Age-Based Task Specialization for Crowdsourced Proofreading

Masatomo Kobayashi, Tatsuya Ishihara, Toshinari Itoko,
Hironobu Takagi, and Chieko Asakawa

IBM Research – Tokyo, 5-6-52 Toyosu, Koto, Tokyo 135-8511, Japan
{mstm,tisihara,itoko,takagih,chie}@jp.ibm.com

Abstract. Crowdsourcing can efficiently produce accessible digital books for people with print disabilities. However, particularly in Japan, the proofreading step tends to be expensive because of language-related issues. The elderly population is a promising source of proofreaders. Our surveys found that they have strong linguistic skills and want to contribute to society. So why do they rarely participate in Internet-based work scenarios such as crowdsourcing? We introduce a collaborative crowdsourcing model that aims to fully utilize the linguistic skills of the elderly by encouraging younger people to support the elderly in overcoming their limited technical skills. We decompose each proofreading task into several types of sub-tasks, where some tasks require more linguistic skills while the other tasks need more technical skills, so that the linguistic and technical tasks can be distributed to older and younger participants, respectively. We also discuss other scenarios that may be suitable for such multi-generational crowdsourcing model.

Keywords: Accessibility, Micro-tasks, Crowdsourcing, Collaboration, Elderly, Intergenerational Communications.

1 Introduction

Digital books in accessible formats such as DAISY (Digital Accessible Information SYstem) can greatly help people with print disabilities quickly and easily access a broad range of information. Print disabilities include blindness, dyslexia, and so on. Converting printed materials into accessible digital books tends to be expensive. One reason is that the OCR (Optical Character Recognition) step causes errors and therefore manual proofreading is needed. For this reason community-based approaches have been introduced to improve the efficiency of the process and to increase the number of accessible digital books. For example, many volunteer proofreaders participate in Bookshare [1] and they have prepared more than a 100,000 DAISY books. CONCERT [2] is a crowdsourcing platform for OCR proofreading with a Web-based interface, where the completed error correction process is decomposed into character-, word-, and page-level micro-tasks. They suggested that their approach could improve the overall productivity of community-based proofreading.

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2013, Part II, LNCS 8010, pp. 104–112, 2013.

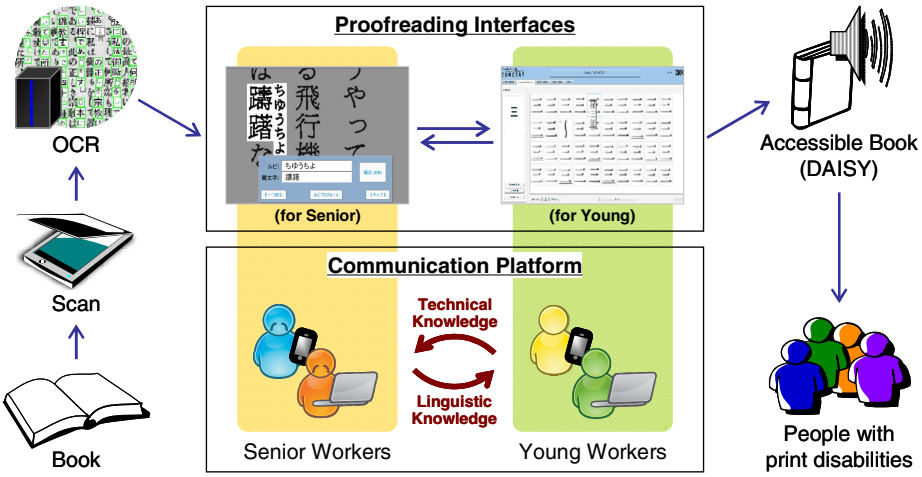


Fig. 1. An overview of the multi-generational collaborative crowdsourcing for proofreading of digital books

For books written in Japanese special care is needed because of the special characteristics of the language. First, there tend to be more OCR errors because words are not separated by spaces and Japanese uses thousands of Chinese-derived characters (*Kanji*) as well as native Japanese and imported Latin characters. Second, the correction of higher-grade *Kanji* characters and their readings is often difficult, even for native Japanese speakers and calls for expert knowledge. Although we have examined a crowdsourcing approach to proofread Japanese books by extending the CONCERT platform [3], finding an appropriate community of workers, who have both linguistic and technical skills, is a challenge.

