

# Lessons Learned from Crowd Accessibility Services

Hironobu Takagi , Susumu Harada, Daisuke Sato and Chieko Asakawa

IBM Research – Tokyo, 5-6-52 Toyosu, Koto-ku, 135-8511 Japan

{takagih, haradas, dsato, chie}@jp.ibm.com

**Abstract.** Crowd accessibility services for people with disabilities, driven by crowd-sourcing methods, are gaining traction as a viable means of realizing innovative services by leveraging both human and machine intelligence. As the approach matures, researchers and practitioners are seeking to build various types of services. However, many of them encounter similar challenges, such as variations in quality and sustaining contributor participation for durable services. There are growing needs to share tangible knowledge about the best practices to help build and maintain successful services. Towards this end, we are sharing our experiences with crowd accessibility services that we have deployed and studied. Initially, we developed a method to analyze the dynamics of contributor participation. We then analyzed the actual data from three service deployments spanning several years. The service types included Web accessibility improvement, text digitization, and video captioning. We then summarize the lessons learned and future research directions for sustainable services.

**Keywords:** Crowd-sourcing, accessibility, digital book, captioning, Web accessibility

## 1 Introduction

Accessibility technologies have advanced significantly over the years, but there still remain many issues that technology alone has not been able to solve. Screen readers and refreshable Braille display technologies opened up a wide array of information to people with visual impairments, but areas such as the automatic conversion of visual information into textual descriptions remain major challenges. Automatic speech recognition and remote captioning technologies have expanded the communication possibilities for people who are deaf or hard of hearing, but for informal speech or conversations in noisy environments, automatic recognition has yet to attain practical levels of reliability.

In recent years, there have been rapid developments of crowd-sourcing technologies to bridge the gaps in mechanical computing technologies, based on harnessing and combining the perceptual, cognitive, and intellectual abilities of human beings. Pioneering services such as the ESP Game [1] have demonstrated the power of this approach. Crowd accessibility services have emerged as a category of crowd-sourcing service focused on supporting people with disabilities. General crowd-sourcing ser-

vices sometimes are designed to train machine intelligence systems instead of supporting humans. For example, building a dictionary for machine translation can be crowd-sourced to people without directly specifying the human beneficiaries or their specific objectives.

In contrast, crowd accessibility services focus on supporting specific people with disabilities. A captioning service seeks to support people who are deaf or hard of hearing, or a book digitalization service can focus on supporting readers and learners with visual disabilities [2, 8]. One category of new services is called “vision as a service”, and it helps blind people to recognize all of the things around them [3, 9]. The existence of human beings who need support strongly distinguish accessibility services from other services.

In various ways, crowd accessibility services can enable dream-like services for people with disabilities. However, it is sometimes difficult to sustain these services because of the small sizes of the target populations. For sustainable services, the supply of workers and the work requests should be balanced, but this is often difficult for crowd accessibility services. This characteristic generally obliges service providers to depend upon unpaid crowd volunteers.

At the same time, crowd accessibility services also have characteristics that present special opportunities. Since crowd accessibility services tend to be focused more on social services and not commercial aspects, we have learned that these contributors tend to be more motivated by altruistic factors rather than by monetary gains or by the diversion and novelty offered by quick and simple tasks, which are among the motivators for traditional crowd-sourcing services.

How can these services succeed and endure while recruiting new and active contributors? How can we motivate them? How can we insure the quality of outputs? Little work has been done to analyze the issues that are specific to crowd accessibility services, since the analysis methods themselves are challenging given the diverse characteristics of the services and the limited availability of data for such analyses.

In this paper, we offer some answers to these questions by analyzing data from three services of our own. After the review of related work, an analysis method is proposed by taking into account the practical availability of data. After examining our data in the observations section, our lessons learned focus on four topics. Before offering our conclusions, a wide range of future research directions are discussed for social infrastructures based on crowd accessibility approaches.

## **2 Related Work**

### **2.1 Crowd Accessibility Services**

Crowd accessibility services are within the broad category of crowd-sourcing services, but are characterized by their focus on people with disabilities as the beneficiaries of the services. Image labeling for accessibility is one of the popular applications for crowd accessibility. Dardailler [7] pioneered the approach of labeling images using the power of the crowd in the ALT-server proposal that stores alternative text on a remote server. The ESP game [1] used the gamification approach in labeling images

to support both image searches and accessibility. This meant the service was not purely a crowd accessibility service, but suitable for general and accessibility applications. Optical character recognition (OCR) technology was combined with crowd-sourcing in the WebInsight system [4]. This was the first approach of combining human intelligence with computer intelligence to improve accessibility. Social Accessibility [17, 18] focused on comprehensive and high quality Web accessibility improvements including image labels. VizWiz [3] explored real-time labeling by recruiting Web workers from the Amazon Mechanical Turk service to answer vision-related questions submitted as photos by blind users using smartphones. [9] is a commercial service to label photos in real-time by combining an image recognition engine with crowds of workers. This service tried to expand the scope from a pure accessibility application to a more general augmented reality application by using general purpose image recognition technologies. In spite of such technical advances such as gamification, OCR integration, authoring tools, and image recognition integration, sustainability still remains an important issue to be addressed by crowd accessibility services. Bigham et al. [5] reviewed the history of crowd accessibility services, summarizing the features and characteristics of 15 services and proposing 13 design dimensions with which to compare the services, such as motivation, accuracy, and reliability. [2, 19].

