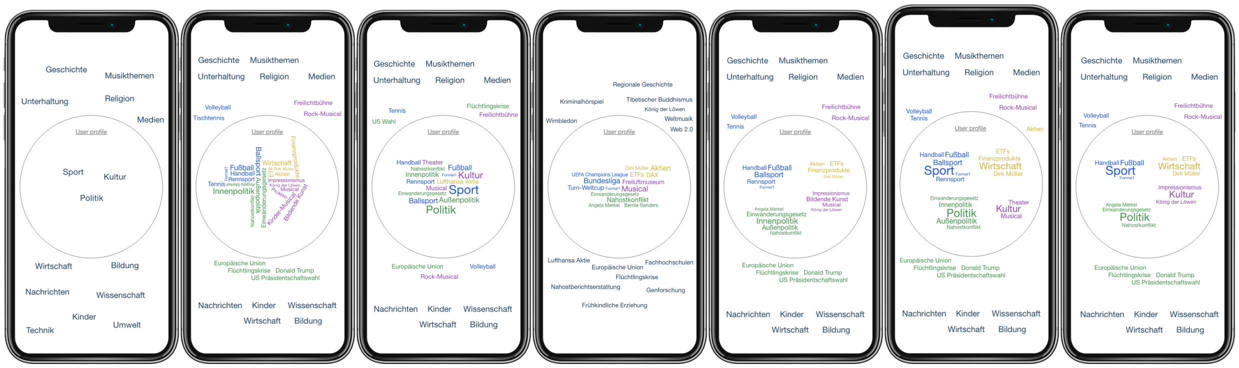

Figure 1. Seven prototypes with different characteristics

Approach and Data Collection

Since there are different variations of think-aloud methods, we considered the three following variations: *Traditional Think Aloud* follows the method of Ericsson and Simon (1993) and does not allow any intervention of the test beyond probing phrases. *Speech Communication Think Aloud* is based on the theories of speech communication (Boren and Ramey, 2000). Here the interviewer can keep the participants talking by picking up the last word that the participant said or with phrases like "um-huh". The *Coaching Think Aloud* approach allows for more verbal feedback. By asking direct questions, e.g., about specific areas of the prototype, the interviewer can guide participants in a certain direction if they are having difficulties (Hertzum, Hansen and Andersen, 2009).

Out of these three variations we opted for the *Coaching Think Aloud* method. As part of the "coaching", we decided to present the seven different prototypes one after another in random order to each participant to make them realize the differences between prototypes and trigger them to talk about the differences. Furthermore, we gave them tasks in which they had to modify the filter bubble (e.g., "you want to delete topic *racing* from your profile", "you want to add the topic *finance products* to your profile", "you want to change the relative importance of topic *refugee crisis*"), so they were forced to think about the visualization and interaction.

We conducted ten interviews in May and June 2020 in German language. To collect comprehensive data covering a diversity of different viewpoints, we aimed at interviewing persons of different ages and backgrounds. Additionally, we wanted to prevent gender bias by interviewing the same number of male and female persons. We interviewed five female and five male persons with an age ranging from 16 to 40 years (16, 18, 22, 23, 23, 25, 26, 29, 32, 40 years) with different backgrounds (2x pupils, 2x students, 1x graphic designer, 2x IT experts, 1x stunt woman, 1x journalist, 1x teacher). We recorded all interviews, transcribed them (smooth verbatim transcript), time-coded paragraphs of coherent statements, and indicated which prototype they refer to.

Data Analysis

To analyze the transcriptions, we applied Qualitative Content Analysis (QCA) to the transcriptions. This method aims at structuring and organizing obvious and latent content. The researcher creates a coding frame by assigning successive parts of the material to its categories. The coding frame is substantial and can be the result itself (Mayring, 2010). An advantage of this method is that one can systematically describe and interpret qualitative data in a rule-guided way. This makes the method both transparent and replicable. Since our research question is largely exploratory in nature, we did not develop categories from theoretical considerations or use a predefined set of categories. Therefore, we chose to conduct an inductive category development as proposed by Mayring (2010). This procedure took place in two steps. First, we performed a free coding with inductively building up *categories*. We decided to have all interviews coded by two researchers, compare results, and resolve potential conflicts through discussion. Second, we built *main categories* by grouping the categories resulting from the previous step. We did not perform a frequency analysis, as the purpose was to gather rich data from different viewpoint, but not their frequencies.

