Note: A Sociomaterial Perspective on Trace Data Collection: Strategies for Democratizing and Limiting Bias

Dipto Das
dipto.das@colorado.edu
Department of Information Science, University of Colorado Boulder
Boulder, Colorado, United States

Arpon Podder
Arponpodder.ap@gmail.com
Khulna University of Engineering and Technology
Khulna, Bangladesh

Bryan Semaan
bryan.semaan@colorado.edu
Department of Information Science, University of Colorado Boulder
Boulder, Colorado, United States

ABSTRACT

Researchers heavily use online platforms for collecting trace data, i.e., data capturing user interaction on and with sociotechnical systems. Human-computer interaction scholars have highlighted the role of reflexivity while analyzing such data in the case of marginalized communities. Drawing on sociomaterial perspectives, we highlight how data collection approaches involving lists of search phrases and APIs can embed researchers' positionality, perspectives, and biases within the datasets. In this note, we reflect on the data collection approaches of two papers that studied the sociohistorically marginalized Bengali communities on the question-and-answer site Bengali Quora. We illustrate how recommendation systems and data labeling workers can be included in the data collection process to democratize and limit bias while broadening and contextualizing the trace datasets for research.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; • **Information systems → Information retrieval**; **World Wide Web**.

KEYWORDS

Data Collection, Bengali, Sociomaterial, Apparatus, Bias

ACM Reference Format:
Dipto Das, Arpon Podder, and Bryan Semaan. 2022. Note: A Sociomaterial Perspective on Trace Data Collection: Strategies for Democratizing and Limiting Bias. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3530190.3534835

1 INTRODUCTION

Data from online platforms, such as social media and question-and-answer (Q&A) sites, are often collected and used by scholars across various domains and disciplines (e.g., social computing [24], affective computing [8], data science [2], software engineering [22], political science [32]) to understand human behavior, re-design technical systems, and make predictions. These trace data are often collected using application programming interfaces (APIs). Researchers generate lists of keywords, phrases, and tags and then use those as the primary inputs into APIs. Nevertheless, little work has reflected upon how researcher biases mediate the preparation and utilization of these keyword lists.

Especially while studying marginalized communities, tensions emerge around exploitation, membership, disclosure, and allyship [20]. As many studies in recent years have focused on marginalized communities online [12, 13, 16], HCI scholars have highlighted how researchers' identities (e.g., race, gender, sexual orientation) can bring certain affinities into perspective which can shape the data collection and analysis process [26]. While researcher bias in data analysis is often talked about, the possibility and sources of biases in data collection are not as heavily discussed. We draw on a sociomaterial perspective–that views social and material to be intertwined, highlighting how the process of trace data collection from sociotechnical systems (e.g., online platforms) is entangled with researchers' positionality and API biases.

In this note, we discuss the cases and lessons from two recent studies–one on sociohistorically marginalized Bengali communities' identity work [11] and another on their experience with governance on Bengali Quora (BnQuora) [9]. Drawing on the sociomaterialist metaphor of apparatus [3, 23], we explicate how the researchers' sociohistoric backgrounds and standpoints affected their choice of search phrases, and influenced their collected dataset in those studies. Recognizing the difficulty of avoiding this reflexivity of the researchers, we discuss how those studies strategized the use of on-platform recommendation system to democratize their data collection apparatuses to be inclusive of more people's views and opinions.

We argue that beyond the above mentioned papers' exploration of sociohistorically marginalized Bengali communities on BnQuora [9, 11], the data collection processes of these papers are interesting from a methodological perspective and have important insights for using trace data in social justice-oriented research focused on marginalized communities. The following section explains concepts like trace data, apparatus, and how data's refractive, reflective, and diffractive readings determine the scope and epistemology of research. In section 3, we reflect on the data collection process from BnQuora as described in two papers by Das and colleagues [9, 11]. First, we discuss how the list of keywords and API functioning as apparatuses influenced the formation of their trace datasets. Later in that section, we examine how their use of BnQuora's on-platform recommendation system, and data labeling participants have widened the scope, in other words, democratized and limited

COMPASS '22, June 29-July 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9347-8/22/06.
https://doi.org/10.1145/3530190.3534835

biases in the datasets. We conclude by underscoring the implications of these strategies for future research.