## 2.2 Methods to Analyze Participation

This section describes some methods to analyze participation in the crowd services from two perspectives, objective (such as statistics and log analysis) and subjective (such as interviews or surveys). Wikipedia is the largest service that is powered by a crowd of volunteers without monetary incentives, which has led to many analyses of its participants. According to an internal article titled "Wikipedians", there are about 16 million registered accounts and about 0.3 million of them edit Wikipedia articles on a monthly basis [20]. The number of edits per account ordered by magnitude follows a Zipf distribution, which forms a straight line on a double-logarithmic graph [10]. Swartz [16] reported that about 73.4% of Wikipedia editing is done by the top 1,400 users, and most of the remaining editing is minor, such as fixing incorrect spellings. This phenomenon is also described as a 90-9-1 rule [11], since 1% of participants make almost all of the contributions, 9% make minor contributions, and the remaining 90%, the "lurkers" [13], only read the results. Stewart et al. [15] showed that the distribution is different in a closed service limited to company employees, reporting values of 33-66-1.

Objective study is key to gaining an overall picture of the participation in these services, though subjective studies reveal many of the psychological factors driving participation. Nonnecke et al. [14] studied why lurkers lurk in crowd-sourcing services. Bryant et al. [6] interviewed active Wikipedians to analyze how they had evolved from lurkers to leading contributors of Wikipedia. A typical user initially visits Wikipedia to obtain information, then begins making minor contributions and learning about the rules and conventions of the community, and finally becomes reg-

**Table 1.** Evaluated Services

	Service type	Primary target users	Start date	Service duration	Scope	Quality assurance
EBIS	OCR correction	Blind and other print disabilities	2011/8/23	2011/12/15	Internal	Expert check
CCES	Captioning	Deaf and hard of hearing	2011/11/11	2011/12/28	Internal	Expert check
Social Accessibility	Web accessibility metadata authoring	Blind and people with visual impairments	2008/7/8	2010/3/31	External	User report

istered as a contributor. Their aim changes from just reading or polishing the articles to improving the quality of Wikipedia itself.

### 3 Evaluated Services

Over the past several years we deployed three crowd accessibility services, as summarized in Table 1. For each service, the following sections provide a summary of the service, characteristics of the supporters, task descriptions and definitions of the task units in the logged data, and methods for quality assurance.

#### 3.1 EBIS

EBIS [8] was originally created to rapidly digitize physical books for the blind and other people with print disabilities, focusing on Japanese books that have up to 10,000 characters in various styles, which makes the process quite difficult compared to languages that use phonetic alphabets. The books are unbound and scanned to create images, and then processed through an OCR engine to generate the initial digital output. The raw output from the OCR engine is not accessible, since it tends to contain various character recognition errors and structural problems. There are several steps, but we focused on a crowd-sourcing process.

EBIS is using a check-by-expert approach as the quality assurance method. A skilled contributor was assigned as a proofreader to each book to identify any errors that might have escaped the eyes of the contributors in the earlier phases. This phase was managed manually and with mail-based information exchange. EBIS used gamification based on points and grades. According to contribution, each user earned some number of points and climbed the grade ladder based on the accumulated points.

The system was announced internally within IBM as part of the company’s centennial ‘celebration of service’ event as one of the encouraged volunteer opportunities. This meant the contributors were unpaid volunteers recruited from a large pool of potential participants.

### **3.2 CCES (Collaborative Captioning Editing System)**

CCES is a crowd-sourcing system that adds captions to digital videos [12]. Adding text captions to digital video tends to be a time consuming task, especially when the task work is being performed ondone for content for which no text transcript is available. The CCES service aims to speed up this process by splitting up the video into short segments, automatically segmenting it at detected breaks between phrases, and offering a user interface through which a captioner can type in the text as they listen to the clip. Each CCES task consists of a 30-second clip of video content that is to be transcribed and submitted by the captioner. The 30-second clip is actually split into roughly 10 or so sub-clips that approximate phrase utterances, and the supporter can review and transcribe them one at a time.

The expert check method was used as the quality assurance method in CCES. The system has a user interface to allow administrators to check and fine-tune captions. For example, administrators can adjust the timing, position on screen, length of one caption, and other attributes that are used for caption color coding, such as the gender of a speaker. The system was also announced internally within IBM as part of the company's centennial 'celebration of service' event. As of the writing of this paper, the CCES service has been running for about one month in its newest version.

### **3.3 Social Accessibility (SA)**

The Social Accessibility [17] service was one of the first crowd accessibility projects we developed at our group. It provides a mechanism through which Web consumers such as blind computer users can identify accessibility problems with certain webpages, and then submit requests for improvements to a central server. The request is then made available to contributors who use a Web interface to create a fix for the accessibility problem by creating metadata that augments the original webpage, perhaps by adding alternative text to images. Since the metadata is stored on our server without altering the original website and the Web consumer can retrieve the metadata using our browser plugin, the service enables quick turnaround and wide reach.

