AuthCrowd: Author Name Disambiguation and Entity Matching using Crowdsourcing

António Correia[*1], Diogo Guimarães[1], Dennis Paulino[1], Shoaib Jameel[2], Daniel Schneider[3], Benjamim Fonseca[1], and Hugo Paredes[1]

[1] INESC TEC and University of Trás-os-Montes e Alto Douro, UTAD, Quinta de Prados, Apartado 1013, Vila Real, Portugal
[2] School of Computer Science and Electronic Engineering, Colchester Campus, University of Essex, UK
[3] Tércio Pacitti Institute of Computer Applications and Research, NCE/UFRJ, Rio de Janeiro, Brazil
[*] Corresponding author. Email address: antonio.g.correia@inesctec.pt

Abstract—**Despite decades of research and development in named entity resolution, dealing with name ambiguity is still a challenging issue for many bibliometric-enhanced information retrieval (IR) tasks. As new bibliographic datasets are created as a result of the upward growth of publication records worldwide, more problems arise when considering the effects of errors resulting from missing data fields, duplicate entities, misspellings, extra characters, etc. As these concerns tend to be of large-scale, both the general consistency and the quality of electronic data are largely affected. This paper presents an approach to handle these name ambiguity problems through the use of crowdsourcing as a complementary means to traditional unsupervised approaches. To this end, we present "AuthCrowd", a crowdsourcing system with the ability to decompose named entity disambiguation and entity matching tasks. Experimental results on a real-world dataset of publicly available papers published in peer-reviewed venues demonstrate the potential of our proposed approach for improving author name disambiguation. The findings further highlight the importance of adopting hybrid crowd-algorithm collaboration strategies, especially for handling complexity and quantifying bias when working with large amounts of data.**

Keywords—*author name disambiguation; crowdsourcing; entity matching; evaluation; scientometrics; task design*

I. INTRODUCTION

The assessment of research activity is a complex endeavor due to the large variability among entities with the same meaning that are expressed in different forms across digital libraries and scientific repositories. In fact, this is a long-standing challenge in the field of scientometrics [1], where a lot of inconsistencies can occur when considering the large-scale quantitative analysis of entities like institutions, authors, countries/regions, and keywords. As a result, such problems can be particularly troublesome in investigations of citation and co-authorship networks [2, 3]. For instance, determining whether two or more papers were written by the same unique author is a non-trivial and error-prone task even with the best algorithms available [4]. This problem has become a significant concern because of the accelerated rate of growth in scientific production, with implications for search engines and database indexing when taking into consideration the ever-larger number of overlapping author names, terms, and taxonomic relations across heterogeneous sources.

In response to such limitations, there is now considerable effort dedicated to developing algorithms and models (e.g., [5]) in order to solve ambiguities every time a new entry is written to the metadata registry of a certain data repository.

Nonetheless, despite the scientific community has avidly pursued a solution for solving this problem by exploring new unsupervised, semi-supervised, and supervised strategies [4], much less attention focuses on improving such disambiguation processes through human collaboration and crowd support. That is, although machine learning (ML) and non-ML (e.g., graph- and heuristic-based) techniques have been widely used for named entity disambiguation problems, the author name disambiguation taxonomy proposed in [4] does not consider the use of human-involved methods and techniques at a large scale (i.e., crowdsourcing) despite some recent advances in this direction [6]. This is where we come back to see the problem in the debate – a failure to incorporate crowd-level inference and decision-making in the loop of scientometrics.

The main driver behind this paper is to demonstrate an alternative reliable solution for the problem of named entity resolution at the author level through the development of new strategies that enable to solve ambiguities in affiliation data. Our goal is to achieve fast and accurate contributions from a crowd of non-experts and experts (i.e., researchers) that could be used to improve the way as current disambiguation tasks are performed. Our insight is that such a diverse crowd can execute tasks like determining whether two papers are written by the same author by using simple online search or their own expertise. Given this context, we are interested in answering the following question: Is crowdsourcing a reasonable method for solving the problem of author name disambiguation? The primary goal of this paper is to address this issue by proposing a crowd-based system and looking at some empirical evidence on how a crowd performs in detecting author and affiliation name ambiguity accurately. With this in mind, we set out to investigate some patterns and causal relationships resulting from human-data interactions and participants' perspectives and experiences on the system interface by considering the impact of changes on user behavior through log analysis.