2 LITERATURE REVIEW

In this section, we discuss the definitions of trace data. By drawing on the optical metaphors such as apparatus, refraction, reflection, and diffraction, we discuss how these different ways of looking at trace data correspond to different research streams. We also provide a general overview of the objectives and principles of recommendation systems.

2.1 Trace and Apparatus

By interacting with and via sociotechnical systems, users create trace data that captures users' performances, often at a fine level of detail [23]. User-generated contents like textual and multimedia posts (e.g., questions, answers, comments, replies) on online platforms (e.g., Q&A sites, social media) and weblogs are good examples of such trace data. These are often utilized in quantitative explorations of systems (e.g., identifying the resource requirements of a server at different times of the day based on the number of visitors). Based on similar data sources, as a qualitative approach, trace ethnography combines the richness of participant observation with the wealth of data logs to reconstruct patterns and practices of users in distributed sociotechnical systems [14]. Drawing on the concept "apparatus", Østerlund and colleagues highlighted a number of methodological and theoretical challenges associated with trace data that rise because of its inseparable nature of the social and material [23].

Barad defines apparatus as the material conditions that determine "what matters and what are excluded from mattering" [3]. By determining the phenomenon of interest, researchers distinguish and explore what are considered central to their research questions and what are considered out of scope for a particular study [23]. These distinctions or cuts matter because traces are seen through these apparatuses. To explain what we "see", feminist scholars Haraway and Barad have used optical metaphors refraction, reflection, and diffraction [3, 17]. Refraction epistemologically gives researchers a positivist-leaning view of data. By creating sharp boundaries around a phenomenon, such reading of trace data considers traces as authentic depictions of the world–free of distortion and homologous to originals. On the contrary, boundaries drawn by a reflection-inspired reading of trace data are fuzzy. Hence, incomplete, blurred, and distorted representations of pre-given objects need to be interpreted to determine meanings. Unlike in refraction and reflection, in the diffraction-based reading of trace data, subjects and objects do not pre-exist instead emerge through practice. Researchers adopt this sociomaterial approach to study the phenomenon and apparatus through one another–how traces ripple through the apparatus.

2.2 Basic Models in Recommendation Systems

User interaction on online platforms simultaneously shapes and is shaped by what contents are furnished by the recommendation systems on those platforms. In doing so, basic recommendation systems usually work in two ways [1]: (a) collaborative filtering methods and (b) content-based recommendation systems. Collaborative filtering methods make recommendations based on the collective power of multiple users' opinions (e.g., ratings) which are often highly correlated across various users and items. In content-based recommendation, the descriptive attributes of items (e.g., keywords) and users (e.g., profiles) are used to make recommendations. In addition to these, specific requirements and constraints are attended to in knowledge-based recommendation systems while generating recommendations. In practice, online platforms combine different recommendation algorithms to put their strengths in addressing various settings and purposes.

Quora provides a personalized home feed (both with questions and answers) experience to its users. For this purpose, the Quora recommendation algorithm (QRA) uses a multi-stage system of personalized learning-to-rank approach that shows contents to an individual user based on topics, social connections with other users, temporal relevance, potential capacity to answer questions, etc. [31]. Though providing personalized experience to users is one of key objectives of recommendation systems, scholars have highlighted the idea of supporting users to explore different views and perspectives through unpersonalized and diverse recommendations [6, 19]. Along that line, while browsing the platform, QRA also presents users with questions related to the current question they are viewing. These recommendations are particularly effective in fostering engagement from logged-out or unregistered users. Without knowing these users' preferences, the recommendations about related questions are unpersonalized. In this case, QRA uses machine learning approaches like different collaborative filtering algorithms (e.g., weighted average least squares, Bayesian personalized ranking) considering different factors such as textual similarity, user co-visit data, topics, quality, and popularity of the questions [31].