The SA service was available to the public, with anyone able to sign up either as a requestor or a contributor. A requestor could submit a request by simply using a browser plugin and pressing a special hotkey when an inaccessible page was encountered. A contributor could view the pending requests with our online interface, and also view the webpage with the problem together with a panel for entering the various kinds of metadata information. A task in SA was a collection of metadata that a contributor submitted for one particular page in one session.

SA employed the user report as its main quality assurance mechanism. Each end user could report incorrect or suspicious errors whenever they used metadata within a seamless user interface with the same tool that applied the metadata to the webpages. Volunteers could also check and improve the quality of the metadata by using the tool that created the metadata, but this was also uncoordinated volunteer work.

## 4 Analysis Method

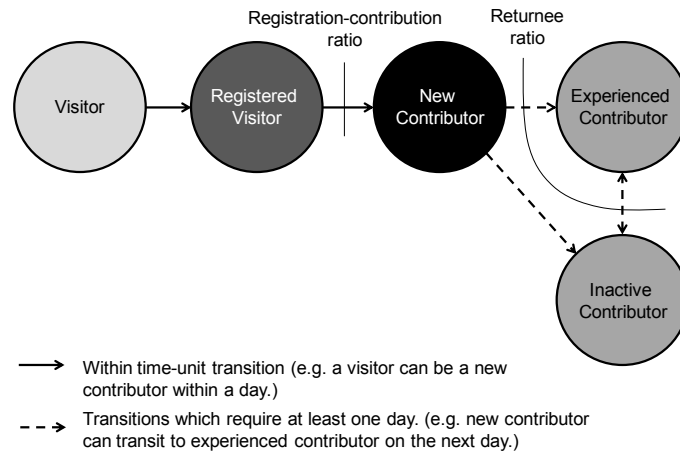
Given the diversity of crowd accessibility services, a standard analysis method itself is hard to develop. Therefore, we have developed an analysis method by focusing on the factors affecting the sustainability of crowd accessibility services. The method consists of two types of analysis to cover various aspects, such as transitions and gaps involving participation statuses (Transition Analysis).

The number of new contributors and the ratio who become committed contributors are the critical factors for assessing sustainable contributor participation. One major problem for each analysis method is the limited availability of data that applies across services. Each output of a comparative analysis method should be based on data that is available for all of the crowd accessibility services.

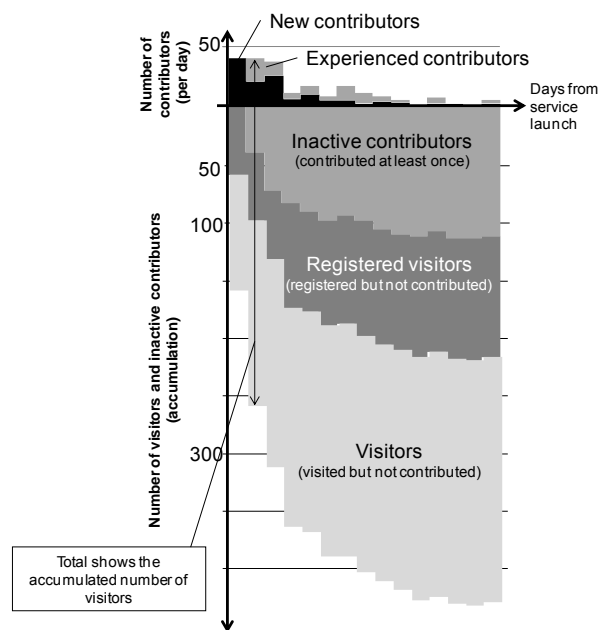
To determine specific challenges for the services, the gaps (or barriers) for participants should be assessed. To increase the number of new contributors, we should determine the gaps facing each new contributor. To increase the number of committed contributors, the gaps in returning should be studied. The transition analysis was designed as a tool to identify these gaps within services. The dynamics of participation can be considered as a series of transitions among participation statuses. Fig. 1 shows our definition of the transitions and Fig. 2 is a bar chart of a model of the statistical transitions among the statuses. The total number of people in all of the statuses refers to the total number of participants at the end of that day.

**Visitors** are people who visited the service, but did nothing. This is the initial state for a participant. If a person finds and visits the service, but decided not to contribute, then the person's status is "visitor". A **registered visitor** is a person who has registered with (or subscribed to) the service but has not yet contributed. This level of involvement is close to the definition of "lurkers" in social media (Section 2.2). In the case of a crowd accessibility service, people may not be simply lurking, but actively trying to figure out ways to contribute. Ideally, this status should be minimized. The number can be easily calculated from the contributor table in each service. Usually, the contributor table records the registration date as a property. This same data can be used to count the number of changes in the status of a registered user.