We are developing a new approach to organize a worker community for the crowdsourced proofreading of digital books by mixing older and younger participants. The literature (e.g., [4]) indicates that semantic memory, including vocabulary and specialized knowledge, does not diminish with age, though the performance of tasks that require close attention declines. Another article [5] suggested that typical senior workers lack technical skills while young workers may lack language skills. The two age groups have complementary knowledge, skills, and abilities. If a whole proofreading task is decomposed into two types of sub-tasks, where one requires more linguistic skill and less technical skill and attention while the other calls for more technical skill and attention and less linguistic skill, then the sub-tasks can be effectively performed by older and younger workers, respectively, and thus the overall proofreading process can be completed by workers who have relatively average skills (Fig. 1). We aim to expand the scope of crowdsourcing by increasing the diversity of workers and providing them with tasks and interfaces tailored to fit their diverse traits. Although in the future detailed personalization will be enabled based on skill assessment technologies, in this particular work we focus on the typical characteristics that are often found in older populations.

This paper is organized as follows. In the next section we review related work. Then we summarize our questionnaire-based survey. We next describe our multi-generational crowdsourcing model for correcting OCR errors with a supplementary social networking platform to support mutual learning and decision making, leading to a discussion of other types of proofreading scenarios. The possibilities for collaboration between desktop and mobile workers are also discussed.

2 Related Work

In addition to CONCERT, there have been many techniques to crowdsource character recognition tasks. For example, Australian Newspapers [6] also provides a Web-based interface for OCR error correction. Workers proofread the OCR results from historic newspapers while viewing the original scanned images. Lang and Rio-Ross [7] used crowdsourcing to transcribe handwritten documents. The reCAPTCHA™ service [8] is a system that distributes small pieces of character recognition tasks as CAPTCHA tests. Digitalkoot [9] tried to make OCR error correction tasks into entertainment by presenting small tasks in game-like interfaces. Overall, we feel that CONCERT's approach seems to be the most suitable for a multi-generational community, since it consists of multiple types of tasks that can benefit from different mixtures of abilities.

For intergenerational online communication, Harley and Fitzpatrick [10] suggested that video blogs can support communication between older and younger generations. Bentley et al. [11] encouraged location-oriented video communication using a map-based interface. Wen et al. [12] investigated intergenerational communication with an online gaming platform. Hiyama et al. [13] proposed a method that allows the elderly to access social media using familiar interfaces based on phone calls and text messages. To encourage multi-generational collaboration, a mechanism is needed to facilitate knowledge sharing beyond basic communication.

3 User Preference Survey

A printed questionnaire was administered to identify the preferences of the elderly for participation in crowdsourcing. We gave the questionnaire to 179 senior citizens in a mid-sized city and 170 of them responded (for a response rate of 95%). All of the respondents were Japanese and consisted of 99 males and 71 females, ranging in age from 54 to 84 (*mean*=67.7, *SD*=5.3). This survey was conducted as a part of a larger survey regarding work and information-communication technologies (ICT).

For the question that asked if they wanted to participate in crowdsourcing, the majority of the respondents (100) answered that they wanted to participate if their requirements were met. The top two requirements were: (i) they be provided with ICT instructions before participation (75); and (ii) they be offered tasks that match their skills and interests (66).

Another question asked them to choose tasks they wanted to support from among candidate tasks we selected from previous studies. The respondents primarily preferred these tasks: 49 of them chose transcription of handwritten documents

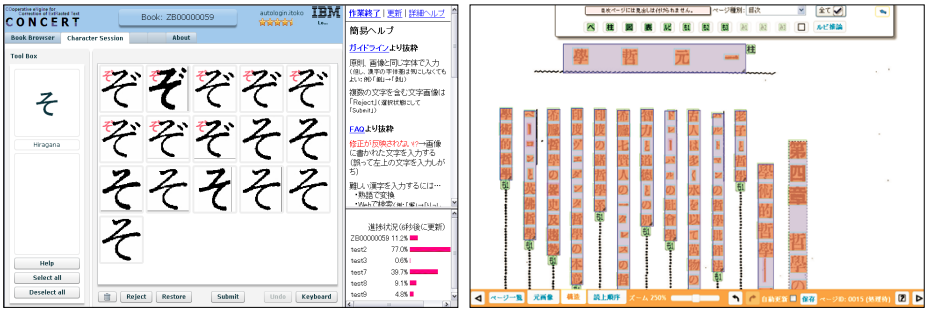


Fig. 2. Screenshots of (a) character-level and (b) page-level interfaces

(e.g., [7]), 46 chose revising sentences and correcting typos (e.g., [14]), 41 chose proofreading digital books for people with visual impairments (e.g., [1]), and 34 chose captioning videos for people with a hearing loss (e.g., [15]). In contrast, they tended to avoid the following tasks: translating English documents into Japanese (e.g., [16]) or adding annotations to make accessible webpages for screen reader users (e.g., [17]) (18 respondents for each). Overall, they prefer the tasks that require skills in their native language and that help people with disabilities, while they did not favor tasks that called for skills in a foreign language or ICT.