Results

In this section, we present the results from applying the QCA to the think-aloud protocols of ten participants who explored seven different prototypes of interactive word clouds representing a user's filter bubble within a recommender system. Mayring (2010) proposes to present the findings of exploratory QCA by presenting and illustrating the coding frame resulting from the analysis. This comprises (1) the coding frame itself, (2) its codes, and (3) their meanings. The final coding frame we constructed consists of two levels comprising 15 categories grouped in three main categories (Table 1). The presentation of the results is structured along the coding frame. The main categories are shown as headings and the original codes of the categories are shown in boldface.

Main category	Categories
A) Text layout of personal interests	A1. Horizontal word alignment
	A2. Cluster-based color coding of words inside the bubble
	A3. Cluster-based color coding of words outside the bubble
	A4. Font size for words inside the bubble
	A5. Font size for words outside the bubble
B) Local arrangement of personal interests	B1. Local thematic clustering
	B2. Distances between thematic clusters inside the bubble
	B3. Position of related words inside and outside the bubble
	B4. Position of main topics outside the bubble
	B5. Distance of topics to center of bubble
C) Level and quantity of information	C1. Presence of main topics
	C2. Presence of subtopics
	C3. Quantity of words
	C4. Tradeoff quantity – level of thematic abstraction

Table 1. Coding frame

Findings regarding Text Style of Personal Interests

To begin with the most intuitive category when thinking of the properties of a word cloud, the *text style* includes all subcategories related to the styling of the words that make up the word cloud. These include the alignment, the font size, and the color of the words.

A1. Horizontal word alignment. Even though only one out of seven prototypes showed non-horizontally aligned words (prototype 2), the corresponding category is the only one that was mentioned by all ten participants. All participants expressed either positive perceptions regarding the horizontal orientation or negative perceptions regarding words that are not horizontally oriented. They explained that horizontally aligned words were more comfortable to read and that they prefer not having to turn their devices to read the words.

A2. Cluster-based color coding of words inside the bubble. This subcategory refers to the color-coding of words within the displayed bubble based on thematic clusters, i.e., all words referring to the same main topic (e.g., sports) are displayed in the same color (prototypes 2, 3, 4, 5, 6, and 7). Nine out of ten participants commented on this subcategory. Eight of these nine assessed color-coding positively and stated that it not only helps to identify thematically related words but also facilitates orientation. Nevertheless, one person did not perceive value in colored words.

A3. Cluster-based color coding of words outside the bubble. This category is similar to A2 but refers to the words outside the bubble (prototypes 2, 3, 5, 6, and 7). All five persons who mentioned the cluster-based color coding of words outside the bubble perceived value by the color codes and named easier understanding and navigation as reasons for their perception.

A4. Font size for words inside the bubble. In typical word clouds, the font size represents the importance of a word within a set of words (as in prototypes 2, 3, 4, 5, 6, and 7), and, analogously, the influence of a topic on the output of a recommender system. Six participants perceived the font size as a representation of the weight of a word in the user profile and found this intuitive and easy to understand. This is opposed by the statement of one person who interpreted the font size as an indicator for a word's level of thematic abstraction.

A5. Font size for words outside the bubble. In contrast to seven participants who commented on the font size of the words inside the user profile, only one person mentioned the font size of the words outside the user profile. According to this person, it was logical to choose a uniform font size for the words outside the user profile because they are not relevant for the user.

Findings regarding Local arrangement of Personal Interests

The following categories refer to the *local arrangement* of the words forming the word cloud. In contrast to typical word clouds, in which the words are usually arranged close together, we prepared prototypes with variations of the local arrangement of words, e.g., in clusters.

B1. Local thematic clustering. Local thematic clustering refers to the arrangement of words within the bubble so that all words that belong to the same main topic are located close together in "sub clouds" of the same color (prototypes 2, 4, 5, 6, and 7). Five out of six participants found this helpful because it creates structure and clarity. In contrast, for prototypes without local thematic clustering (prototypes 1 and 3), one person complained that the words were not arranged according to their thematic similarity. This would make it more difficult to find interesting topics. One participant complained that the ratio of the topics influencing the recommendations was unclear.