The remainder of the paper is organized as follows. A brief review of prior related work is presented in Section II. Section III describes the general architecture of AuthCrowd system as a crowd-based approach intended to support author name disambiguation. Section IV provides the experimental settings and results of a case study. A discussion of the main findings achieved in this paper is also presented in this section, including some design recommendations. Finally, the paper concludes in Section V with a summary of the main insights.

II. RELATED WORK AND BACKGROUND

Named entity disambiguation is a complex resolution task which requires applying knowledge to distinguish entity mentions and link them to their corresponding entries [7]. Thus, as a variance of the entity disambiguation problem, author name disambiguation can be understood as the task of disambiguating unique named entity mentions whereby two or more authors share identical names [5]. Some errors discussed in the literature during the last years range from problems in transcribing large document collections [1] to namesake alias, homonymy or polysemy (when the same name corresponds to multiple authors), and name variability or synonymy (when an author appears under different names) [8]. Other common issues reported include missing identifiers, lack of standardized schemas, and inconsistencies in data representation [1, 9].

A step in the direction of understanding the complexity of the disambiguation tasks in scientometrics has been done by Tekles and Bornmann [10], who compared a collection of unsupervised approaches used in the past and demonstrated great results when using rule-based scoring and clustering. Most of the existing methods that are available to handle author named disambiguation try to build a similarity graph using clustering techniques and classifiers able to identify and predict relationships between entities that share the same name (including co-authorship networks) [6]. Besides, context-based word and network embedding methods have gained a lot of attention in the academia and industry due to their ability to learn representations while applying neural networks to disambiguate word senses and predicting future occurrences [11]. To account for this, Wang and co-authors [5] proposed a unified framework for author name disambiguation based on adversarial representation learning. This network embedding approach is particularly useful in large-scale heterogeneous networks based on low-dimension representations.

While research on author name disambiguation has focused extensively on unsupervised models and algorithms that use similarity measures without the need of pre-labeled training data, such kind of techniques can be affected by limitations on the accuracy due to missing attributes [1]. On the other hand, supervised models require pre-labeled data to train ML algorithms able to group similar entities based on clustering techniques. Concurrently, the use of crowdsourcing in such tasks remained less investigated and has not been addressed at the crossroads of bibliometric-enhanced IR extensively [12]. Within this context of application, Cheng and co-authors [13] published one of the first known studies using discriminative feature labeling and crowdsourcing for author name disambiguation. In [14], the name ambiguity problem was addressed through the use of crowdsourced topic annotations, while a human-in-the-loop visualization system [15] was proposed to support end-users in validating and iteratively refining ambiguous cases. In a more recent study designed in part to create a large-scale and manually-labeled dataset of author names, Xiao et al. [6] framed their findings under a crowdsourcing annotation framework. Nevertheless, in spite of the positive results obtained with the experiments, a lot of open-ended issues remain to be investigated taking into consideration the potential of crowdsourcing and human-AI collaboration strategies in the context of name ambiguity.

III. AUTHCROWD SYSTEM

Regarding the difficult task of cleaning the raw data during the preprocessing stage of a scientometric workflow [12], we propose a system that directly addresses the inadequacies of name ambiguity by crowdsourcing disambiguation tasks. Thus, we assume that a crowd-based approach can offer stable performance for coping with data quality dimensions such as consistency, accuracy, completeness, and uniqueness [16]. The architecture of the AuthCrowd system is shown in Fig. 1 and builds upon the acquisition of publication records as the input information, primarily for its disambiguation and correct formatting. The integration with digital libraries is the starting point and essential for the functioning of the whole system. As soon as the disambiguation module obtains relevant results to proceed with the analysis, a crowdsourcing campaign is created using a crowdsourcing platform that serves as a link between the AuthCrowd system and the pool of crowd participants available to contribute in the human-in-the-loop cycle of disambiguation.

