

“Holy Starches Batman!! We are getting walloped!”: Crowdsourcing Comic Book Transcriptions

Christine Samson¹, Casey Fiesler², and Shaun K. Kane¹

Department of Computer Science
University of Colorado Boulder
Boulder, CO 80309 USA

{christine.samson, shaun.kane}@colorado.edu

Department of Information Science
University of Colorado Boulder
Boulder, CO 80309 USA

casey.fiesler@colorado.edu

ABSTRACT

Comic books are among the most popular forms of popular media, but most comics are not provided in an accessible format. Creating an accessible transcript of a comic book may be more challenging than simply describing the images, as comics involve complex interplay between words and images, and often feature long-running and complex storylines. In this poster we describe a pilot study exploring the feasibility of crowdsourcing transcriptions of comic book pages. We recruited 60 crowd workers and asked them to transcribe a page of a comic book; half were told that the description was for a blind person, and half were not. We found that people who knew that they were transcribing for a blind person produced longer, more detailed descriptions. Our results also suggest that comic book knowledge may have at least some small impact on description detail.

CCS Concepts

• Human-centered computing → Accessibility technologies

Keywords

comics; crowdsourcing; fandom; visual impairment

1. INTRODUCTION

Comics are a unique source of literature and entertainment that combine text and graphics to tell a story. However, the foundation of comics—combining graphics and text—poses a problem to blind readers. Most mainstream comics are not produced in an accessible format. Furthermore, translating comic pages to an accessible representation can be challenging, as comic storytelling often involves a complex interplay between words and images [4].

In this exploratory study, we consider how we might use crowdsourcing to generate descriptions of graphics for blind comics readers when official transcripts are not available. We engaged 60 crowd workers to describe two pages from a comic book [7]. We examine how two variables might impact these descriptions: (1) how much of a comic fan the worker is; and (2) whether the worker knows the purpose of the task (i.e., that the description is for a blind reader).

The results of this study suggest important directions for future work, including how to leverage fan communities rather than paid workers for crowdsourcing. This study presents a first step towards the design of a system to support transcriptions of media from fan communities.

2. RELATED WORK

Making visual content accessible to people with vision impairments typically involves translating the visual content to audio (e.g., speech or sonification) or tactile (e.g., Braille and tactile graphics) formats [3]. Each solution has its limitations: translating to audio provides broad accessibility, but removes the spatial and visual elements that are essential to comic storytelling [5]. Translating visual art into tactile format typically requires an expert to perform the translation, and works best with simple images [4]. Our preliminary experiments in creating tactile comics have shown that the amount of text in a typical comics page prevents the text from being overlaid as Braille without significantly increasing page size. In this work, we focus on creating text transcriptions of the content, which could later be presented in the user’s preferred modality.

A significant challenge in this work is translating images into text. While solutions exist for automatically captioning images (e.g., [1]), these systems require significant training data sets, and no such training set currently exists for comic images. Another approach is to recruit crowd workers to transcribe images, which can be effective for simple images such as charts and graphs, especially when workers are given clear instructions [6]. This work contributes an exploration of crowdsourced image transcription for comic book pages, which may present challenges beyond translating charts, such as identifying characters or interpreting the story.

3. METHODS

For this study, we deployed a survey on Mechanical Turk (MTurk). Crowd workers (Turkers) were shown a two-page spread from a Batman comic book (Figure 1), and were given the instruction: “Your task is to write a text description of the actions that occur in this comic page.” We specified to Turkers that they did *not* need to transcribe any text, but rather only describe the action taking place.

We also asked Turkers to identify the characters that appear in the comic (by both their hero/villain identity and secret identity, if known—e.g., Batman/Bruce Wayne), and to report their familiarity with Batman comics (whether they had read Batman comics or seen Batman films). Finally, there were two conditions for the survey: *blind*, in which we instructed the Turker that their description was for a blind comics fan, and *non-blind*, in which we included no such instruction. We recruited a total of 60 Turkers who were paid \$1 each. Turkers were randomly assigned into the blind or non-blind conditions upon accessing the survey, resulting in 27 Turkers in the blind condition and 33 in the non-blind condition.



Figure 1. Comics panel that we showed to crowd workers [7]. Image (c) DC Comics.

One challenge of our data analysis is that we do not currently have a metric for what is a “good” or “bad” transcription, or what constitutes an appropriate level of detail for a transcribed comic. For this preliminary study, we used word count as a metric for level of detail for descriptions, as descriptions with more words suggest that the transcriber put more effort into the task and included more detail in their transcription.

