10-31-2011

Using Amazon Mechanical Turk to Transcribe Historical Handwritten Documents

Andrew Lang

Joshua Rio-Ross

October 31, 2011

Using Amazon Mechanical Turk to Transcribe Historical Handwritten Documents

Andrew Lang
Joshua Rio-Ross

code{4}lib JOURNAL

Issue 15, 2011-10-31

Using Amazon Mechanical Turk to Transcribe Historical Handwritten Documents

The developing "information age" is continually unraveling new ways of discovering, presenting and sharing information. Most new academic material is digitally formatted upon its creation and is thus easy to find and query. However, there remains a good deal of material from times prior to the "information age" that has yet to be converted to digital form. Much of this material can be found in library collections—whether academic, public or private—and thus remains available only to a limited number of locals or willing-and-able sojourners. Using OCR technology, most typeset documents can be digitized and made available online; and there are several projects underway to do exactly this. However, there remains little to be done for handwritten materials. Those who own collections of handwritten documents are increasingly wanting to make the content thereof available to the general public. Unfortunately, traditional transcription models typically prove to be expensive or inefficient and pdf snapshots are not searchable. We have developed a model for digital transcription using Google Docs and Amazon's Mechanical Turk. Using this model, one can use an online workforce to efficiently transcribe handwritten texts and perform quality control at a cost much lower than professional transcription services. To illustrate the model we used Amazon's Mechanical Turk to transcribe and then proofread the Frederick Douglass Diary which we have made available on a public searchable wiki. The total cost of transcription and proofreading for the 72 page diary was less than $25.00 with some pages being transcribed and proofread for as little as $0.04. Our results show that using Amazon's Mechanical Turk holds great promise for providing an affordable transcription method for hand-written historical documents making them easily sharable and fully searchable.

by Andrew S.I.D Lang and Joshua Rio-Ross

Introduction

Transcribing often hard-to-read historical handwritten documents is a painstakingly slow and costly undertaking. This has meant that the majority of important writings are not available for public study and those

that are available are often only viewable as digitally scanned images. After more than 100 years of effort, only slightly more than fifty percent of James Madison's papers have been transcribed and published, [1] while the transcription of Thomas Jefferson's papers, begun in 1943, will take another 15 years to complete. [2] [3] The remaining untranscribed portion of Madison's and Jefferson's writings are only available as "digitally scanned images of microfilmed copies of handwritten documents." [2]

Transcription can be outsourced to a professional transcription service, but this is very costly, usually costing somewhere between $7.00 and $15.00 per hour depending on the level of service. So it is not surprising that many archivists are turning to crowdsourcing for transcription. Most current attempts at crowdsourcing are run in-house, where they provide and preserve software for end users, upkeep servers, build and maintain web sites, and have editors proofread results. All of this requires a certain level of technical knowledge and support staff, meaning the projects can become time-consuming and costly. When done correctly, such crowdsourcing efforts have been remarkably successful, but they do not come without obstacles. Daniel Stowell, director and editor of the Papers of Abraham Lincoln, is quoted in a recent New York Times article on crowdsourcing transcription, saying that volunteers produced so many errors and gaps that "we were spending more time and money correcting them as creating them from scratch." [3]

Our transcription project, the Written Rummage project, has itself been a rummage for the most efficient method of using crowdsourcing and various other Internet resources to digitally transcribe handwritten documents. Our primary objective was the successful and accurate transcription of the selected Frederick Douglass Diary documents to searchable, digital text. However, the project would be of little general interest if it could not eliminate the cost barrier that usually stands between editors and accurate transcription. We therefore resolved to develop a procedure that is cheaper to implement than other transcription services. While crowdsourcing is inherently crowded with workers, managing the workers should require relatively few people so long as transcriptions can be completed efficiently. Thus, the last objective was that Written Rummage be expedient.

Academic Crowdsourcing

Crowdsourcing is the process by which a task is outsourced to an undefined group of people (the crowd) rather than contracting professionals to accomplish that task. Crowdsourcing is used in both business and academic settings and the crowd can either be unpaid volunteers or be incentivized with micro-payments or other benefits. Crowdsourcing is therefore especially useful for tasks that need to be repeated numerous times but are beyond the skills of computers–such as classifying images, gathering data, and transcribing handwritten text.