The platform consists of three core modules: (i) *Task Design*, (ii) *Validation*, and (iii) *Aggregation*. This is in line with Bhatti and colleagues [17] who argue that these are key elements in crowdsourcing, together with other aspects such as quality control and incentive mechanisms. In general terms, *Task Design* consists of elaborating the crowdsourcing task interface that will be shown to crowd participants. In this context, the PYBOSSA[1] platform was chosen as it enables the implementation of a task presenter tailored for crowdsourcing campaigns. This platform comes with an API that allows its integration for managing tasks and importing the results. Furthermore, PYBOSSA gives support to enrich the task presented to the user by including a built-in mechanism that passes information to the task presenter. Thus, the outputs of the completed tasks can be retrieved and aggregated during the task run. Concerning the AuthCrowd system, the task interface is built based on the presentation of one or more scientific publications together with response fields (predefined choice or open answers), and indicators about the confidence level of each decision.

Consequent upon this, the *Validation* process allows to select the crowd participants who are really focused on performing such tasks (e.g., using ground truth questions). Then, the responses provided are used to infer the result to be returned from the campaign during the *Aggregation* process. At this level, different aggregation techniques can be applied, as scrutinized in [17]. Some examples of these aggregation methods include but are not limited to majority voting (the best result is the most frequent answer among the crowd workers), average of results (a single aggregated answer is obtained from the average outputs), HoneyPot (a non-iterative
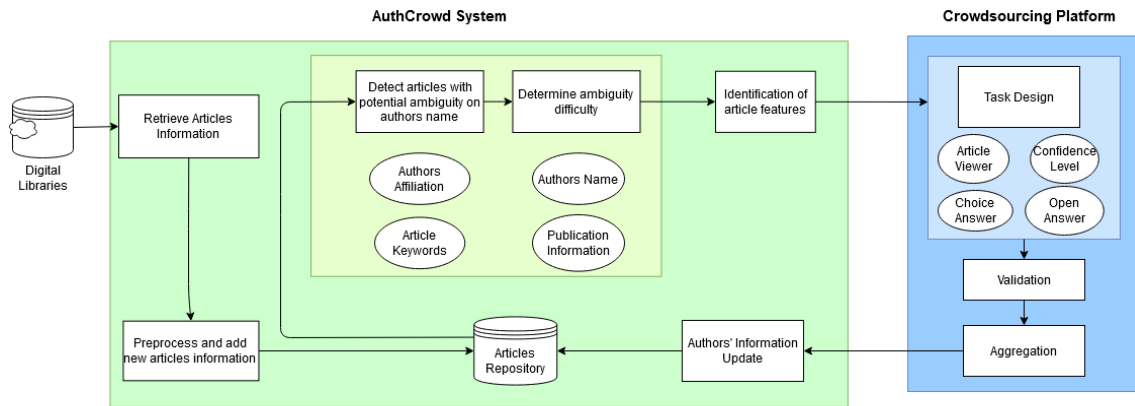
---

[1] https://pybossa.com.

Fig. 1. Architecture and workflow of the AuthCrowd system.

aggregation technique intended to capture deviant or malicious behaviors through a pre-processing step that excludes responses from untrusted crowd workers), and expert label injected crowd estimation (a probabilistic approach that relies on the inputs provided by a crowd of experts). After the completion of the crowdsourcing campaign, the results obtained are integrated into the AuthCrowd system's database using a supervised model approach. As an iterative process involving human oversight during the entire cycle, we must assume that all the steps are prone to errors and thus we need to ensure that a requester who launched the campaign is able to relaunch another campaign with different task properties when the results obtained are below the expected outcomes.