B2. Distances between thematic clusters inside the bubble. All except for one person mentioned the local distances between thematic clusters (as implemented in prototypes 5, 6, and 7). Here it was noticeable that this happened without exception when the participants explored the first prototype which visualized the distances (prototype 5). Eight participants evaluated it positively mainly for reasons of improved clarity. The one who did not like it justified this by saying that this arrangement makes it more difficult to see the relationships between the words.

B3. Position of related words inside and outside the bubble. This subcategory is related to the previous subcategory and refers to prototypes that arrange words outside the bubble closer to thematically related words inside the bubble (prototypes 5 and 6). All four participants who commented on it had a positive perception. They explained that it would improve orientation in the search for interesting topics to supplement the profile and provide additional information about the recommender's assumptions.

B4. Position of main topics outside the bubble. In prototypes 2, 3, 5, 6, and 7, the main topics outside of the bubble were moved far to the edges of the screen. Two participants found this intuitive.

B5. Distance of topics to center of bubble. Two participants mentioned that they could not figure out whether the distance of a word to the middle of the user profile had a meaning or not.

Findings regarding Level and Quantity of Information

The main category *level and quantity of information* comprises all categories dealing with either the level or the quantity of information presented in the word cloud. This refers to different variables such as the level of abstraction or the total number of words. The higher the level of abstraction, the less detailed the words are and vice versa. Our prototypes showed several variations of this, from a word cloud consisting only of main categories to word clouds comprising words in several levels of detail.

C1. Presence of main topics. Prototypes 6 and 7 visualized main topics as cluster representatives (e.g., *sports* or *culture*). Six out of eight participants perceived the main topics representing the clusters as helpful for easy identification of the clusters and orientation within the word cloud. The other two participants were unsure whether a main topic inside the bubble is related to the relevance of all related subtopics or not.

C2. Presence of subtopics. Three participants mentioned the presence of subtopics (visualized in prototypes 2, 3, 4, 5, 6, and 7). Subtopics refer to topics around the second level of abstraction within a taxonomy. Two participants commented positively on the presence of subtopics. They said this would show which subtopics the user profile contains and clarify the underlying taxonomy. In contrast, for one person the presence of intermediate topics led to confusion about the relevance of the main topics.

C3. Quantity of words. Six participants commented on the number of words that make up the word cloud. Five of them perceived either a lower number of words (prototypes 1 and 7) as positive or a higher number (prototype 2 and 6) as negative and justified this with a less overloaded picture and an easier decision-making process. In contrast, one participant complained of a lack of information when fewer words are displayed.

C4. Tradeoff between quantity and level of thematic abstraction. Our prototypes showed different levels of thematic abstraction. Some prototypes showed a rather detailed level of thematic words (prototypes 2, 3, 4, 5, and 6), others showed less information (prototypes 1 and 7).

While five participants criticized an information overload (on prototypes with detailed information) and a too detailed level for adjusting their profile, one participant found that more detailed words provided a more appropriate level of differentiation. One participant explained that words outside the bubble should initially be unspecific to maintain clarity. When selecting the topics to be added to the profile, one should be able to specify the main topics such that more detailed words can be added to it, that is, detailed words inside the user profile and main topics outside of it.

In contrast, for prototypes that showed a high degree of abstraction (prototypes 1 and 7), four out of seven participants complained that they could not differentiate sufficiently when selecting relevant topics. The other three assessed it positively for reasons of more clarity and the avoidance of information overload. One of them also pointed out that the interaction would require less cognitive effort because a whole cluster could more easily be removed by deleting only one main word from the user profile instead of multiple detailed words.

All in all, prototypes 6 and 7 can be identified as the preferred ones since they fulfill most requirements that can be derived from the participants' perceptions.

Discussion

The results of the study indicate that all participants had a good overall understanding of the prototypical visualizations. Even though we did not explain the meaning of any property of the word cloud, the participants were immediately able to conclude their meanings. This suggests that word clouds can be considered as a valid approach to make filter bubbles in recommender systems understandable and controllable in an intuitive way. Yet, the design requirements of such interactive filter bubble word clouds differ from those of common, static word clouds.