A **new contributor** is a person who is contributing to the service for the first time and an **experienced contributor** is someone who has contributed to the service at least twice. In Fig. 2, all of the contributors are new contributors on the first day, and after the second day there will be a mix of returnees (experienced contributors) and new contributors. These two types of contributors indicate different steps in the service. The gap between registered visitors and new contributors reflects the difficulties in starting to contribute after subscribing to the service. An iceberg chart visually shows the time changes of the ratio between registered visitors and newcomers (bottom-side dark gray and upper-side black areas). The gap can be also measured as a ratio:



**Fig. 1.** Transition Diagram among Participation Status



**Fig. 2.** Iceberg Visualization of Transitions (Contributor Status Chart)

**Registration-contribution ratio** = (total number of contributors) / (total number of registered visitors)

In contrast, the number of returnees (experienced contributors) is a metric to assess the degree of activity of committed contributors. By checking the changes over time

in the numbers of new contributors and experienced contributors, we can visually assess the ratio of returnees. Thus, the returnee ratio from new contributors to experienced contributors can be seen as a metric to assess the difficulty of the transition to become a committed contributor.

**Returnee ratio** =  $((\text{total number of contributors}) - (\text{number of one-day contributors})) / (\text{total number of contributors})$

The definitions of new contributors and experienced contributors take into account the ease-of-calculation based on the properties of output data in a usual crowd accessibility service. For example, metadata is the output in SA, and each piece of metadata has properties including a link to its creator and its creation date. Any service with functions to track “who” worked on “which” task units will have such a data table.

An **inactive contributor** is a person who has not done any work on a given day. This definition means that the active contributor statuses (new and experienced) are temporary and that most of the contributors are in this inactive contributor status.

#### 4.1 Task Analysis

The number of completed tasks in a unit of time is a clear quantitative metric for the activity level of a service. This task analysis focuses on measuring and visualizing the dynamics of completed tasks by new contributors and experienced contributors separately.

The definition of a task unit varies according to the target of a service. One transcribed line of text can be one task unit for captioning services, and one confirmed character can be one task unit for an OCR correction service. The definition can also be affected by the characteristics of the task management for the service. For example, when an OCR correction service manages correction work in units of “one page” then each page can be a task unit. A page can be an acceptable unit as long as the granularity of the task unit is sufficiently fine for the analysis.

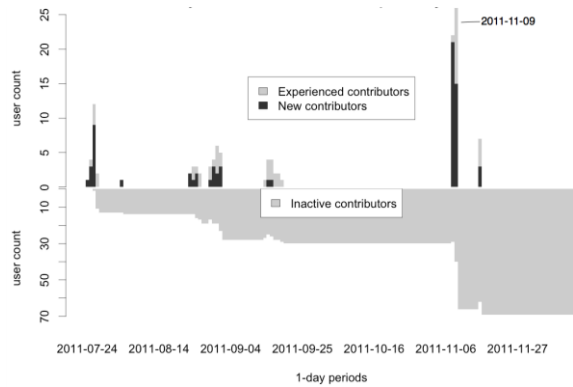
## 5 Observations

We examined three of our crowd accessibility services using the analysis methods described in Section 3. In this section, we will present the analysis of the data for each service. The results will be summarized in the following section as a set of lessons learned.

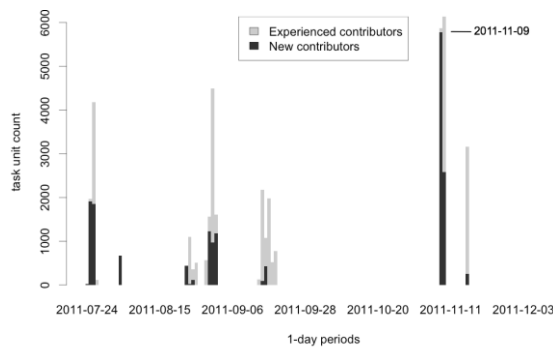
### 5.1 EBIS

Fig. 4 shows the analysis results for EBIS. Fig. 4(b) shows several multi-day peaks in the completed tasks. The Within each of these periods, contributors successfully processed 10 to 25 books. The task shortage was strongly linked to the very success of the service. Contributors actively worked and completed all of the available tasks





**Fig. 4(a).** Contributor Status Chart for EBIS (2011/7/24 ~ 2011/12/14)



**Fig. 4(b).** Aggregate Task Count Chart for EBIS (2011/7/24 ~ 2011/12/14)

in a short period of time. As seen in the aggregate task count chart in Fig. 4(b), the EBIS service had other periods when there were no tasks being performed by the contributors. The charts show that work typically occurred in spurts spanning two to four days. The contributor data shows that majority of the contributors worked for no more than three days, roughly corresponding to the average duration of a spurt. The contributor status charts in Fig. 4(a) also show that when the second spurt occurred around November 8, 2011, almost all the contributors were new first-time contributors, meaning that almost none of the previous contributors returned.

The first spike on November 8 was created by an announcement to all employees on the front page of a portal. The small peak on November 17 was created without any announcement, but just by people who discovered the availability of the service. That is why the majority of those contributors were returnees, as shown in Fig. 4(b). The last peak was created by an email announcement to experienced contributors about the availability of new books. In this way, the service successfully engaged with the experienced contributors and they completed the tasks much more quickly than the administrators could prepare additional books.

The underlying cause of the task shortage was the bottleneck due to expert phases in the process. The availability of books was limited by the performance of the experts, and this meant the contributors had to wait until each batch of processed books was proofread before being able to continue with the next batch. This can also be seen in Fig. 4(b), showing that when a new batch of books became available on November 8, 2011 (after a 2-month hiatus), almost all of the tasks were completed by new contributors and few of the previous contributors returned.