A good example of a successful academic crowdsourcing project is Zooniverse, [4] which now consists of several separate crowdsourcing projects but originally began with just the Galaxy Zoo project. Zooniverse now has close to half a million volunteers and has led to several academic papers and scientific discoveries. [5] [6] One recent addition to the set of Zooniverse projects is Old Weather, where volunteers transcribe the location and weather from British Royal Navy ship log books from around the time of World War I. [7] To encourage participation, Old Weather leverages Google Maps technology to allow volunteers to follow the course of a ship over time as they enter data. Old weather also provides an experience where volunteers can learn about the ships, the log books and the people who kept them. As volunteers contribute they also move up in rank, with the top contributor for each ship being designated captain.

Another project of transcriptional interest is the Open Dinosaur Project, where volunteers are asked to transcribe dinosaur limb bone measurements from academic papers. Citizen scientists usually need little impetus to get them to contribute to a “dinosaur project”, because “dinosaurs are cool”, but the organizers have offered an additional incentive for participation stating that “all participants will be included as junior authors on the resulting scientific paper.” [8]

Both Old Weather and the Open Dinosaur Project have a large user base of volunteers because participation is fun, interesting, and incentivized; and in the case for the Old Weather project liberally funded. [9] An up-to-date list crowdsourcing projects of general interest can be found on Wikipedia. [10]

Crowdsourcing Digitization Projects

While the Old Weather project and the Open Dinosaur Project both use crowdsourcing for transcription, the resulting digitization is not the primary focus of these projects. The former focusing on climate change and the latter on evolution. Both projects do show that to be successful in the increasingly crowded crowdsourcing arena it helps for a project to be either of great interest, well funded, or fun. If your transcription project is none of the above, then it will take greater effort to be successful using volunteers for transcription.

One technique to increase participation in more esoteric transcription projects is “community engagement”, where organizers use various enticements to attract volunteers. Recent success stories “have been particularly adept at using social media, developing refined mechanisms for ensuring that contributions are quality assured, working with large data sets, and creating interfaces that interact in a way that reduces complexity and confusion.” [11]

A good example of using social media and other incentives for the volunteer transcription of handwritten documents is the Transcribe Bentham project, which is a “participatory project” transcribing the manuscript papers of Jeremy Bentham based at University College London. [12] The Transcribe Bentham project has had great media exposure and also a lot of success with “encouraging undergraduates and school pupils studying Bentham’s ideas… to use the site to enhance their learning experience.” [13] The project is also being aided by the Arts and Humanities Research Council who awarded the project a grant of £262,673.00. [14]

There are many other noble efforts in using crowdsourcing for the transcription of historical handwritten documents, with new ones seemingly appearing every other day. [15] One of the more recent projects aims to help transcribe the soon to be released US Census data by using an ingenious hybrid of automation and crowdsourcing, automatically recognizing handwritten text using word spotting. The current government proposed system plans to pay for “transcribing the handwritten content of the images, a task that will take thousands of trained laborers anywhere between 6 and 12 months.” [16] How successful this and other projects will be remains to be seen, though

newcomers are being aided by Open Source transcription tools such as Scripto. [17]