IV. CASE STUDY AND DISCUSSION

Here the main objective is to conduct a crowdsourcing case study to validate the proposed prototype system when performing crowdsourced author name disambiguation tasks. In particular, we are interested in discovering more about the usefulness of certain task parameters and interface elements taking into consideration their possible effects on the crowd performance for the particular problem at hand. To this end, we report on some results and observations derived from a set of experiments carried out using the AuthCrowd system.

A. Experimental Design

An experimental design was used to study the influence of different task properties on the outputs of a crowdsourcing campaign in the context of author name disambiguation and user-driven metadata correction. Grounding our research in a set of well-known task characteristics with different levels of complexity [18], we formulate our hypotheses about how these factors may influence the performance of participants in such scenarios as follows:

- *Article view*: The type of visual representation that is shown to the user at the article level may affect the performance of crowd workers;
- *Confidence level*: The confidence degree specified by users during the execution of a task taking into account their own perception of the accuracy of a response can help to estimate what will be their performance;

- *Response format*: The type of responses that is presented to the crowd workers (e.g., open question, true or false) can influence their performance.

In order to evaluate these properties, a total of eight disambiguation tasks were created. We designed two different types of tasks of varying complexity. The first type consists of determining if two articles are written by the same author. This arrangement allows the simultaneous presentation of two different articles, where each participant is asked if a certain author has authored or co-authored both publications. For this type of task, we studied the property of viewing the article using two distinct representations, where the list of authors can appear either in a separated form or interspersed with the affiliation. Moreover, we also studied the degree of confidence in the decision through a scale from 1 to 5 (5 being the highest value). The second type of task is based on identifying certain properties of affiliation and authorship metadata, including two tasks to determine who is the corresponding author and one task intended to transcribe a complete author's affiliation using an open-answer response format. With respect to the complexity and duration of execution, the tasks were designed to be easy to understand in a way so that they would not take more than five minutes from the pre-questionnaire to the post-questionnaire. In Fig. 2 it is shown an example of a task interface designed for the crowdsourcing experiment.

Fig. 2. Main interface of the crowdsourced task designed for identifying if two articles are written by the same author.

The interface was developed in HTML5 and Javascript. From a visualization perspective, the task window shows a question related with the disambiguation problem as depicted by Label 1 in Fig. 2. Each task presents at least one *Article Viewer*, which shows the full-text of the publication (Label 2). To help navigate through an article, the interface provides buttons to change the current page and allows zooming in and out on the current display (Labels 3 and 4, respectively). Consequent upon this, users can be asked to complement their choice of answer with a subjective opinion regarding their level of confidence in the decision made (Label 5). Finally, the task can be submitted through the answer buttons as shown in Label 6 (Fig. 2).

B. Procedure

The target population of this study comprised people with an interest and/or background/technical expertise in scientific writing. Recruitment was carried out by sending research cooperation request letters through email invitations to mailing lists from universities and research centers. The invitation contained an explanation of the case study and detailed instructions on how to proceed with the experiment through a link to a crowdsourcing campaign hosted on PYBOSSA. Therefore, participants were able to register in the system and no compensation was provided for their participation.

Each participant had to fill a pre-questionnaire prior to the study, assessing general knowledge about their previous experience on name ambiguity problems (familiarity) and their proficiency in performing online tasks, along with an understanding of the participants' demographic profile. The first two questions were presented on a 4-point Likert scale ranging from strong agreement to strong disagreement. Table I lists the tasks performed by participants, including the type of task, the author name to be disambiguated, the type of visualization used, and the level of difficulty of a task.