Our findings may inform broadcasters, publishers, and media houses how to design interactive word clouds to make their recommendations understandable and controllable. Designers of interactive filter bubbles in digital journalism can refer to three main categories to give users control over their filter bubbles and, as a result, over the diversity in their information stream. Referring to *Text Style* (first main category), we have indications that designers of interactive filter bubble interfaces should align all words horizontally and ensure that they are always displayed in a legible text size. While these aspects are more of a basic requirement, more text style properties directly contribute to understandability and controllability of filter bubbles. The results suggest that filter bubble interfaces should color-code words based on thematic clusters to facilitate the orientation of users. We also have a strong indication that a common feature of static word clouds, the font size representing the weight of a word within a set of words, can be used to visualize the importance of a user's interest and the weight of a topic for the recommender system. The second main category we identified concerns the *local arrangement* of the displayed words. According to the results, words should be arranged in thematic "sub clouds", i.e., words of thematic similarity for smaller clouds that are locally separated. The third main category that filter bubble interaction designers should refer to is the *level and quantity of information* to be visualized in the word cloud. Here the results are less clear how to design the tradeoff between number of words and detail of information, which indicates that this category is more complex to deal with. Yet, this category might be the most critical one since the main challenge is to develop effective visualization techniques that can be easily integrated without contributing additionally to the overload issue (Nagulendra and Vassileva, 2014, 2016). While main topics (words of the highest level of thematic abstraction) seem to be good cluster representatives, some users want to modify their filter bubble on a more detailed level. At the same time, they prefer neither a more detailed nor a more imprecise level. Limiting the number of words displayed seems to make sense according to our results. This could be interpreted as an indication that the word cloud should allow users to navigate through different levels of abstraction, starting with the highest.

Conclusions, Limitations and Future Research

While users need algorithmic systems to filter content and manage information overload, phenomena such as filter bubbles occur, resulting in people lacking a diversity of viewpoints without being aware of reality (Nagulendra and Vassileva, 2016; Pöchhacker et al., 2017). Our work analyzes user perceptions of specific properties of interactive word clouds in terms of their ability to enhance the understandability of filter bubbles and controllability of diversity. The goal was to give the user transparency and control over the filter bubble and to create a basis for building trust in the system.

The results show clear tendencies in the perception of certain properties. From this, we can conclude how word clouds should be designed as interactive interfaces so that users can intuitively understand and control filter bubbles. In doing so, we contribute to research since only a few studies focused on how to design word clouds as an interactive visualization interface for recommender systems. The study builds a foundation for the development of effective visualization techniques that address the handling of diversity in digital artifacts.

Yet, this work is not free of limitations. One limitation is the relatively small number of participants in our study. Although the participant we interviewed last did not mention any new categories, we cannot rule out the existence of others. A second limitation concerns the distribution of the age of the participants. Since the oldest person was 40 years old, conclusions drawn from the results of this study might not apply to older persons. Another limitation of this study is the variation of word cloud properties represented in the prototypes. The participants mainly commented on properties that changed from one prototype to another. There might be properties that are relevant, but which were not mentioned because they are not represented in the seven prototypes. In addition, we conducted our test only on smartphone displays. Referring to the Elaboration Likelihood Model (ELM) and the Heuristic-Systematic Model (HSM), users on larger screens might have different perceptions via different perceptual pathways. Therefore, it is a limitation that our research does not clarify the ways in which users' information processing affects the understandability and possible differences in perception between large and small screens.

Our research contributes to theory building regarding the understandability of filter bubbles as a foundation for trust in recommender systems. Our findings might lay the foundation for both design theories in design science and confirmatory research approaches. Within design science, our findings can give a first indication for the development of design principles. Regarding confirmatory research approaches, our exploratory study may serve as a basis for the formulation of hypotheses and the design of quantitative surveys. Further research may also analyze the influence of controllable filter bubbles (e.g., in the form of interactive word clouds as designed in this research) on the overall acceptance of recommender systems.