3 SOCIOMATERIAL CONSIDERATIONS IN TRACE DATA COLLECTION

Das and Semaan [11] explored the Bengali communities' experience with colonization–the practices through which foreign forces migrated and disrupted the social structures and lives of native and indigenous people [21]. They studied how the colonially marginalized Bengali communities engage in collaborative identity decolonization work–the process of reclaiming their local and indigenous identities, on BnQuora. In their prior work, Das and colleagues [9] studied the Bengali communities' experience with the governance on BnQuora–how the sociotechnical mechanisms of the platform support and impede their identity expression and performance. Both these papers adopted a trace ethnography approach to understand users' strategies and experiences by studying Q&A threads data from BnQuora.

3.1 How Data Collection Apparatuses Shape Research Endeavors

In this subsection, we will discuss how trace data collection tools (e.g., lists of keywords and API), which are often deemed standard components for data collection in many disciplines [2, 22], can reflect researchers' perceptions and biases. Using Das and colleagues' works [9, 11] as examples we illustrate how these components function as apparatus, i.e., influence and shape how researchers can view trace data–thus, question the assumption of objectivity of lists of keywords and APIs.

3.1.1 List of Keywords as Apparatus. A traditional approach for collecting trace data from online platforms is to use a list of keywords and search phrases. The terms and phrases that are related to the topic and context of the study are included in that list. Researchers make certain decisions about what keywords are relevant to the context or purpose of the study and what are not. In doing so, the researchers effectively have agency over the sampling and data collection from the target online platform.

While collecting the Q&A threads that are related to identity decolonization work, Das and Semaan [11] prepared a list of keywords based on the colonial history of the Bengal region. Different (a) related concepts, (b) historic figures, (c) events, (d) places, and (e) emergent political ideologies during the colonial period were considered relevant while preparing the list. Similarly, in [9], the initial dataset was collected by using a list of search keywords that included terms and phrases identifying (a) features of the platform (e.g., moderation and stages[1]), (b) narratives describing how people were experiencing governance, and (c) users' linguistic, national, or religious identities.

In both papers [9, 11], Das and colleagues tried to identify and include search terms on relevant concepts (e.g., colonialism, nationality). What these concepts, narratives, and experience can entail can be somewhat abstract and fuzzy. Again, given the long history of colonization in the region, preparing an exhaustive list of all historic figures, events, places, and emergent political ideologies during the colonial periods is difficult. Therefore, while the lists of keywords in both papers are quite extensive, Das and colleagues had to prioritize certain historic figures, events, places, and political ideologies [11] and certain platform features, identity categories, etc. [9] over others.

Thus, Das and colleagues' understanding of the social, historic, cultural, and political context of the Bengali communities or that of the sociotechnical scaffolds of content moderation and governance on BnQuora, determined the inclusion and exclusion criteria of keywords. These lists led to a purposive sampling [29] and collection of Q&A threads from the BnQuora platform. By reflecting what they thought as "important" onto the list of keywords, they performed agential cuts in the data collection. These cuts have determined what these papers could focus on and study about the user interaction on BnQuora. For example, while both Bangabandhu Sheikh Mujibur Rahman and Ziaur Rahman were influential figures in the liberation movement of Bangladesh, the list of keywords in Das and Semaan's work [11] included the former but not the latter. One can interpret such inclusion and exclusion while studying the decolonization of Bengali identity as the researchers' bias towards or prioritizing the political ideology of one of them over the other. In both papers [9, 11], to determine in which Q&A threads the Bengali users were undergoing through identity decolonization work or different communities were talking about their experience with the governance, is difficult. It is likely that many threads where these discussions on identity decolonization and/or user experience were taking place might have not been included into the trace dataset because of some important keywords and phrases not being present in the list of search terms. Thus, the traces collected through the apparatus, i.e., the list of keywords present a reflection–partial and incomplete representation, of pre-given objects (i.e. discussions, phenomenon) leading to interpretivist accounts of the process in both papers [9, 11].