TABLE I.   OVERVIEW OF TASK DESIGN AND ATTRIBUTES.

task id	type of task [a]	author name	article #1		article #2		difficulty [c]
			paper id	visualiz. [b]	paper id	visualiz.	
1	T1	Alireza Abbasi	P1	NAS	----	----	C
2	T2	David Randall	P2	NAR	P3	NAS	A
3	T2	Marcos Borges	P4	NAS	P5	NAS	B
4	T3	Niels van Berkel	P6	NAS	---	---	C
5	T2	Weiming Shen	P7	NAS	P8	NAR	A
6	T1	Colin Venters	P9	NAS	----	----	C
7	T2	David Randall	P10	NAS	P11	NAR.	B
8	T2	Liam Bannon	P12	NAR	P13	NAR	A

[a.] Type of task that the participants were asked to perform: identifying the corresponding author (**T1**), determining if two documents are written by the same author (**T2**), and finding information about the affiliation of the main author (**T3**).

[b.] Overall criteria of the type of visualization at the author level: name and affiliation of the authors are shown in the same row (**NAR**), or in separate rows (**NAS**).

[c.] Levels of difficulty for each task assessed as: hard (if the author has a different affiliation in the two articles) (**A**), medium (if the author has similar names with little variations in both articles) (**B**), and easy (if the participant had some previous experience working with scientific literature) (**C**).

A log with the actions taken by participants was kept confidential for research purposes only by recording the number of times a user clicked on the interface during task execution. After the completion of the tasks, each user was asked to fill in a post-questionnaire adapted from the validated questionnaire for user interaction satisfaction (QUIS) [19]. In addition, we used a control question to check whether a participant knew some of the author names displayed in any of the eight crowdsourcing tasks.

C. Findings

A total of twenty-four participants took part in the crowdsourcing experiment individually. The experiment was conducted online, and no group interactions were studied. Furthermore, the sample of this study consisted in individuals who had a higher education diploma or any other degree in any polytechnic or university. An analysis of the demographic data collected shows that the majority of responses (58.34%) came from individuals with some prior background in disambiguation issues. The pre-questionnaire also reveals that a large proportion of the sample has proficiency in performing online tasks, where around seventy-one per cent of the respondents strongly agreed with this item.

1) Task Results

The observations revealed that all of our participants were able to successfully work on the AuthCrowd system. Fig. 3 plots the results of the analysis for each of the five tasks assessing the ability to determine if two articles were written by the same author. Looking at the results in detail, we notice that two tasks achieved a mean accuracy higher than 75%. We found a significant association between correctness and a mean average confidence level for positive answers. In such cases, the greater levels of accuracy were strongly associated with a high number of clicks made by users as measured by totalizing the counts (>=6 clicks).
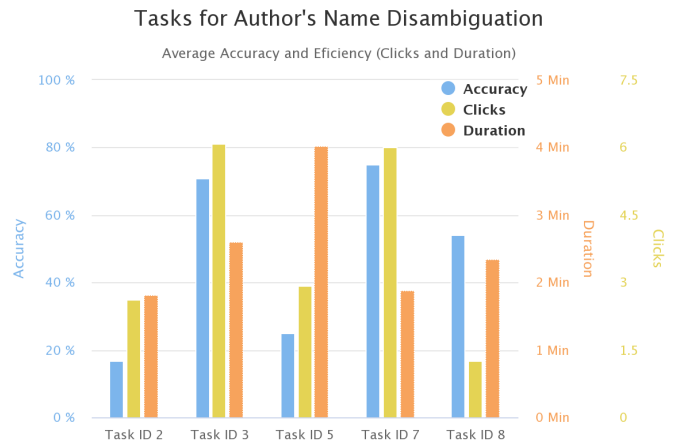

Fig. 3.   Column chart of average results on crowdsourcing tasks containing disambiguation problems at the author level.

When considering the tasks with the lower accuracy levels, a possible explanation could lie in the fact that the authors had different affiliations in the periods of enrollment when these papers were published. To a large extent this can also explain why the participants used a lower number of clicks and a minimum hold time in these tasks since they possibly had the first impression that the two articles were written by two

different authors. Concerning the aspects related to the task duration, we also found a possible relationship between the time spent in a task and the level of difficulty and type of visual representation. Among the tasks performed during the experiment, it is observed that the final task carried out by participants obtained satisfactory results although the lowest number of clicks which can be partially related with the order in which the tasks were displayed.

