Towards Understanding Desiderata for Large-Scale Civic Input Analysis

Mahmood Jasim
University of Massachusetts Amherst
Amherst, MA 01003, USA
mjasim@umass.edu

Ali Sarvghad
University of Massachusetts Amherst
Amherst, MA 01003, USA
asarv@umass.edu

Enamul Hoque
York University
Toronto, ON M3J 1P3, Canada
enamulh@yorku.ca

Narges Mahyar
University of Massachusetts Amherst
Amherst, MA 01003, USA
nmahyar@umass.edu

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
http://dx.doi.org/10.1145/3334480.3382964

Abstract
Advancement in digital civics and the emergence of online platforms have enabled vast amounts of community members to share their input on various civic proposals. The intricacy of the community input analysis process, coupled with the increased scale of community engagement, makes community input analysis particularly challenging. Civic leaders, who gather, analyze, and make critical decisions based on community input, struggle to make sense of large-scale unstructured community input due to lack of time, analytical skills, and specialized technologies. In this qualitative study, we investigated civic leaders' requirements that can accelerate the community input analysis process and help them to gain actionable insights to make better decisions. Our interviews conducted with 14 civic leaders revealed a dichotomous nature of requirements based on their roles and analysis practices. The interviews also revealed the civic leaders' desire to understand the community's opinions beyond sentiments and how text analysis and visualization can bring structure and enable sensemaking of community input. This study is our first step towards exploring the design of community input analysis technologies for civic leaders that can contribute to democratic decision-making in digital civics.

Author Keywords
Digital Civics; Community Input Analysis; Civic engagement

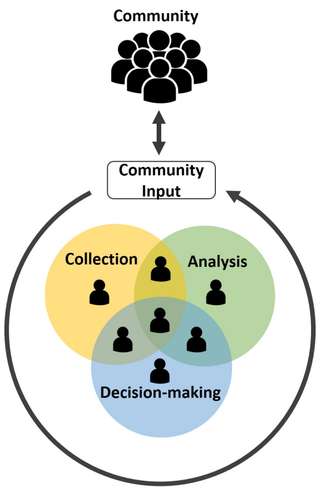

Figure 1: Large-scale community input collection and analysis is a complex process with many actors involved. In our study, we refer to these actors as *Civic Leaders*. Civic leaders have different roles, such as, *Decision-Makers* who make decisions on planning and policies, *Community Envoys* who work as an interface between decision-makers and community members to engage community members and collect community input and *Analysts* who analyze and interpret community input. Very often civic leaders assume more than one role that increases the complexity of the community input analysis process.

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); Interaction Design;

Introduction

Community engagement is paramount in the practice of participatory democracy [14]. The advancement in digital civics via several online platforms has enabled civic leaders to widen their outreach and collect extensive community input (e.g., CommunityCrit [21], Decidim [2], and DemocracyOS [1]). This raw unstructured community input goes through multiple layers of analysis and decision-making before being transformed into a concrete set of policy decisions [22]. The *Community Envoys* start this process by engaging community members, gathering community input, and often participating in analyzing and summarizing raw community input to generate reports. These reports are forwarded to decision-makers. When community envoys are not involved with analysis, they send raw community input to *Analysts* after collection. The analysts rigorously examine and interpret the community input and forward the analysis results to decision-makers. The *Decision-Makers* further analyze the data to make the final decision. In reality, these roles (community envoys, analysts, and decision-makers) are fluid and not exclusive to certain individuals as one person can assume multiple roles in the community input analysis process (Figure 1). For example, an analyst makes decisions about which community input to integrate into the analysis results. Similarly, decision-makers further analyze the results received from analysts before making the final decision. This interweaving of roles present unique challenges to design technologies for community input analysis due to the sensitivity of the community input and different analysis practices [22]. Moreover, as the scale of community input increases, regardless of their roles, civic leaders involved with community input analysis face obstacles in effectively analyzing, communicating, and utilizing the massive amount of unstructured community input [17, 22]. The challenges stem from their scarcity of time, lack of expertise with analytical tools, and the absence of specialized technologies tailored to their requirements [22]. In our study, we use *Civic Leaders* as an umbrella term to refer to various actors in community-based decision-making such as, *Decision-Makers*, *Community Envoys*, and *Analysts*.