3.1.2 Application Programming Interface (API) as Apparatus. Researchers who study data from online platforms usually collect their data using APIs [7, 30]. While some platforms (e.g., Twitter, Reddit, StackExchange) offer API for data collection, researchers often create their own APIs and web scraping tools for collecting data from different online platforms. While Quora does not provide an API, their terms of services[2] permit the use of web crawlers and scraping tools under a few conditions. Some examples of available data collection tools for Quora are: quoras [10], pyquora [28], quora-scraper [4], quorapy [5], etc.

Both papers that we are discussing in this note used the quoras API [10] in addition to Quora's web interface to collect Q&A threads data (questions, answers, and comments) from Quora. The first author of these papers accessed both the web interface and the quoras API using his Quora credentials [9, 11]. Since Quora provides personalized Q&A threads to its users [31], by using a researcher's Quora credentials, their access and privileges on the platform (e.g., which Quora stages or spaces the researcher is a member of and from which of these they can view posts) determines what contents are accessible during data collection. This influenced what Q&A threads researchers could retrieve and include during purposive sampling–potentially introducing researcher biases in their initial datasets. Moreover, while Quora offers multimodality, due to the quoras API's limitations, both papers only collected and analyzed textual data from the platform [10]. By using the list of keywords as input to the API, Das and colleagues formed the initial datasets for both papers. The API acting as an apparatus determined and put a boundary around what trace data get included and analyzed in the studies on the Bengali communities' collaborative identity decolonization work [11] and their experiences on the platform [9].

3.2 Strategies for Democratizing and Limiting Bias

Here, we discuss how the incorporation of unpersonalized recommendations can broaden the scope, democratize and limit researcher bias in data. However, while these recommendations can broaden the scope of datasets, relevance of those should be considered before including them into the final datasets. We reflect on how Das and colleagues balanced between broadening the scopes and bounding the contexts in their datasets [9, 11].

3.2.1 Recommendation Systems as Apparatus: Democratization through Unpersonalized Recommendation. Das and colleagues, in both their explorations–of BnQuora users' strategies towards identity decolonization work [11] and the communities' experience with the governance practices on the platform [9] included QRA's unpersonalized recommendations about related questions into their datasets as a form of snowball sampling [15].

[1]Das and colleagues [9] translated the Bengali word *"moncho"* as "stage". Quora stages are somewhat equivalent to Facebook groups and subreddits.

[2]https://www.quora.com/about/tos

As the QRA bases these recommendations on factors such as user co-visit data and popularity [31], we argue that they capture and reflect the aggregated interest and perceptions of the user mass on Q&A threads' relevance to the topic of interest. Unlike the list of search keywords and phrases that reflected the sociohistoric understandings of the researchers or unlike the API that reflected the prior activities of an individual researcher, the QRA trains itself based on interaction data from a much larger number of users. Hence, the inclusion of QRA's unpersonalized recommendations on related questions broadened the scope of the data. For example, Das and Semaan mentioned that while the term "Hindustan" (a colloquial endonym of "India") was not present on their initial list of keywords, unpersonalized recommendations on related questions retrieved several Q&A threads on this keyword [11]. Later, they included the Q&A threads on those topics in the dataset while considering the QRA's recommendations on related questions.

Acting as an apparatus that reflects the opinions of Quora's vast number of users, recommendation system on Quora contributed to democratizing the datasets in both studies [9, 11]. Here, the QRA informed by the user population of BnQuora serves as the apparatus. We can describe the users living in a sociohistorically marginalized society and using the BnQuora platform as the phenomena of interest. In both studies, the apparatus and the phenomena co-configure each other. Thus, the documentary traces that they generate are not pre-given rather created, leading to diffractive approaches for understanding the data [23].