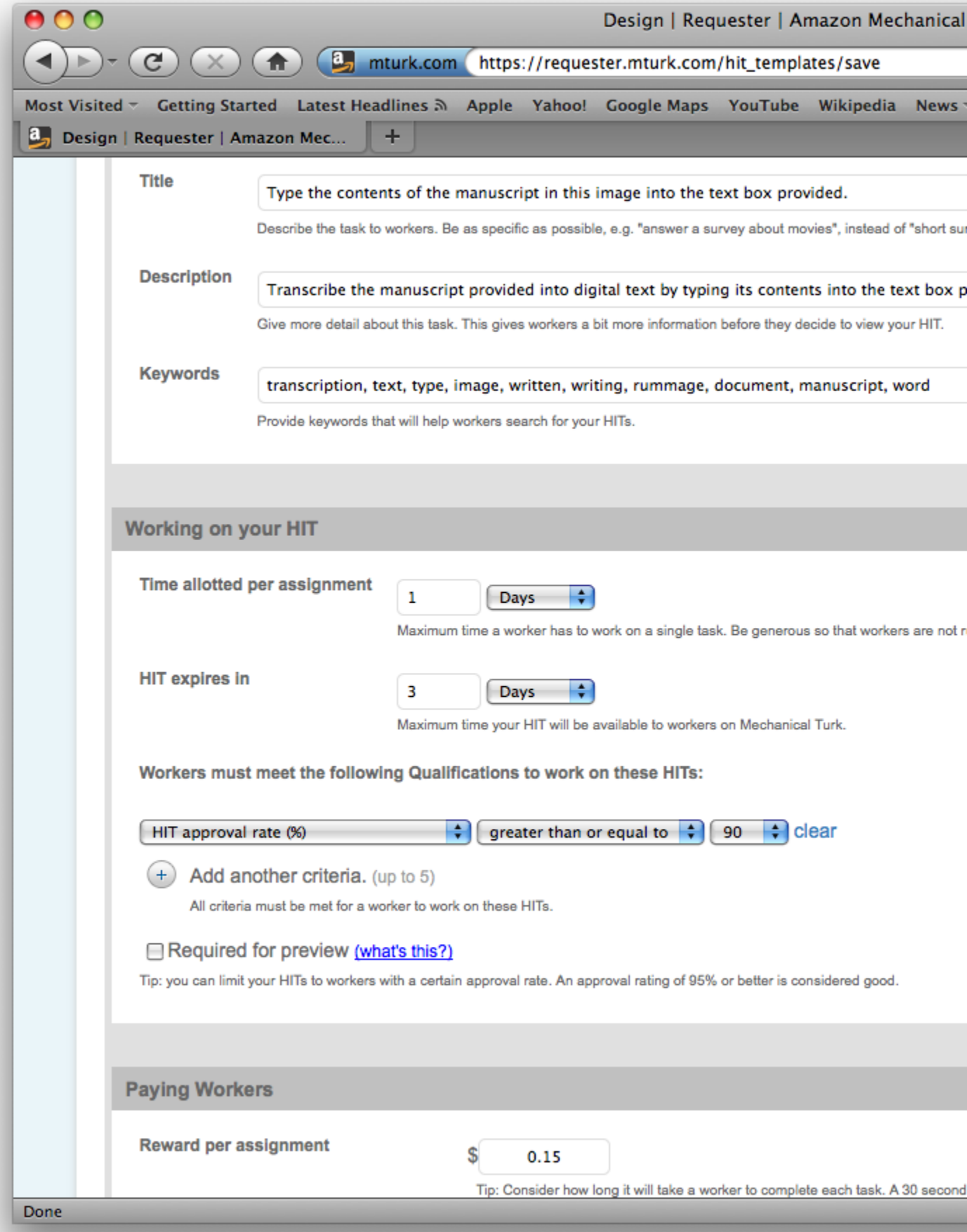
Using Mechanical Turk to Transcribe Handwritten Manuscripts

Several methodologies were tried, scrapped, and/or modified in the nascent stages of the Written Rummage project. The primary hurdles to overcome were developing a way to ensure transcription accuracy—a problem encountered by other transcription projects—and tweaking various aspects of management to expedite the project. What emerged is a technique that uses two separate transcriptions for each document, such that the first transcription is proofread in the process of the second transcription. The procedure is as follows.

Assuming that some storage of manuscripts to be transcribed exists in PDF format—in the present case, the Frederick Douglass Papers—the first step is to prepare the Mechanical Turk Human Intelligence Task (HIT), a simple task performed by an anonymous workforce for micropayments, in this case, transcribing a page of handwritten text.

Amazon's Mechanical Turk, often referred to as MTurk, is among the premier crowdsourcing resources on the internet. MTurk's framework can be used to submit various tasks to a crowd for a prescribed amount of compensation, which Amazon also takes a percentage of. One must have an account with MTurk as a "Requester" to do this. [18] Once we do, we can click "Design" on the toolbar. From here, a HIT can be designed from a number of templates, depending upon what sort of task is to be accomplished by the crowd, see Figure 1. For transcription purposes, text specifying the requirements of the task, a link to the manuscript image, and a text/comment box for the worker to type the transcription are all the necessary components.