In our previous work, we deployed a community engagement tool, CommunityCrit [21] and successfully gathered large numbers of community input on an urban development project in San Diego (Figure 2). However, the civic leaders we collaborated with struggled to examine the input using their current analysis approach. This is a recurring problem [22, 13] in the digital civics domain, where engaging and empowering community members to contribute their opinions is a crucial component of the digital democracy equation [12, 26]. However, the other component, that has been mainly overlooked, is how to enable civic leaders to understand and analyze the enormous amount of community input effectively to accelerate the decision-making process [13, 22]. Prior work indicated that current practice around analyzing the public input process is tedious and time and manpower intensive, which is often done through manually coding or using qualitative data analysis tools (e.g. [5, 3]) for thematic analysis [22]. In terms of technology use in digital civics, previous research enabled opinion sharing, consensus building, and provided quick overviews of community input. For example, ConsiderIt [18] builds a pro-con list to augment personal deliberation to help identify common grounds from diverse opinions. Opinion Space [10] provides overviews to understanding community opinion by allowing users to browse online opinions. DemocracyOS and Pol.is focus on community deliberation by allowing voting and aggregation of community input [1, 4]. While these tools provide surface-level overviews of community opinions, they do not support the

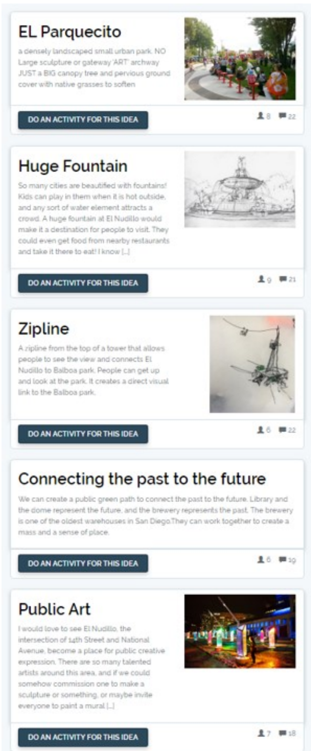

Figure 2: A screenshot of CommunityCrit [21], an existing community engagement platform. The figure depicts several ideas proposed by the community members regarding an urban development project in San Diego. *Proposals* are main agendas for civic discussions where community members can share their comments, new ideas, thoughts, and opinions.

detailed exploration of community input that is robust enough for civic leaders to sublimate large-scale community input into concrete actionable insights. Furthermore, civic leaders seldom have extensive training to use complex data analysis tools [22], which makes these tools ineffective in practice.

Prior work has called for attention towards a lack of specialized technologies for analyzing community input in the digital civics domain [13, 22]. Harding et al. advocated for human-centered research to better understand and adapt to civic leaders' requirements [13]. We address this call by closely working with and conducting interviews with 14 civic leaders, who are experienced in analyzing large-scale community input. Our thematic analysis revealed various desiderata that can help them effectively understand, analyze, and utilize community input by curating and designing technologies to satisfy civic leaders' requirements based on various roles and expertise, surfacing community's opinions that go beyond sentiment analysis, and utilizing text summaries and visualization to bring structure to community input.

Interview Study

To learn civic leaders' requirements, we conducted interviews with 14 civic leaders over four months. We recruited interviewees by emailing a wide range of civic leaders and used the snowball method [6] to find more participants. To capture multiple perspectives and community input analysis practices, we approached several organizations who grapple with the challenges of analyzing and utilizing large amounts of community-generated data. Table 1 present civic leaders we interviewed, their primary roles, and their approaches towards analyzing community input. Our interviewees are associated with different organizations within the United States, working in projects with a common goal towards community engagement and including community opinions in civic decision-making. The interviews explored several agendas

ID	Primary Role	Data Analysis Approach(s)
P1	Decision-Maker	Manual
P2	Decision-Maker	Google Docs and Spreadsheets
P3	Decision-Maker	Outsource
P4	Community Envoy	Manual
P5	Community Envoy	Manual
P6	Community Envoy	Outsource
P7	Community Envoy	Manual
P8	Community Envoy	Manual
P9	Analyst	Google Docs
P10	Analyst	Manual
P11	Analyst	Atlas.Ti and Dedoose
P12	Analyst	Atlas.Ti and Dedoose
P13	Analyst	Manual
P14	Analyst	Manual

Table 1: This table shows participants of the interview study, their primary roles, and their civic input analysis approaches. Their roles include Decision-Makers who make key policy decisions, Community Envoys who work to foster change on behalf of communities and Analysts who analyze and interpret the results. However, all of them could be engaged in other activities.