Extrapolating to the second type of tasks concerned with identifying the corresponding author and finding information about the affiliation of the main author, we explored the performance of users and then we came to a set of polarized results within the spectrum of corresponding author identification. For instance, as shown in Fig. 4, more negative results were obtained in the first task of the experiment when considering the accuracy of crowd participants. This may have been caused by the inexperience of the users and the lack of clear instructions for dealing with the interface of the system.
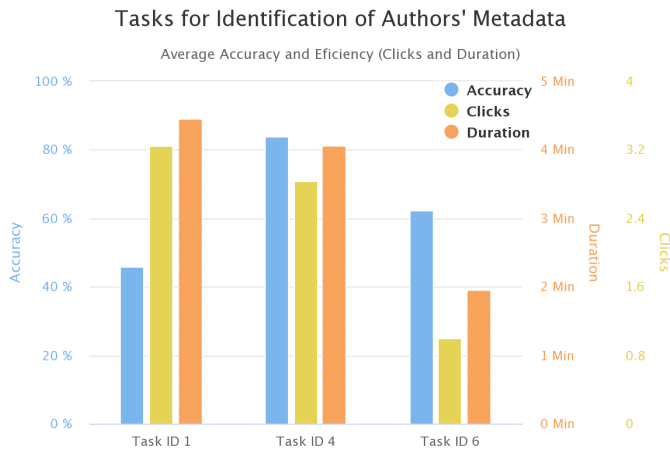

Fig. 4. Experimental results of the participants' accuracy in three different crowdsourced tasks related to identifying affiliation and authorship metadata.

On the other hand, some significant results were found associated with the task of identifying the affiliation of the first author of a publication, reaching an overall accuracy score higher than 80%. A particular aspect of this task is that it is the only one where the participant needed to fill a text box. In view of this, the cause of such result might be related to the features of highlighting and copying text in the article viewer directly to the answer text box.

2) User Click Behavior

In this study, we also examined the effect of participants' click behavior to gain a more insightful perception on their actions during the task execution. As it can be observed from Fig. 5, the average number of clicks on the article viewer buttons when manually changing the pages of a document (*ArVisPdf*) decreases progressively as the participant performs the crowdsourcing tasks. As the accuracy did not exert a visible effect on these button actions, a possible insight into this phenomenon is given by the fact that participants might have started to get more experience with the interface. Through log analysis, we also found that the article viewer

zoom buttons (i.e., resolution) had a higher number of average clicks (*AvZoomButtons*) in tasks where the main goal is the recognition and validation of authorship between two options. Based on our results, we can assume that this phenomenon could have occurred because of the type of visual representation used (i.e., NAS on the left side of the screen, accompanied by NAR on the right). Response buttons (*AvRespButtons*) were pressed equally and remained constant during the entire experiment with an expected value of at least one click per task. Although each participant was allowed to submit one response only, the transitions between tasks might have a delay of a few seconds so the user may have tried to click in the answer buttons multiple times.
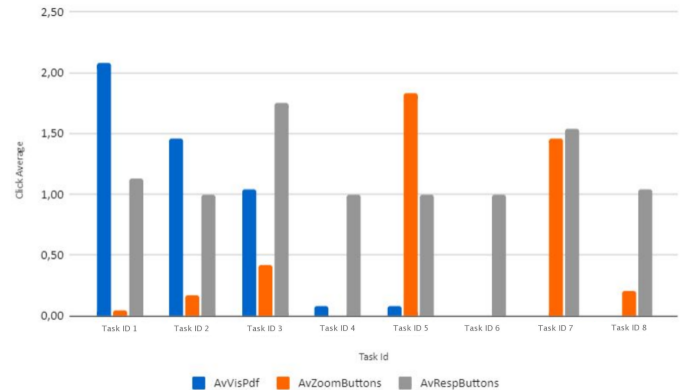

Fig. 5. Mean proportion of clicks per participant during the experiment across each of the crowdsourcing tasks.