The "Design" stage is also when the Requester designates how much a worker is compensated for each completed HIT. The lower the compensation, the slower the transcription process takes, since workers have a market of tasks to select from. Written Rummage has found that $0.08 per HIT is enough to ensure quick acceptance and completion of HITs at this stage—usually a set of seven (7) HITs can be transcribed in three or four days.



Title Type the contents of the manuscript in this image into the text box provided.

Describe the task to workers. Be as specific as possible, e.g. "answer a survey about movies", instead of "short su

Description Transcribe the manuscript provided into digital text by typing its contents into the text box p

Give more detail about this task. This gives workers a bit more information before they decide to view your HIT.

Keywords transcription, text, type, image, written, writing, rummage, document, manuscript, word

Provide keywords that will help workers search for your HITs.

Working on your HIT

Time allotted per assignment 1 Days

Maximum time a worker has to work on a single task. Be generous so that workers are not r

HIT expires in 3 Days

Maximum time your HIT will be available to workers on Mechanical Turk.

Workers must meet the following Qualifications to work on these HITs:

HIT approval rate (%) greater than or equal to 90 clear

+ Add another criteria. (up to 5)

All criteria must be met for a worker to work on these HITs.

☐ Required for preview (what's this?)

Tip: you can limit your HITs to workers with a certain approval rate. An approval rating of 95% or better is considered good.

Paying Workers

Reward per assignment $ 0.15

Tip: Consider how long it will take a worker to complete each task. A 30 second

Figure 1. Template for HIT design in Amazon Mechanical Turk.

In order to implement the HIT, a Comma-Separated Values (CSV) file has to be loaded containing the links to each of the respective PDF files the Requester wants to have transcribed. One of the convenient features of MTurk is the ability to publish multiple HITs at once using a .CSV file. A list of any order X will in turn instruct the server to produce X HITs corresponding to each linked item, see Figure 2. Typically a batch of seven HITs is most manageable.