by asking open-ended questions regarding civic leaders' practices and challenges of analyzing community input. We began by asking *"What do you want to learn from the community input and why?"* We also asked, *"What method or technology do you use to analyze the data?"*. Furthermore, we asked them questions around how to address the challenges of processing these inputs, such as, *"What could help you in analyzing community data?"*. Finally, we touched upon the desired attributes of potential analysis technologies that would help them understand the community input, asking *"What features would you want in civic data analysis technologies?"*. Based on these interviews and prior work on community engagement [10, 19] we performed rapid prototyping to further discussion around design requirements in followup interviews with all participants. Figure 3 presents some example prototypes. We shared the prototypes with

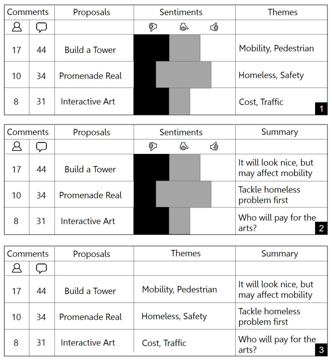

Figure 3: Example prototypes we shared with interview participants before the followup interviews. Example 1 (at the top) showed various information such as the proposal title, the number of commenters and comments on each proposal. Based on civic leaders' self-reported familiarity with simple visualizations, we added a bar chart depicting positive, neutral, and negative comments to address their requirements to know the community's opinions. Also, it showed the most important topics of discussions as per civic leaders' requirements. Example 2 replaced topics with snippets of text summaries of the comments made on these proposals to reflect civic leaders' current practice of using text summaries. Example 3 replaced the bar charts to show the topics and text summaries.

.

the participants by email before the followup interviews. All interviews were semi-structured and conducted over the telephone that lasted between 30-45 minutes and all participation was voluntary. Each interview was conducted with a minimum of two members of our research team to divide the roles of interviewer and note-taker. The interviews were audio-recorded and extensive notes were taken for later analysis. Over 900 minutes of interview audio was collected and transcribed. The transcripts and notes were thematically analyzed using iterative open coding method [7] where we looked for civic leaders' challenges and requirements. The analysis revealed several themes and critical design requirements. We also extracted representative quotes from participants to support our analytical claims.

Findings

We present the key findings from our interviews and highlight the requirements for civic data analysis technology.

Designation dictates dichotomous desiderata: We found an important distinction on requirements among the civic leaders we interviewed based on their designated roles and practices. The interviewees who are primarily involved with decision-making, wanted high-level summaries of community input to help them to concentrate on critical issues effectively and efficiently without having to interpret every input individually. In contrast, the others who are primarily involved in analyzing community input thought that summaries of community input alone without considering individual input might oversimplify the underlying narrative, leading to the incomplete interpretation of community input. They mentioned *"Overviews can suppress minority viewpoints"* (P13), which may lead to marginalization. To do so, they required exploration functionalities to drill down to individual input and examine each community input in detail. Some analysts noted, *"you have to refer "to the community" to understand their specific needs and "what works for them"* (P10) which is important *"especially with regards to lower-income neighborhoods"* (P13). Another analyst mentioned that summaries make is difficult to come to *"any conclusions about the comments"* (P8). While both groups found summaries useful for high-level exploration of community input and for focusing on specific agenda, depending on their roles, practices, and perspective towards community input analysis process, they were divided on to what extent summarization should be utilized in community input analysis. They wanted to negate the biases in decision-making by including all community input while maintaining reasonable time and labor requirements for community input analysis and decision-making.

Surfacing community's opinions that go beyond sentiments: A common requirement among interviewees was to understand the community's emotions towards a proposal. They emphasized the importance of understanding the *"reactions of the community towards the circumstances around them and the future prospects"* (P6). Another participant (P9) mentioned, "Decision-makers need to know which direction public opinion lies". In the follow-up interviews, we discussed what kind of opinions they want to extract from the community input. Some participants mentioned their familiarity with sentiment analysis in extracting opinions from the text. However, they were concerned about how sentiment analysis groups several viewpoints into singular categories (Positive or Negative). As a result, they found sentiment analysis to be ineffective in surfacing nuanced community emotions. One of them mentioned, *"We aren't usually thinking positive, neutral or negative, we are usually thinking: what are the different categories in the way people are thinking about something"* (P11). We further asked them what kind of emotions they want to extract from community input and mentioned some popular categories in emotion analysis, such as Anger, Disgust, Surprise, Happiness, etc. However, the participants mentioned that some of these categories seldom appear in civic discussions, saying, *"Emotions like Disgust and Surprise are not relevant and useful to civic discussions [...]. Since the community is aware of the discussion proposals, I don't see how that [Surprise] factors in"* (P2). Rather, the civic leaders wanted to learn whether the community is excited or angry towards a civic issue, or if they are concerned, happy, or if they do not care at all.