3) Perceptions on the System Usage

After performing the tasks, participants filled out a post-use questionnaire based on Likert questions to provide their feedback on AuthCrowd. Since the feedback from users represents an important factor in the evaluation of a crowdsourcing system, the participants were presented with an adapted version of the QUIS [19] tailored to the analysis of their views and experiences regarding the use of AuthCrowd for disambiguating author names and identifying metadata descriptors as expressed in bibliographic records. Table II summarizes the responses to the individual questions rated by participants from 1 to 5, according to the strength of agreement among users.

TABLE II. AVERAGE LIKERT SCORES FROM POST-USE QUESTIONNAIRE; 1: STRONGLY DISAGREE, 5: STRONGLY AGREE.

Question [a]	Average Likert score
Overall reaction to the AuthCrowd system itself	3.96
Ease of use	4.04
Perceived utility	4.17
Interface	3.88
Reading characters on the screen	3.17
Highlighting simplifies task	4
Organization of information	4.38
Terminology related to task	4.21
Prompts for input	3.96
Performing tasks is straightforward	4.29
AuthCrowd speed	4.38
AuthCrowd reliability	4.29
Correcting your mistakes	4
Designed for all levels of users	3.50

[a.] The questionnaire used for validating AuthCrowd was designed on the basis of the four-factor model provided by Chin and co-authors [19], which forms the foundation of the questionnaire for user interface satisfaction.

Overall, respondents expressed positive views about the prototype's usefulness and ease of use. Consistent with these findings, the general reaction to the AuthCrowd system was favorable on average, although there were some lower values regarding the participants' ability to read characters on the screen and the lack of appropriate guides for all levels of users. However, as reported by one respondent, *"zoomin in when comparing two papers side to side [...] causes the papers to overlap in the screen"*. At a certain point, this overlapping can become an obstacle and pose significant difficulties to read some crucial text to properly answer the questions. On the other hand, participants found the system fast and the organization of information very simple and easy to understand. Despite the positive perceptions about the velocity of the system, the waiting time between pages was reported to be a little bit too long, as stated by a participant. Another aspect that can be observed from the data is that the terminology used in the system is consistent and complies with the domain standards. Moreover, participants expressed generally positive thoughts on the feature of highlighting text as a means of supporting the decision-making process. The results of our post-questionnaire also revealed the same positive opinions with regard to the reliability of the system, the ability of correcting mistakes, the straightforwardness of the tasks, and the overall impression of the interface.

V. CONCLUSIONS AND FUTURE WORK

This paper examines the problem of name ambiguity in bibliometric-enhanced IR, in particular within the author name disambiguation domain, by exploiting the potential benefits of using crowdsourcing for such type of content analysis. Under this perspective, we developed the "AuthCrowd" prototype, a crowd-powered system where authors and the general public worldwide are able to contribute by fixing duplicates, missing data, incorrect references, misspellings, and synonyms. Thus, in this study, we sought to examine whether differences in interface design features and task presentation within a crowdsourcing campaign would yield performance gains. At first sight, we believe that this study can draw useful insights into the design of crowdsourcing systems for disambiguation purposes with different levels of complexity. Based on our results, we see significant potential in scaling up such tasks using quality control strategies based on both ML and crowdsourcing (e.g., gold standards, crowd worker reputation) for ensuring accuracy. This can be particularly useful to avoid several problems that arise from data incompleteness and inconsistency hindering its practical application to obtain trustworthy representations based on observational data while understanding the rationale behind decisions.

ACKNOWLEDGEMENTS

REFERENCES

[1] C. A. D'Angelo and N. J. van Eck, "Collecting large-scale publication data at the level of individual researchers: A practical proposal for author name disambiguation," *Scientometrics*, vol. 123, pp. 883-907, 2020.