In this research, we focused on biased diversity exposure with regards to content (called content filter bubble). Future research can bring an opinion perspective and political orientation into the diversity considerations, raising the question of how to categorize opinions and how to integrate them into the bubble visualization.

REFERENCES

Agichtein, E., E. Brill and S. Dumais. (2006). "Improving Web search ranking by incorporating user behavior information" (pp. 19–26).

Aytekin, T. and M. Ö. Karakaya. (2014). "Clustering-based diversity improvement in top-N recommendation." Journal of Intelligent Information Systems, 42(1), 1–18.

Bernstein, A., C. de Vreese, N. Helberger, W. Schulz, K. Zweig, C. Baden, … T. Zueger. (2020). "Diversity in News Recommendations." ArXiv:2005.09495 [Cs].

Borchers, A., J. Herlocker, J. A. Konstan and J. Riedl. (1998). "Ganging up on Information Overload." Computer, 31(4), 106–108.

Boren, T. and J. Ramey. (2000). "Thinking aloud: reconciling theory and practice." IEEE Transactions on Professional Communication, 43(3), 261–278.

Bozdag, E. and J. van den Hoven. (2015). "Breaking the filter bubble: democracy and design." Ethics and Information Technology, 17(4), 249–265.

Bradley, K. and B. Smyth. (2001). "Improving Recommendation Diversity."

Bruun, A. and J. Stage. (2015). "New approaches to usability evaluation in software development: Barefoot and crowdsourcing." Journal of Systems and Software, 105, 40–53.

Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User-Adapted Interaction, 12(4), 331–370.

Chen, M.-H., C.-H. Teng and P.-C. Chang. (2015). "Applying artificial immune systems to collaborative filtering for movie recommendation." Advanced Engineering Informatics, 29(4), 830–839.

Das, A. S., M. Datar, A. Garg and S. Rajaram. (2007). "Google news personalization: scalable online collaborative filtering." In: Proceedings of the 16th international conference on World Wide Web (pp. 271–280). Banff, Alberta, Canada: Association for Computing Machinery.

De Nart, D. and C. Tasso. (2014). "A Personalized Concept-driven Recommender System for Scientific Libraries." Procedia Computer Science, 38, 84–91.

di Sciascio, C., V. Sabol and E. E. Veas. (2016). "Rank As You Go: User-Driven Exploration of Search Results." In: International Conference on Intelligent User Interfaces (pp. 118–129). New York: ACM.

Diao, Q., M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang and C. Wang. (2014). "Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS)." In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 193–202). New York, New York, USA: Association for Computing Machinery.

Ericsson, K. A. and H. A. Simon. (1993). Protocol Analysis: Verbal Reports as Data (revised edition). Cambridge, Mass: MIT Press.

Hannak, A., P. Sapiezynski, A. Molavi Kakhki, B. Krishnamurthy, D. Lazer, A. Mislove and C. Wilson. (2013). "Measuring personalization of web search." In: International conference on World Wide Web (pp. 527–538). Rio de Janeiro, Brazil: Association for Computing Machinery.

Hertzum, M., K. D. Hansen and H. H. K. Andersen. (2009). "Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload?" Behaviour & Inf, Technology, 28(2), 165–181.

Hirschmeier, S. and V. Beule. (2018). "Compliance of Personalized Radio with Public-Service Remits." In: Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP) Adj.