We received various informative comments as feedback from the contributors. One category of comments was about the effectiveness of the gamification feature that allowed the contributors to climb up a grade ladder by completing more tasks (Section 3.1). Some of contributors mentioned that the indication of the remaining task units to go up a level was highly motivating. Not only because of the explicit gamification feature, but many contributors mentioned the task itself gave them an impression of a “game” and that helped them to concentrate on the tasks. Many contributors mentioned that the user interface was similar to the “brain training games” on mobile game consoles (such as the Nintendo DS).

## 5.2 CCES

Fig. 5(a) shows the contributor status chart for CCES. It can be seen that throughout the duration of the service, there was a relatively steady flow of contributors performing tasks each day, both new and experienced contributors. A survey was conducted to get feedback from contributors. Among all of the contributors, 92.2% of the contributors answered good (51.0%) or very good (41.2%). The other options are 5.9% neutral, 0% bad and 0% very bad with 2.0% no answer. This high ratio of acceptance backs up the result of the steady flow of contributors.

The dips around November 26, 2011 and December 4, 2011 coincided with weekends, reflecting the fact that most contributors performed the tasks on weekdays. The graph also shows that each day there were at least some new contributors joining to perform tasks (indicated by the black bars). Such a participation pattern indicates a successfully operating service, with contributors’ task output remaining relatively steady and new contributors steadily joining the service. Fig. 5(b) shows the same tendency as Fig. 5(a). New and experienced contributors share the tasks in a balanced way. The direct interpretation of the observed data is that the service was not sticky enough to encourage contributors to complete many tasks, even for experienced contributors. Instead, many new people joined the service and each quickly completed a small number of tasks. The phenomenon was not statistically significant and we clearly need more data to study such differences.

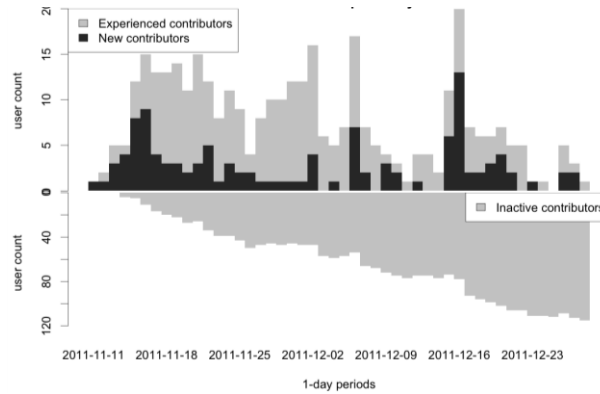


Fig. 5(a). Contributor Status Chart for CCES (2011/11/11 ~ 2011/12/28)

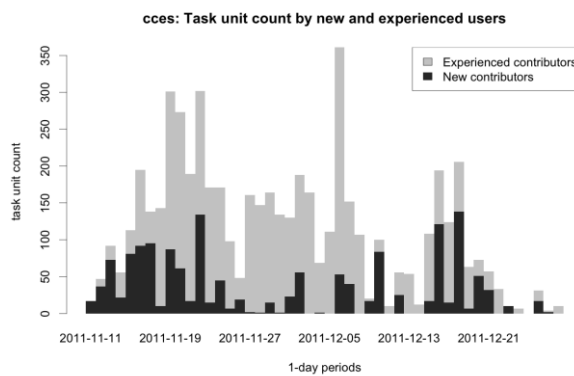
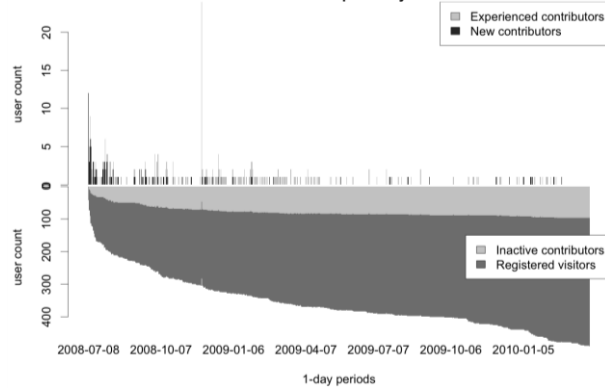


Fig. 5(b). Aggregate Task Count Chart for CCES (2011/11/11 ~ 2011/12/28)

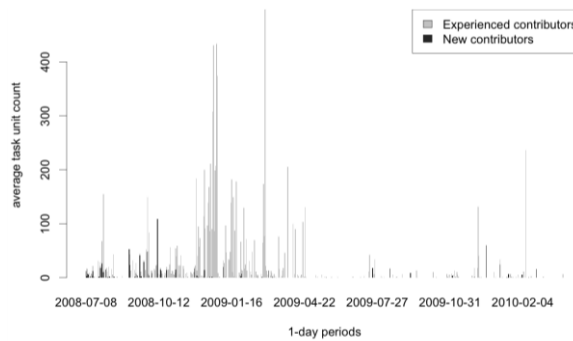
### 5.3 Social Accessibility (SA)

Two key characteristics of the SA service that distinguish it from the other services we have deployed were the length of time the service was available and the contributor pool. As seen in Fig. 6(b), the SA service continued to observe contributor activity for well over six months. When the SA service was launched, it was announced not only within Japan but also globally, resulting in nearly 700 registered users during the deployment period (Table 2). Among the three crowd accessibility services we deployed, SA had the largest number of contributors (98), with the largest number of contributors who contributed for more than two days (52).