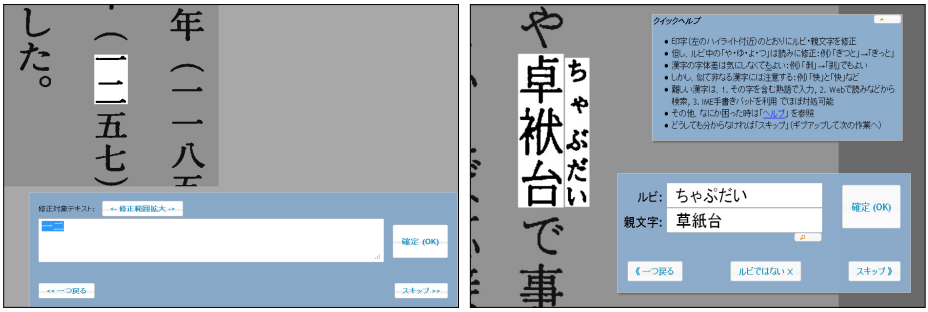
These results encourage us towards involving senior citizens in proofreading tasks on behalf of people with print disabilities. We are now working on a multi-generational collaborative approach to allow young workers with good ICT skills to assist seniors.

4 Multi-Generational Collaborative Crowdsourcing

4.1 Crowdsourcing Sub-tasks for Proofreading

The main idea of crowdsourced proofreading is that the time required for processing a book can be drastically reduced by allowing multiple workers to process decomposed sub-tasks in parallel. There are several methods for decomposing a proofreading process.

The crowdsourcing platform that we used in [3] involved an interface for character-level OCR correction tasks (Fig. 2-a) and a page-level interface for correction and annotation (Fig. 2-b). In the character-level interface, which was described in [2], each worker sees a set of scanned images of characters as thumbnails that were recognized by the OCR as a certain character. The character that the OCR recommended as the "answer" is shown in the upper-left area in the screen. The worker is asked to choose any character that is different from the "answer", i.e., any incorrectly recognized characters. The worker is allowed to optionally type a correct answer. The submit button confirms the results and proceeds to the next set of scanned characters.



(a) (b)

Fig. 3. Screenshots of (a) phrase correction and (b) ruby correction interfaces

In the page-level interface, each worker is asked to add structural metadata to the page. The worker adds the metadata by selecting a metadata type (e.g., heading, table of contents, etc.) from the toolbar and drawing a box on the scanned page to specify the target for the metadata.

The character-level task exploits the human capability to detect mismatched patterns at a glance, but does not require much linguistic knowledge. The page-level task requires a certain level of visual attention and knowledge about the conventions of page structures.

We added two additional interfaces that require less visual attention and ICT skill but more linguistic knowledge. The phrase correction task (Fig. 3-a) shows each worker a small piece of a scanned page and the recognized text in an editable field with large fonts, which may contain OCR errors. The worker is asked to correct any errors by editing the text. The ruby correction task (Fig. 3-b) uses a similar interface. It shows a pair of some main text and the corresponding ruby text in editable fields, where the ruby text typically represents the reading of the base text (for details about ruby, see [18]). Either or both of the base and ruby texts may contain OCR errors. Complex *Kanji* characters are often recognized incorrectly and advanced knowledge is required to specify the correct readings of such complex characters. Workers are allowed to skip any task if they do not have sufficient knowledge about the characters.

As shown in Fig. 1, our main idea of multi-generational crowdsourcing is to route the tasks that require linguistic knowledge to senior participants, who are expected to have better linguistic knowledge as mentioned in Section 1, while the tasks that call for ICT skills or visual attention go to younger participants, who have better ICT skills and visual attentions. Thus each worker does not need to be an expert proofreader who has all of the key skills. Of course a young worker might do the senior-favored tasks if that worker has special aptitude, and vice versa. Among these four interfaces, the character-level interface would generally be more suitable for younger participants while the phrase and ruby correction interfaces would be more suitable for senior participants. The page-level interface is most suitable for experts only, since it requires both visual attention and knowledge about publications.