Jones, N., A. Brun, A. Boyer and A. Hamad. (2011). “An Exploratory Work in Using Comparisons Instead of Ratings.” In: C. Huemer & T. Setzer (Eds.), E-Commerce and Web Technologies (pp. 184–195). Berlin, Heidelberg: Springer.
Just, N. and M. Latzer. (2016). “Governance by algorithms: reality construction by algorithmic selection on the Internet:” Media, Culture & Society.
Knijnenburg, B. P., M. C. Willemsen, Z. Gantner, H. Soncu and C. Newell. (2012). “Explaining the user experience of recommender systems.” User Modeling & User-Adapted Interaction, 22(4–5), 441–504.
Kunaver, M. and T. Požrl. (2017). “Diversity in recommender systems – A survey.” Knowledge-Based Systems, 123, 154–162.
Liang, T.-P., H.-J. Lai and Y.-C. Ku. (2006). “Personalized Content Recommendation and User Satisfaction: Theoretical Synthesis and Empirical Findings.” Journal of Management Information Systems, 23(3), 45–70.
Lops, P. de G., Marco; Semeraro, Giovanni. (2011). “Content-based Recommender Systems: State of the Art and Trends.” In: F. R. Ricci Lior; Shapira, Bracha; Kantor, Paul B. (Ed.), Recommender Systems Handbook (pp. 73–106). Boston, MA, USA: Springer.
Lu, J., D. Wu, M. Mao, W. Wang and G. Zhang. (2015). “Recommender system application developments: A survey.” Decision Support Systems, 74, 12–32.
Martin, C. H. (2013, October 28). “Music Recommendations and the Logistic Metric Embedding.”
Mayring, P. (2010). Qualitative Inhaltsanalyse: Grundlagen und Techniken (11th ed.). Weinheim: Beltz.
Munson, S. and P. Resnick. (2013). “Encouraging Reading of Diverse Political Viewpoints with a Browser Widget.” In: Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM.
Nagulendra, S. and J. Vassileva. (2014). “Understanding and controlling the filter bubble through interactive visualization: a user study.” In: Proceedings of the 25th ACM conference on Hypertext and social media (pp. 107–115). Santiago, Chile: Association for Computing Machinery.
Nagulendra, S. and J. Vassileva. (2016). “Providing awareness, explanation and control of personalized filtering in a social networking site.” Information Systems Frontiers, 18(1), 145–158.
Pariser, E. (2011). The Filter Bubble: What The Internet Is Hiding From You. Penguin UK.
Plattner, H., C. Meinel and L. Leifer (Eds.). (2011). Design Thinking: Understand – Improve – Apply. Berlin Heidelberg: Springer-Verlag.
Pöchhacker, N., M. Burkhardt, A. Geipel and J.-H. Passoth. (2017). “Interventionen in die Produktion algorithmischer Öffentlichkeiten: Recommender Systeme als Herausforderung für öffentlich-rechtliche Sendeanstalten.” kommunikation @ gesellschaft, 18.
Resnick, P., R. K. Garrett, T. Kriplean, S. A. Munson and N. J. Stroud. (2013). “Bursting your (filter) bubble: strategies for promoting diverse exposure” (pp. 95–100). Presented at the Proceedings of the 2013 conference on Computer supported cooperative work companion, ACM Press.
Ricci, F. R., Lior; Shapira, Bracha. (2011). “Introduction to Recommender Systems Handbook.” In: F. R. Ricci Lior; Shapira, Bracha; Kantor, Paul B. (Ed.), Recommender Systems Handbook (pp. 1–30). Boston, MA, USA: Springer.
Slaney, M. and W. White. (2006). “Measuring Playlist Diversity for Recommendation Systems.” In: Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia (pp. 77–82).
Tintarev, N., M. Dennis and J. Masthoff. (2013). “Adapting Recommendation Diversity to Openness to Experience: A Study of Human Behaviour.” In: S. Carberry, S. Weibelzahl, A. Micarelli, & G. Semeraro (Eds.), User Modeling, Adaptation, and Personalization (pp. 190–202). Berlin, Heidelberg: Springer.
Tintarev, N. and J. Masthoff. (2012). “Evaluating the effectiveness of explanations for recommender systems.” User Modeling and User-Adapted Interaction, 22(4), 399–439.
Tsai, C.-H. and P. Brusilovsky. (2017). “Providing Control and Transparency in a Social Recommender System for Academic Conferences.” In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP ’17 (pp. 313–317). Bratislava, Slovakia: ACM Press.
Tsai, C.-H. and P. Brusilovsky. (2019). “Explaining recommendations in an interactive hybrid social recommender.” In: Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 391–396). Marina del Ray, California: Association for Computing Machinery.
Turcotte, J., C. York, J. Irving, R. M. Scholl and R. J. Pingree. (2015). “News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking.” Journal of Computer-Mediated Communication, 20(5), 520–535.
Webster, A. and J. Vassileva. (2007). “The keepup recommender system.” In: Proceedings of the 2007 ACM conference on Recommender systems (pp. 173–176). Minneapolis, MN, USA: Association for Computing Machinery.