The SA service was also able to stimulate a surge in contributor activity when an improved user interface was announced on November 27, 2008 (Fig. 6(a)). The surge was mainly due to experienced contributors who returned to the service as shown in



**Fig. 6(a).** Contributor Status Chart for SA (2008/7/8 ~ 2010/3/29)



**Fig. 6(b).** Aggregate Task Count Chart for SA (2008/7/8 ~ 2010/3/29)

**Table 2.** Contributor Status Transition Table

	Registered visitors	Registration-contribution ratio	Contributors (at least once)	Returnee ratio	Experienced contributors
EBIS	340	21%	71	41%	29
CCES	209	56%	116	52%	60
Social Accessibility	690	14%	98	53%	52

Fig. 6(b). This highlights the important impact that usability enhancements can have in stimulating contributions.

While the SA service continued to yield contributions for over six months, the conversion from registered users to contributing users was relatively low compared with the other services, 14% (Table 2). From the contributor data, a majority of contributors contributed less than 10 days total. This suggests that the prolonged continuation of contributions may in large part have been due to a very small number of high-

ly dedicated and productive contributors. This view seems to be supported by the contributor status chart (Fig. 6(a)) which shows that for the last two-thirds of the service period, the average number of contributors per day was between one and three.

The contributor chart also shows that during this period while the contributor count remained low, the number of registered visitors was continuing to increase. What this suggests is that new people were registering for the service, but were not converting into contributors. This represents missed opportunities, and may point to some limitations of the service that were turning away new contributors or inhibiting them from taking the first step.

## **6 Lessons learned**

### **6.1 Importance of Task and User Interface Design**

One of success factors we found was the ease of completing tasks. One example is in CCES. One of the reasons for the high registration-contribution ratio (56%, Table 2) compared with other services may be due to the easy to understand task and productive user interface. The task goal of “captioning” may be familiar for people from seeing captioned TV programs. In addition, the interface makes it easy for a contributor to repeatedly listen to a short clip of the video and to type in the text, and provides convenient support for pausing and briefly rewinding the playback as the user types without requiring more explicit control by the user.

EBIS seems to have a similar situation. The registration-contribution ratio is lower than CCES but some of that is probably due to the periodic task shortages resulting from the task analysis (Fig. 4(b)). The success of the design can be seen from the rapid and large quantity of task completions, as mentioned in Section 5.1. Among the concerns expressed by the contributors, one of their highest priorities was “a rapid response from the server”. This highlights the importance of having a streamlined user interface and a responsive system to fully benefit from the potential of large numbers of human participants.

In SA, several factors might cause the low registration-contribution ratio (14%, Table 2), such as shortages of user requests, a side effect of its being a public service (while the other services are all private), and the long term of its deployment. Another factor may be the steeper learning curve compared with CCES and EBIS, since the metadata required by the SA service was quite specific to the Web accessibility domain, and good results depended upon some knowledge of concepts such as heading navigation, alternative text, etc. Among the tasks in the services, character correction (EBIS) and captioning (CCES) are relatively simple tasks and do not require unusual expertise, which made the design of their tasks and their user interfaces easier than SA.

### **6.2 Necessity of Task Shortage Management**

The importance of the management of tasks, especially for handling task shortages, was one of most important lessons we learned. SA used a request-answer concept as the basic crowd-sourcing model, but that means the model depends on the number of

requests arriving from people with disabilities. That made the system especially vulnerable to task shortages.

EBIS experienced a task shortage (Section 5.1), but the cause was different from the SA situation. The process for quality assessment (Section 3.1) became the bottleneck of this service, and resulted in task shortages. The long gaps between short spurts of contributions were due to the fact that the contributors would quickly finish processing a batch of books in a period of a few days, but the next batch could not be made available to the contributors until the post-processing had been completed for those books. This post-processing step involved a single proofreader who carefully read through the entire book. Such proofreading can only be performed by skilled contributors when high quality is a concern, and this step could not be eliminated because the partner organizations wanted high-quality results. The gap between the crowd-handled work (with EBIS) and the proofreading was large. There are other alternative methods for such quality assurance phases, but no perfect alternative exists. Majority voting is a popular method but it sometimes requires large number of contributors if the quality criteria are high.

In contrast, CCES was managed in a better way. Right from the launch of the service, the service maintainers carefully monitored the consumption rate of the content by the contributors. For instance, during the first few days after deployment, the maintainers noticed that English content was receiving relatively little attention, and decided to augment the input data feed with more Japanese content during the first content update on the first weekend (November 18). The task count on the following day (November 19) more than doubled, apparently due to the increased availability of new content.

In summary, the lessons we learned fit into four areas: (1) tasks should be designed not to cause task shortages, (2) bottleneck phases should be minimized, (3) task shortages should be carefully monitored, and (4) methods for responding to task shortages should be predefined.