3.2.2 Data Labeling Workers as Apparatus: Balancing between Broader scope and Bounded Context. While unpersonalized diverse recommendations support exploring of new ideas [6, 19], datasets need to be bounded within a context for research. Das and Semaan [11] found that not all Q&A threads in the search results retrieved by using the list of keywords and phrases and the ones recommended by the QRA during their data collection process were relevant to the context of their study. For example, searching with the term "colonial" retrieved two particular questions–while one was related to British colonization, the other was about colonization on Mars. While the first question was relevant to the context of Das and Semaan's study [11] on Bengali communities' identity work in relation to British colonization, the second was not considered relevant. Therefore, as an effort to contextualize the dataset, they used an additional data relevance labeling step between retrieving a potentially relevant Q&A thread (from keyword searching or from QRA) and including that into the dataset.

Whereas in-depth qualitative or quantitative analysis of trace data requires critical thinking, expertise in specific methods (e.g., inductive thematic analysis, statistical models), tasks like labeling whether a data instance is related to the context of the study is less skill-intensive and are often done by crowdsourcing [18, 27]. However, the identities of communities or individuals performing the data labeling tasks can shape the final dataset [25, 27]. The web data, in this case, is subject to data labeling workers' interpretations to determine meanings–reflective reading of trace data. As the data labeling workers navigate through the fuzzy boundaries of phenomena of interest, their perspectives play a role in the inclusion/exclusion of trace data. The ones carrying out the data labeling task, thus, collectively function as an apparatus through which researchers see and study the data. Given that their lived experiences influence the trace dataset, we emphasize the importance of considering the positionality of the individuals who participate in data labeling, similar to that in Das and Semaan's work [11].

4 CONCLUSION

Sociocultural and political perceptions of researchers and their relation with the context of a study and the corresponding community can embed certain biases in collecting and analyzing trace data by preparing lists of keywords and using APIs. In this note, we have discussed how the inclusion of on-platform recommendation systems in the data collection process can extend the scope of the data. Besides, we explain how data labeling workers can help contextualize the extended trace datasets. In addition to limiting biases in the dataset, future researchers can repeat this process of including algorithmically recommended related data multiple times to prepare a large dataset. This repetitive process can pave the way for answering various research questions using large-scale trace ethnography and data science-based approaches. If such an on-platform unpersonalized recommendation system is not available, researchers can consider using open-source recommendation libraries on the platform's public data dump to get similar recommendations on related traces to broaden the scope of datasets. During this process, they should also reflect on the positionality of the data labeling participants and emphasize the interpretivist nature of such studies while resisting the narrative of studies using online trace data to be completely unbiased and objective.

ACKNOWLEDGMENTS

This research was partially supported by CRII grant no. 1657429.

REFERENCES

[1] Charu C Aggarwal et al. 2016. *Recommender systems*. Vol. 1. Springer.
[2] Uchenna Akujuobi and Xiangliang Zhang. 2017. Delve: a dataset-driven scholarly search and analysis system. *ACM SIGKDD Explorations Newsletter* 19, 2 (2017), 36–46.
[3] Karen Barad. 2007. Meeting the universe halfway. In *Meeting the universe halfway*. Duke University Press.
[4] Youcef Benkhedda. 2020. quora-scraper. https://pypi.org/project/quora-scraper/. PyPI: https://pypi.org/project/quora-scraper/, last accessed: March 27, 2022.
[5] Zacchaeus Bolaji. 2019. quorapy. https://github.com/djunehor/quorapy. PyPI: https://pypi.org/project/quorapy/, last accessed: June 18, 2020.
[6] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2019. Optimizing the recency-relevance-diversity trade-offs in non-personalized news recommendations. *Information Retrieval Journal* 22, 5 (2019), 447–475.
[7] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.
[8] Dipto Das and Anthony J Clark. 2018. Sarcasm detection on facebook: A supervised learning approach. In *Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct*. 1–5.
[9] Dipto Das, Carsten Østerlund, and Bryan Semaan. 2021. " Jol" or" Pani"?: How Does Governance Shape a Platform's Identity? *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
[10] Dipto Das and Bryan Semaan. 2020. quoras: A Python API for Quora Data Collection to Increase Multi-Language Social Science Research. In *The 23rd ACM Conference on Computer-Supported Cooperative Work and Social Computing*. ACM.
[11] Dipto Das and Bryan Semaan. 2022. Collaborative Identity Decolonization as Reclaiming Narrative Agency: Identity Work of Bengali Communities on Quora. In *CHI Conference on Human Factors in Computing Systems*. 1–23.
[12] Bryan Dosono and Bryan Semaan. 2020. Decolonizing tactics as collective resilience: Identity work of AAPI communities on Reddit. *Proceedings of the ACM on Human-Computer interaction* 4, CSCW1 (2020), 1–20.