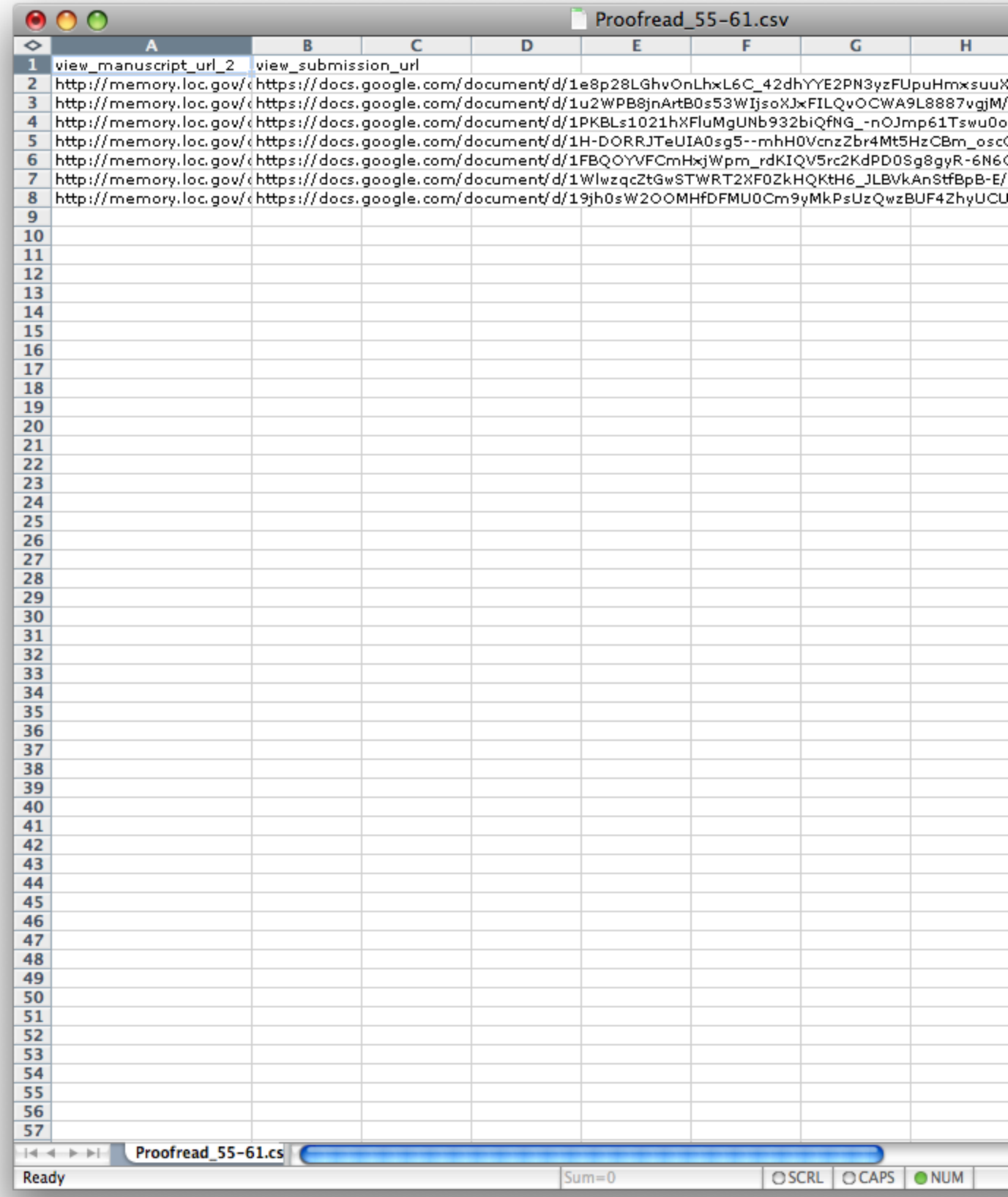

Proofread_55-61.csv
A
B
C
D
E
F
G
H
view_manuscript_url_2
view_submission_url
http://memory.loc.gov/
https://docs.google.com/document/d/1e8p28LGhvOnLhxL6C_42dhYYE2PN3yzFUpuHmxsuuX
https://docs.google.com/document/d/1u2WPB8jnArtB0s53WIjsoXJxFILQvOCWA9L8887vgjM/
https://docs.google.com/document/d/1PKBLs1021hXFluMgUNb932biQfNG_-nOJmp61Tswu0o
https://docs.google.com/document/d/1H-DORRJTeUIA0sg5--mhH0VcnzZbr4Mt5HzCBm_oscC
https://docs.google.com/document/d/1FBQOYVFCmHxjWpm_rdKIQV5rc2KdPD0Sg8gyR-6N6C
https://docs.google.com/document/d/1WlwzqcZtGwSTWRT2XF0ZkHQKtH6_JLBVkAnStfBpB-E/
https://docs.google.com/document/d/19jh0sW2OOMHfDFMU0Cm9yMkPsUzQwzBUF4ZhyUCU
Proofread_55-61.cs
Ready
Sum=0
SCRL
CAPS
NUM

Figure 2. One of our CSV files used to submit multiple transcription HITs.

Once the desired CSV file is made, we click “Publish” on the toolbar, select the desired template to implement—“first transcription,” or the like—upload the CSV file, confirm, name and “publish” the batch. MTurk will subtract the appropriate amount of pending funds from the Requester’s account, and all that is left to do is wait. We thereafter visit the “Manage” screen to check on the HITs progress, see Figure 3.