### **6.3 Engagement with Continuous Contributors**

We learned that the top contributors played an important role across the services. In CCES and SA, they contributed to the services over many days, and achieved some of the highest average task completions per day. While it may not be prudent to depend on the emergence of such top contributors in every service, it is better for service owners to prepare for and take advantage of such contributors when they do appear. Identification is the first step to working with top contributors. The next step is to reach out to these top contributors. The service owners can derive various insights from the top contributors for improving their service. It may be possible to receive quite specific feedback regarding the service from the perspective of an experienced contributor that could be used to help improve the service for the rest of the contributors. They may also have developed various methods or strategies for approaching the tasks that may benefit other contributors as well.

For example, during the deployment of the SA service, we were able to meet with one of the top contributors by chance. We were able to extract various insights from

him. He was much older than we had expected, actually one of the early baby boomers. He was a retired office worker, had lots of spare time and was trying to find something meaningful for his life. SA fit well with his current objectives. His motivations for participating in the service were very personal. The service owner team could clearly understand a role model contributor.

## **7 Discussion of Future Research Directions**

### **7.1 Mechanisms to Propagate the Sense of Contributions**

One of the most committed contributors to SA was neither an engineer nor a Web professional, but he had passion and time for volunteer work (Section 6.3). Since SA was a request-based service, he commented that he could directly understand the users' problems and the value of his work. His comments suggest the importance of the special sense of altruistic contribution as an incentive in accessibility services. Crowd accessibility services are characterized by the existence of human beneficiaries, the people with disabilities who are always behind the system. Because of this characteristic, we learned that a personal sense of contribution to those who need help can be an important incentive factor. In the services we studied, CCES notified the contributors when a video was posted with their captions, EBIS notified them when a book they had worked on was uploaded to the digital library, and SA had a function to notify a contributor "when someone used the metadata you created" by using a blinking icon [18]. One of the topics we have not explored was the cultural effects of the propagation of the sense of contributions. Most of the contributors for these three services were Japanese, and therefore cultural difference among countries could not be examined in this study. We suspect that the sense of altruism may differ among cultures, so further research activities in this area would be important.

### **7.2 Mechanisms for Accessibility Skill Development**

Accessibility improvements often require specific skills for contribution. SA required basic knowledge of Web accessibility, such as appropriate alternative texts according to contexts and appropriate heading levels, and so the task familiarity for contributors was low (Section 6.1). The tasks for CCES and EBIS were more familiar than the SA work but still required specific knowledge that affected the output quality. For example, the notation of sound effects in captions, and techniques to find appropriate kanji characters from character tables. These tasks require knowledge and skill to extend them to environments for people with disabilities. Section 6.1 discussed the importance of easy-to-use interfaces and easy-to-understand tasks, but systematic mechanisms to develop "accessibility skills" in contributors offer another big area to explore.

### **7.3 Challenges of User-side Service-quality Assurance**

One of the specific challenges for crowd accessibility services is the difficulty of quality assessment on the user side. SA relied on user reporting as its most important quality assurance method (Section 3.3). User reporting is a popular method used by many social services in such forms as the “illegal content report” of most video sharing services. However, for crowd accessibility services this can be difficult to implement. For SA the quality of an alternative text can be assessed by some straightforward criteria (such as spelling mistakes, lack of clarity, etc.), but this assessment is not possible for many kinds of problems. For example, if a malicious volunteer added an alternative text such as “Cancel” to an OK button, it would be extremely difficult to detect what was wrong by pushing the button. The end users need support because of their limited cognitive abilities, and that also limits their ability to assess the quality. This is one of the characteristic challenges in some types of crowd accessibility services.

### **7.4 Influence of Unintended Fun Factors for Increasing Engagement**

We saw cases in which the unintended fun factor contributed to motivate contributors in serious accessibility services. In the CCES, the video clips were extracted from longer and sometimes enjoyable videos about substantive topics. From the comments in the survey, contributors found themselves becoming interested in the content itself as they performed the transcription tasks. A similar unintended fun factor also worked in EBIS, since in the final check the contributors read through an entire book and they could enjoy it. We found that the sense of contribution was the primary motivation for the contributors to accessibility services, but the unintended fun factor strengthened engagement with each service, thus increasing the participation. Compared to the intentional fun factors like gamification, unintended fun can easily be overlooked. The influence of unintended fun for serious accessibility services is worth investigating in future research.

### **7.5 Possibilities for Senior Citizens as Contributors**

Another implication is the importance of encouraging potential contributors. For the leading contributor mentioned earlier, we had not expected such prominent involvement by a senior citizen. This anecdotal evidence suggests that senior citizens may be especially good candidates as volunteers for crowd accessibility services. Based on this experience, we studied several crowd-sourcing tasks including accessibility work for seniors ranging in age from 60 to 80. Half of them answered that they would be willing to do such tasks “without any fee”, in contrast to a smaller percentage of younger participants who agreed to the no-fee model.