[13] Brianna Dym, Jed R Brubaker, Casey Fiesler, and Bryan Semaan. 2019. " Coming Out Okay" Community Narratives for LGBTQ Identity Recovery Work. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–28.

[14] R Stuart Geiger and David Ribes. 2011. Trace ethnography: Following coordination through documentary practices. In *2011 44th Hawaii international conference on system sciences*. IEEE, 1–10.

[15] Leo A Goodman. 1961. Snowball sampling. *The annals of mathematical statistics* (1961), 148–170.

[16] Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, and Gillian R Hayes. 2016. Digital footprints and changing networks during online identity transitions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2895–2907.

[17] Donna J Haraway and Thyrza Nichols Goodeve. 2018. *Modest_witness@ second_millennium*. routledge.

[18] Jeffrey H Kahn, Renee M Tobin, Audra E Massey, and Jennifer A Anderson. 2007. Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American journal of psychology* 120, 2 (2007), 263–286.

[19] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th acm conference on recommender systems*. 11–14.

[20] Calvin A Liang, Sean A Munson, and Julie A Kientz. 2021. Embracing four tensions in human-computer interaction research with marginalized people. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 2 (2021), 1–47.

[21] Ania Loomba. 2015. *Colonialism/postcolonialism*. Routledge.

[22] Nicole Novielli, Fabio Calefato, and Filippo Lanubile. 2018. A gold standard for emotion annotation in stack overflow. In *2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR)*. IEEE, 14–17.

[23] Carsten Østerlund, Kevin Crowston, and Corey Jackson. 2020. Building an apparatus: Refractive, reflective, and diffractive readings of trace data. *Journal of the Association for Information Systems* 21, 1 (2020), 10.

[24] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.

[25] Morgan Klaus Scheuerman, Jacob M Paul, and Jed R Brubaker. 2019. How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–33.

[26] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 5412–5427.

[27] Shilad Sen, Margaret E Giesel, Rebecca Gold, Benjamin Hillmann, Matt Lesicko, Samuel Naden, Jesse Russell, Zixiao Wang, and Brent Hecht. 2015. Turkers, scholars," arafat" and" peace" cultural communities and algorithmic gold standards. In *Proceedings of the 18th acm conference on computer supported cooperative work & social computing*. 826–838.

[28] Christopher Su. 2016. pyquora. https://github.com/csu/pyquora. PyPI: https://pypi.org/project/quora/, last accessed: June 18, 2020.

[29] Harsh Suri et al. 2011. Purposeful sampling in qualitative research synthesis. *Qualitative research journal* 11, 2 (2011), 63.

[30] Yazhe Wang, Jamie Callan, and Baihua Zheng. 2015. Should we use the sample? Analyzing datasets sampled from Twitter's stream API. *ACM Transactions on the Web (TWEB)* 9, 3 (2015), 1–23.

[31] Lei Yang and Xavier Amatriain. 2016. Recommending the World's Knowledge: Application of Recommender Systems at Quora. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 389–389. Supplementary materials: talk: https://www.youtube.com/watch?v=VJOtr47V0eo, presentation: https://www.slideshare.net/nikhildandekar/scaling-recommendations-at-quora-recsys-talk-9162016.

[32] Ussama Yaqub, Nitesh Sharma, Rachit Pabreja, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2020. Location-based sentiment analyses and visualization of Twitter election data. *Digital Government: Research and Practice* 1, 2 (2020), 1–19.