4.2 Communication Platform

The communication platform is a key component for collaborative crowdsourcing, allowing co-workers to share information that helps other people. They also need to discuss problems that appear while doing tasks [17]. In a multi-generational collaboration for proofreading, the platform has two primary roles. First, it helps the participants share the problems they encounter during their tasks and make decisions about how to address the problems. For example, in a preliminary study using our prototype system, the workers found that the instructions we gave them did not tell them how to deal with special characters such as symbols and spaces. They needed to have discussions about the rules for those characters. The second role is allowing senior workers to ask how to use the interfaces. Our questionnaire-based survey presented in Section 3 and informal interviews showed that they need instructions. A previous study [19] confirmed that senior citizens prefer to learn ICT based on instructions from manuals or teachers, rather than by trial-and-error approaches. Having a platform that allows them to ask for instructions whenever they need help will encourage senior citizen participation. The platform can also allow younger participants to ask the seniors for linguistic answers. Also it may support free conversations so that older and younger participants can know about each other.

5 Implementation and Preliminary Performance Evaluation

We implemented the proofreading interfaces as Web-based applications. The participants can do the tasks using their own desktop or laptop computers running a standard Web browser. The current implementation has the character-level, phrase correction, and ruby correction interfaces. The page-level interface was not included, since it requires expert skills and thus is not expected to be suitable for crowdsourcing to workers with ordinary skills. A portal page was created to allow the participants to access each task interface and the communication platform. For the communication platform, standard open-source software for social networking and learning management were used to allow the participants to communicate with each other and share instructional materials. Note that post-processing is needed to make fully-accessible digital books, so that expert proofreaders can verify the output of the crowd and correct the kinds of errors that are not handled by the crowdsourcing interfaces. We plan to invite the experts to use the communication platform to share their knowledge. People with print disabilities will also be involved to give the participants feedback from the end-users.

Prior to the deployment to multi-generational communities, we examined the performance of crowdsourced proofreading as a preliminary evaluation of the proofreading interfaces. More than 100 volunteers from the authors' company participated in the session. They were all adults less than 65 years old, and had good linguistic knowledge and expert-level ICT skills. The session lasted for about 4 months and processed a total of 50 books. Most of the books were from 100 to 300 pages long. The crowd took up to 4 days to process one book, with an average working time of 2

days. This was quite rapid compared to traditional proofreading processes that often take several months for each book.

6 Discussion

6.1 Multi-generational Crowdsourcing for General Word Processing

In Section 5, we discussed an OCR error correction scenario based on the multi-generational crowdsourcing model. The model can be used for broader applications as long as larger tasks can be decomposed into sub-tasks and each sub-task requires the knowledge, skills, and abilities of specific types of workers.

One example is crowdsourced word processing. Bernstein et al. [14] proposed to decompose word processing tasks such as proofreading and editorial suggestions into three sub-tasks: Find, Fix, and Verify. The last task, Verify, asks workers to point out inappropriate options from improvement candidates. This would be more suitable for seniors because it does not require as much ICT skill but uses linguistic knowledge. The Find and Fix sub-tasks, where workers are asked to find areas for improvements and generate the improvement candidates, calls for more visual attention and ICT skill.

Another strategy to exploit multi-generational communities might be adding an additional step, such as Classify, after Find. The new step might ask workers to classify the candidate areas for improvement into *easy* and *difficult* tasks. The *easy* tasks, e.g., basic typos, can be dispatched to younger workers while the *difficult* tasks, e.g., sensitive expressions, would be sent to the more senior workers.

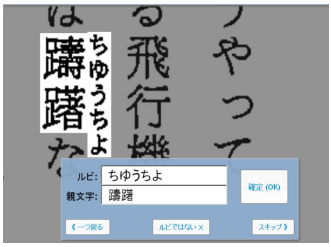
6.2 Collaborative Crowdsourcing between Desktop and Mobile Workers

We are considering supporting participation with mobile devices such as smartphones and tablet devices, although the current prototype system supports only desktop browsers. It is known that mobile touchscreen interfaces are preferred by the elderly [20] and these kinds of devices are becoming increasingly popular among senior citizens.

However, it is also known that Japanese text entry using software keyboards is difficult and confusing for senior citizens because of the frequent mode-switching while inputting thousands of characters. Modern products support voice input based on speech recognition technologies, but this is often inadequate for complex *Kanji* characters and their readings. Hence advanced collaboration strategies are needed.