[2] C. K. Singh, D. Vasques Filho, S. Jolad, and D. R. O'Neale, "Evolution of interdependent co-authorship and citation networks," *Scientometrics*, vol. 125, no. 1, pp. 385-404, 2020.

[3] A. Correia, S. Jameel, D. Schneider, B. Fonseca, and H. Paredes, "The effect of scientific collaboration on CSCW research: A scientometric study," in *Proceedings of the IEEE International Conference on Computer Supported Cooperative Work in Design*, pp. 129-134, 2019.

[4] I. Hussain and S. Asghar, "A survey of author name disambiguation techniques: 2010-2016," *Knowledge Engineering Review*, 32, e22, 2017.

[5] H. Wang, R. Wan, C. Wen, S. Li, Y. Jia, W. Zhang, and X. Wang, "Author name disambiguation on heterogeneous information network with adversarial representation learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 238-245, 2020.

[6] Z. Xiao, Y. Zhang, B. Chen, X. Liu, and J. Tang, "A framework for constructing a huge name disambiguation dataset: Algorithms, visualization and human collaboration," *arXiv preprint arXiv:2007.02086*, 2020.

[7] S. Cucerzan, "Large-scale named entity disambiguation based on Wikipedia data," in *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 708-716, 2007.

[8] D. K. Sanyal, P. K. Bhowmick, and P. P. Das, "A review of author name disambiguation techniques for the PubMed bibliographic database," *Journal of Information Science*, 0165551519888605, 2019.

[9] P. Km, S. Mondal, and J. Chandra, "A graph combination with edge pruning-based approach for author name disambiguation," *Journal of the Association for Information Science and Technology*, vol. 71, no. 1, pp. 69-83, 2020.

[10] A. Tekles and L. Bornmann, "Author name disambiguation of bibliometric data: A comparison of several unsupervised approaches," *Quantitative Science Studies*, pp. 1-19, 2019.

[11] B. Zhang and M. Al Hasan, "Name disambiguation in anonymized graphs using network embedding," in *Proceedings of the ACM Conference on Information and Knowledge Management*, pp. 1239-1248, 2017.

[12] A. Correia, S. Jameel, D. Schneider, H. Paredes, and B. Fonseca, "A workflow-based methodological framework for hybrid human-AI enabled scientometrics," in *Proceedings of the IEEE International Conference on Big Data*, pp. 2876-2883.

[13] Y. Cheng, Z. Chen, J. Wang, A. Agrawal, and A. Choudhary, "Bootstrapping active name disambiguation with crowdsourcing," in *Proceedings of the ACM International Conference on Information and Knowledge Management*, pp. 1213-1216, 2013.

[14] X. Sun, J. Kaur, L. Possamai, and F. Menczer, "Ambiguous author query detection using crowdsourced digital library annotations," *Information Processing & Management*, vol. 49, no. 2, pp. 454-464, 2013.

[15] Q. Shen, T. Wu, H. Yang, Y. Wu, H. Qu, and W. Cui, "Nameclarifier: A visual analytics system for author name disambiguation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 141-150, 2016.

[16] M. Venanzi, J. Guiver, G. Kazai, P. Kohli, and M. Shokouhi, "Community-based Bayesian aggregation models for crowdsourcing," in *Proceedings of the International Conference on World Wide Web*, pp. 155-164, 2014.

[17] S. S. Bhatti, X. Gao, and G. Chen, "General framework, opportunities and challenges for crowdsourcing techniques: A comprehensive survey," *Journal of Systems and Software*, 110611, 2020.

[18] R. T. Nakatsu, E. B. Grossman, and C. L. Iacovou, "A taxonomy of crowdsourcing based on task complexity," *Journal of Information Science*, vol. 40, no. 6, pp. 823-834, 2014.

[19] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 213-218, 1988.