Our questionnaire included two questions to approximate Turkers’ level of fandom and familiarity with the characters: (1) whether or not they reported having read a Batman comic book or had only seen a Batman movie or TV show; and (2) whether they were able to correctly identify the names and secret identities of the characters in the comic.

Using word count as a dependent variable, we performed one-way ANOVAs to determine whether these factors (blind versus non-blind survey condition, reading or not reading comics, and ability to recognize characters) had a significant impact on word count.

4. FINDINGS

Descriptions varied greatly in length and level of detail: word count ranged from 30 to 885 (mean 274, median 238). Brief descriptions included single sentences such as “batman get shot to an aircraft” or “Holy Starches Batman!! We are getting walloped!” Longer descriptions described the sequence of actions in the comic, and sometimes described each panel separately.

In analyzing this preliminary survey data, we focused on level of detail of the descriptions, approximated by word count. We hypothesized that: (1) Turkers who were told that they were describing for a blind person would provide longer descriptions; and (2) Turkers with more comics knowledge would provide longer descriptions.

Our first hypothesis was (marginally) accepted with a significance of $p = 0.05$. The mean word count of descriptions provided by participants in the *blind* condition was 323, and for those in the *non-blind* condition, 235. In other words, Turkers who were told that they were writing for blind readers produced longer descriptions than those who were not told this.

Our second hypothesis was that comics “fans” would produce more detailed descriptions. We operationalized this in two ways, by whether they reported reading comics, and whether they could correctly identify characters. For both of these metrics, our data

shows a trend in the direction we expected, though does not rise to the level of significance. Those who reported that they read comics provided descriptions with a mean word count of 304, and those who do not, 226 ($p=0.09$); those who can identify all characters correctly provided descriptions with a mean word count of 350, and those who could not, 261 ($p=0.16$).

5. DISCUSSION & CONCLUSION

These preliminary findings provide us with several potential avenues for future work.

First, it is important to note that using word count to determine what might be a “good” or “bad” description is obviously inadequate (since *amount* of detail may not be the most important or useful feature of a transcription for a blind comics reader). One next step is to use the descriptions generated by this survey to determine what kinds of descriptions are useful to blind comics fans, so that in future iterations of this work we can better determine “good” versus “bad” descriptions.

Second, though our findings regarding the effect of “fandom” level on transcription detail were not significant in this small sample, the directionality of our results are encouraging for continuing this line of inquiry moving forward. An obvious next direction is to look *outside* of MTurk for fans. Prior work shows that fan communities tend to be particularly oriented towards social justice, and in particular are concerned about issues of accessibility [2]. One motivation of this work is the intuition that not only might transcribers recruited from fandom communities provide *better* transcriptions than crowd workers, but that they might perform transcription and other accessibility-related tasks for altruistic reasons, rather than for pay.

Finally, this work begins what we hope to be a new research exploration in converting comic book stories into a more accessible format. We intend to develop a system that will allow a user to request crowd transcriptions for their comics, and perhaps even to specify the level of detail that they wish to include, so that dedicated fans could explicitly request more detailed descriptions. We may also extend this approach to transcribing other forms of media, such as movies and TV episodes. We hope that this work will contribute to making popular media more accessible to everyone.

6. REFERENCES

- [1] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *Proc. CVPR '09*, 248-255.
- [2] Fiesler, C., Morrison, S., & Bruckman, A.S. (2016). An archive of their own: a case study of feminist HCI and values in design. *Proc. CHI '16*, 2574-2585.
- [3] Jacko, J. A., Vitense, H. S., & Scott, I. U. 2002. Perceptual impairments and computing technologies. In *The Human-Computer Interaction Handbook*. L. Erlbaum Associates Inc.
- [4] Jayant, C., Renzelmann, M., Wen, D., Krisnandi, S., Ladner, R., & Comden, D. (2007). Automated tactile graphics translation: in the field. *Proc. ASSETS '07*, 75-82.
- [5] McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York: William Morrow Paperbacks.
- [6] Morash, V. S., Siu, Y. T., Miele, J. A., Hasty, L., & Landau, S. (2015). Guiding novice web workers in making image descriptions using templates. *ACM TACCESS*, 7(4), 12.
- [7] Parker, J., & Case, J. (2013). *Batman '66, # 1*. DC Comics.