Bringing structure to text around visualization: The interviewees mentioned difficulties keeping up with the increasing scale of unstructured community input, resulting in an expensive, tedious, and time-intensive analysis process. Most of them wanted ways to provide some sort of structure for the community input to gain a high-level understanding. In the follow-up interviews, we probed further into what kind of structure they want, to efficiently analyze community input. Most of them mentioned visual summaries are useful to show which civic proposals or topics are generating the most discussions, or which are attracting community members the most. They articulated their familiarity with using simple charts (e.g., bar chart) in their current practice. They thought high-level visual summaries can be useful in accelerating the analysis process through faster exploration. However, some noted that visual summaries could be insufficient in contextualizing the community input. As a result, they also wanted text summaries of community input. This preference for text summary is rooted in their current data analysis practices, as they thought text summaries match their *"traditional methodology"*(P3) for public input analysis. The interviews and discussions around the prototypes highlight the importance of combining text analysis and visualization to surface actionable insights from community input.

Defining themes on the fly: The participants wanted to know the main supporting themes of discussions on the civic proposals. However, they objected against the fixed categorization of such themes by drawing parallels with product review analysis, where fixed aspects of products are surfaced (e.g., price, weight, etc.). In contrast, they wanted dynamic extraction of underlying themes from community input and gradually add or remove themes as they continue to gain further insight into the community input aligned with the spirit of qualitative thematic analysis. One participant mentioned, *"Sometimes it's easy to have predetermined categories. But often, we found, some of the most surprising things came through kind of open-coding and that often attracted interest"* (P2). The interview discussions reveal the importance of designing technologies aligned with civic leaders' practices.

Discussions and Future Work

Community input analysis is at the risk of becoming a bottleneck for participatory democracy due to the time and effort needed to make informed decisions based on large-scale unstructured community input. Therefore, it is crucial for HCI researchers to engage in designing novel technologies for civic leaders to help them analyze and more effectively utilize larges-scale community input in civic decision-making.

In our study, we found contrasting requirements among civic leaders on what information to extract and present for analysis. Due to the tedious and time-intensive nature of the civic data analysis process, some civic leaders want summaries of the community to balance the understanding of community input and making timely decisions. Others, however, want to drill down to individual community input as they think summaries increase the risk of suppressing unpopular opinions and may lead to marginalizing smaller groups while incentivizing vocal ones. It is crucial for HCI researchers to ensure both user groups are supported for community input analysis. Previous work on civic engagement and data analysis practices show that one of the primary reasons for resentment towards available qualitative data analysis tools [5, 3]) was the gap between what the civic leaders wanted to analyze and what they were able to analyze, leading to conflicts between analysts and decision-makers [22]. Our study extends their work [22, 13] by identifying civic leaders' contrasting viewpoints towards summaries and their requirements to understand the community's emotions. We advocate for designing technologies that provide adequate functionalities to analyze multifaceted community input for both user groups while maintaining consistency and synchronization throughout the community input analysis process.

To understand the community's perspectives, civic leaders need specific information regarding the community's emotions towards civic agendas that go beyond sentiment polarities (positive, negative, or neutral). This is aligned with existing research that emphasizes how emotions are important and influential drivers for decision-making [20]. Furthermore, using popular categories for emotion analysis [9, 23] without understanding the requirements of civic leaders in the digital civic domain is not helpful due to irrelevant categorization of the community's emotions. This is aligned with previous research in online discussions and learning environments where researchers proposed domain-specific categorization of emotions, arguing that popular emotion categories do not translate well to all domains [8, 24]. Our work identified that emotion categories such as anger, concern, happiness, or excitement are more important for civic decision-making rather than emotions such as disgust or surprise.

Our findings suggest that text summaries can help accelerate the civic input analysis process as it resembles civic leaders' current community input analysis practices. This is aligned with existing research that show how users prefer interfaces to mirror their existing practices and if novel features do not work as expected, they revert back to their old habits [25]. It is especially applicable to civic leaders who seldom have expertise in working with complex data analysis tools. Although there are recent successes in automatic text summarization, these methods are still heavily domain-dependent and may not generalize to digital civic domain due to lack of structure and ambiguity, and presence of sarcasm and euphemism in community input [11]. We also found that visualization is suitable for providing a high-level summary of community input which is aligned with previous work in other domains [15, 16]. We extend their work in the digital civics domain by positing that visualization can be used as a scaffolding to structure community input. Combining simple visualization and text analysis can establish a structure that provides an easy to perceive summary of community input that can be used to weave relevant community input information around it. Consequently, text summaries can be linked with this visualization to provide contextual information while maintaining a way to shuffle between the two as required. Our work calls to action for collaboration between HCI, natural language processing, and visualization researchers to find novel ways to combine text analysis and visualization to enable large-scale community input analysis.