We expect a two-dimensional classification of interfaces will address these problems (Table 1). The mobile interfaces only ask whether or not the OCR result is correct, whereas the desktop interfaces also ask for the correct answers when the OCR results are incorrect. Desktop workers might see only the tasks marked by the mobile workers as “incorrect”. As a result, the number of tasks for the desktop workers is reduced while the mobile workers are not troubled by text entry.

Table 1. Two-dimensional classification of collaborative crowdsourcing interfaces for older and younger workers, and desktop and mobile workers

	Desktop	**Mobile**
Senior	 Linguistic Knowledge, Editable	 Linguistic Knowledge, Decisions Only
Young	 Visual Attention, Editable	 Visual Attention, Decisions Only

7 Conclusion

This paper described a collaborative crowdsourcing strategy for proofreading OCR results, which aims to efficiently exploit the capabilities of both older and younger participants. The tasks that require linguistic knowledge will be dispatched to senior workers while the tasks that call for visual attention or technical skills should be dispatched to young workers. A collaboration platform is also provided to support mutual learning and decision making. We plan to deploy the proposed system in a multigenerational community to assess the effectiveness of our collaborative crowdsourcing model in practical use.

Acknowledgments. This research was partially supported by the Japan Science and Technology Agency (JST) under the Strategic Promotion of Innovative Research and Development Program.

References

1. Bookshare, `http://www.bookshare.org/`
2. Neudecker, C., Tzadok, A.: User Collaboration for Improving Access to Historical Texts. Liber Quarterly 20(1), 119–128 (2010)

3. Ishihara, T., Itoko, T., Sato, D., Tzadok, A., Takagi, H.: Transforming Japanese Archives into Accessible Digital Books. In: Proc. JCDL 2012, pp. 91–100. ACM (2012)
4. Ouchi, Y., Akiyama, H. (eds.): Gerontology – Overview and Perspectives, 3rd edn. Univ. of Tokyo Press (2010) (in Japanese)
5. Staffing Mature Worker Survey, http://www.goldenworkers.org/images/publication/mature_workers_survey_2012_adecco.pdf
6. Holley, R.: Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers. National Library of Australia Staff Papers (2009)
7. Lang, A.S., Rio-Ross, J.: Using Amazon Mechanical Turk to Transcribe Historical Handwritten Documents. The Code4Lib Journal (2011)
8. von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M.: reCAPTCHA: Human-Based Character Recognition via Web Security Measures. Science 321(5895), 1465–1468 (2008)
9. Chrons, O., Sundell, S.: Digitalkoot: Making Old Archives Accessible Using Crowdsourcing. In: Proc. HCOMP 2011 (2011)
10. Harley, D., Fitzpatrick, G.: YouTube and Intergenerational Communication: the Case of Geriatric1927. UAIS 8(1), 5–20 (2009)
11. Bentley, F.R., Basapur, S., Chowdhury, S.K.: Promoting Intergenerational Communication through Location-Based Asynchronous Video Communication. In: Proc. UbiComp 2011, pp. 31–40. ACM (2011)
12. Wen, J., Kow, Y.M., Chen, Y.: Online Games and Family Ties: Influences of Social Networking Game on Family Relationship. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 250–264. Springer, Heidelberg (2011)
13. Hiyama, A., Nagai, Y., Kobayashi, M., Takagi, H., Hirose, M.: Question First: Passive Interaction Model for Gathering Experience and Knowledge from the Elderly. In: Proc. PerCol 2013. IEEE (in press, 2013)
14. Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K.: Soylent: a Word Processor with a Crowd Inside. In: Proc. UIST 2010, pp. 313–322. ACM (2010)
15. Sobhi, A., Nagatsuma, R., Saito, T.: Collaborative Caption Editing System – Enhancing the Quality of a Captioning and Editing System. In: Proc. CSUN 2012 (2012)
16. Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality using Amazon's Mechanical Turk. In: Proc. EMNLP 2009, pp. 286–295. ACL (2009)
17. Takagi, H., Kawanaka, S., Kobayashi, M., Ito, T., Asakawa, C.: Social Accessibility: Achieving Accessibility through Collaborative Metadata Authoring. In: Proc. ASSETS 2008, pp. 193–200. ACM (2008)
18. Ruby Annotation, http://www.w3.org/TR/ruby/
19. Leung, R., Tang, C., Haddad, S., McGrenere, J., Graf, P., Ingriany, V.: How Older Adults Learn to Use Mobile Devices: Survey and Field Investigations. TACCESS 4(3), Article 11 (2012)
20. Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M., Ifukube, T.: Elderly User Evaluation of Mobile Touchscreen Interactions. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 83–99. Springer, Heidelberg (2011)