## 7.6 Possibilities for Internal Organizational Crowd-sourcing

Two out of the three services were deployed within enterprise environments. (See Section 3 and Table 1.) These environments were not open to the public, so they can be regarded as constrained environments. However, our experiments showed some advantages in deploying crowd applications within an enterprise. The first advantage is the potentially large number of contributors in a large corporation, making the deployment similar to a general public environment, but with better control. We can announce, maintain, and analyze the results by using same framework as used for public services. Second, contributors within a corporation have a strong rationale not to act maliciously or to submit low-quality work in their enterprise environment partly because of their lack of anonymity. Such activities can be tracked, which is a strong incentive for the contributors to strive for high quality work. The third point is that services will be more sustainable due to the official support from the corporation as part of its CSR (Corporate Social Responsibility) efforts.

## 8 Conclusion

This paper discussed some lessons learned through an analysis of three crowd-accessibility services. After reviewing the related work, we developed visualization techniques that reflect the status changes of the contributors and of the completed tasks. The lessons learned were considered in the observation section by highlighting four topics, (1) design of the tasks and the user interface, (2) task shortage management, (3) contributor engagements, and (4) data collection and monitoring methods. Beyond these lessons learned, challenges and future research directions were discussed. The sense of contribution can be an important incentive especially for accessibility services. Better mechanisms for developing accessibility skills are imperative to improve the quality of the services. User side quality assurance poses unique challenges for accessibility services. The influence of unintended fun factor is worth investigation to increase engagement with the services. Also, senior citizens and corporate workers can be untapped new contributing resources.

Crowd accessibility has a huge potential to improve many situations for people with disabilities by combining human intelligence and computer intelligence. We are studying various ways to make the approach more sustainable in our society. A standard analysis method is itself challenging and methods should be evolved to unveil more specific problems within services. We hope that this paper will contribute to improving crowd accessibility strategies toward the level of more reliable and sustainable services.

## References

1. von Ahn, L. and Dabbish, L. Labeling images with a computer game. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04, 319–326. (2004)
2. Benetech. Accessible books and periodicals for readers with print disabilities - Bookshare - accessible books for individuals with print disabilities. <http://www.bookshare.org/>

3. Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S. and Yeh, T. VizWiz: nearly real-time answers to visual questions. In Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, UIST '10, 333–342. (2010)
4. Bigham, J. P., Kaminsky, R. S., Ladner, R. E., Danielsson, O. M. and Hempton, G. L. WebInSight: making Web images accessible. In Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '06, 181–188. (2006)
5. Bigham, J. P., Ladner, R. E. and Borodin, Y. The design of human-powered access technology. In Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '11, 3–10. (2011)
6. Bryant, S. L., Forte, A. and Bruckman, A. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, GROUP '05, 1–10. (2005)
7. Dardailler, D. The ALT-server (“An eye for an alt”). <http://www.w3.org/WAI/altserv.htm> (1997)
8. Ishihara, T., Itoko, T., Sato, D., Tzadok, A. and Takagi, H. Transforming Japanese archives into accessible digital books. In Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '12, 91–100. (2012)
9. IQ Engines. IQ Engines: image recognition and visual search. <http://www.iqengines.com/>
10. Nielsen, J. Zipf distribution (power law) of Website popularity (Alertbox Sidebar). <http://www.useit.com/alertbox/zipf.html> (1997)
11. Nielsen, J. “90-9-1” rule for participation inequality: lurkers vs. contributors in internet communities (Jakob Nielsen’s Alertbox). [http://www.useit.com/alertbox/participation\\_inequality.html](http://www.useit.com/alertbox/participation_inequality.html) (2006)
12. Sobhi, A., Nagatsuma, R. and Saitoh, T. Collaborative Caption Editing System – Enhancing the Quality of a Captioning and Editing System. In Proceedings of the 28th Annual International Technology and Persons with Disabilities Conference. CSUN 2012. (2012)
13. Nonnecke, B. and Preece, J. Lurker demographics: counting the silent. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00, 73–80. (2000)
14. Nonnecke, B. and Preece, J. Why lurkers lurk. In Proceedings of the Americas Conference on Information Systems. (2001)
15. Stewart, O., Lubensky, D. and Huerta, J. M. Crowdsourcing participation inequality: a SCOUT model for the enterprise domain. In Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP '10, 30–33. (2010)
16. Swartz, A. Who writes Wikipedia? <http://www.aaronsw.com/weblog/howwriteswikipedia> (2006)
17. Takagi, H., Kawanaka, S., Kobayashi, M., Itoh, T. and Asakawa, C. Social Accessibility: achieving accessibility through collaborative metadata authoring. In Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '08, 193–200. (2008)
18. Takagi, H., Kawanaka, S., Kobayashi, M., Sato, D. and Asakawa, C. Collaborative Web accessibility improvement: challenges and possibilities. In Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '09, 195–202. (2009)
19. Valerie, Cherek and Ann Harrison. Bookshare crosses 100,000-member milestone - Bookshare - accessible books for individuals with print disabilities. [http://www.bookshare.org/\\_/aboutUs/2010/07/100KmemberMilestone](http://www.bookshare.org/_/aboutUs/2010/07/100KmemberMilestone) (2010)
20. Wikipedia. Wikipedia: Wikipedians - Wikipedia, the free encyclopedia. <http://en.wikipedia.org/wiki/Wikipedia:Wikipedians> (2011)