This interview study is our first step towards enabling civic leaders to effectively and efficiently analyze large-scale community input. In this study, we focused on identifying the requirements for community input analysis technologies. Although our sample size for this study is limited to 14 participants, in the future, we will engage a broader spectrum of civic leaders to iteratively refine the design requirements to design and develop novel community input analysis technologies that combine text analysis and visualization. Furthermore, we will engage in longitudinal studies to investigate how such technologies can advance the community input analysis process. We will further examine their potential impact on democratizing the civic decision-making process.

REFERENCES

[1] 2019. DemocracyOS. http://democracyos.org/. (2019).

[2] 2019. Free open-source participatory democracy for cities and organizations. https://decidim.org/. (2019).

[3] 2019. NVivo. https://www.qsrinternational.com/nvivo/nvivo-products. (2019).

[4] 2019. Pol.is. https://pol.is/gov. (2019).

[5] ACM. 2019. ATLAS.ti. (2019). https://atlasti.com/.

[6] Patrick Biernacki and Dan Waldorf. 1981. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research* 10, 2 (1981), 141–163.

[7] Philip Burnard. 1991. A method of analysing interview transcripts in qualitative research. *Nurse education today* 11, 6 (1991), 461–466.

[8] Sidney D'mello and Arthur Graesser. 2007. Mind and body: Dialogue and posture for affect detection in learning environments. *Frontiers in Artificial Intelligence and Applications* 158 (2007), 161.

[9] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.

[10] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion Space: A Scalable Tool for Browsing Online Comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1175–1184.

[11] Mahak Gambhir and Vishal Gupta. 2017. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review* 47, 1 (2017), 1–66.

[12] John Gastil and Robert C Richards. 2017. Embracing Digital Democracy: A Call for Building an Online Civic Commons. *PS-Political Science and Politics* 50, 3 (2017), 758–763.

[13] Mike Harding, Bran Knowles, Nigel Davies, and Mark Rouncefield. 2015. HCI, civic engagement & trust. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2833–2842.

[14] Brian W. Head. 2007. Community Engagement: Participation on Whose Terms? *Australian Journal of Political Science* 42, 3 (2007), 441–454.

[15] Enamul Hoque and Giuseppe Carenini. 2015. ConVisIT: Interactive Topic Modeling for Exploring Asynchronous Online Conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 169–180.

[16] Mengdie Hu, Huahai Yang, Michelle X. Zhou, Liang Gou, Yunyao Li, and Eben Haber. 2013. OpinionBlocks: A Crowd-Powered, Self-improving Interactive Visual Analytic System for Understanding Opinion Text. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 116–134.

[17] Quentin Jones, Gilad Ravid, and Sheizaf Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research* 15, 2 (2004), 194–210.

[18] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012a. Supporting Reflective Public Thought with Considerit. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 265–274.

[19] Travis Kriplean, Michael Toomim, Jonathan Morgan, Alan Borning, and Andrew Ko. 2012b. Is this what you meant?: promoting listening on the web with reflect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1559–1568.

[20] Jennifer S. Lerner, Ye Li, Piercarlo Valdesolo, and Karim S. Kassam. 2015. Emotion and Decision Making. *Annual Review of Psychology* 66, 1 (2015), 799–823. DOI:http://dx.doi.org/10.1146/annurev-psych-010213-115043 PMID: 25251484.

[21] Narges Mahyar, Michael R James, Michelle M Ng, Reginald A Wu, and Steven P Dow. 2018. CommunityCrit: Inviting the Public to Improve and Evaluate Urban Design Ideas through Micro-Activities. (2018).

[22] Narges Mahyar, Diana V Nguyen, Maggie Chan, Jiayi Zheng, and Steven P Dow. 2019. The Civic Data Deluge: Understanding the Challenges of Analyzing Large-Scale Community Input. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, 1171–1181.

[23] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*. Elsevier, 3–33.

[24] Reijo Savolainen. 2015. Expressing emotions in information sharing: a study of online discussion about immigration. (2015).

[25] Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. 2018. More text please! understanding and supporting the use of visualization for clinical text overview. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 422.

[26] Vasillis Vlachokyriakos, Clara Crivellaro, Christopher A. Le Dantec, Eric Gordon, Pete Wright, and Patrick Olivier. 2016. Digital Civics: Citizen Empowerment With and Through Technology. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1096–1099.