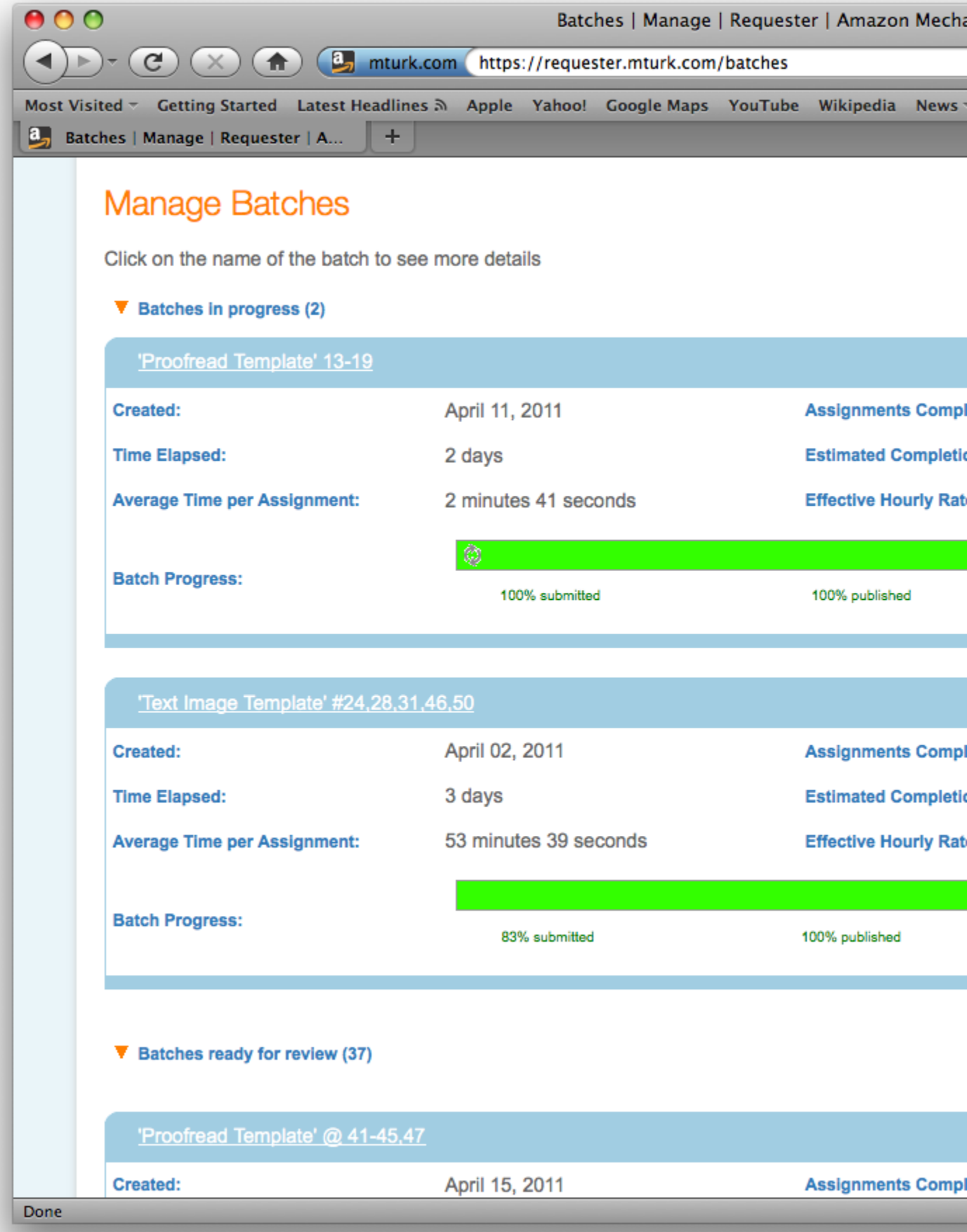

Batches | Manage | Requester | Amazon Mecha
mturk.com
https://requester.mturk.com/batches
Most Visited
Getting Started
Latest Headlines
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News
Batches | Manage | Requester | A...
Manage Batches
Click on the name of the batch to see more details
Batches in progress (2)
'Proofread Template' 13-19
Created:
April 11, 2011
Assignments Compl
Time Elapsed:
2 days
Estimated Completio
Average Time per Assignment:
2 minutes 41 seconds
Effective Hourly Rat
Batch Progress:
100% submitted
100% published
'Text Image Template' #24,28,31,46,50
Created:
April 02, 2011
Assignments Compl
Time Elapsed:
3 days
Estimated Completio
Average Time per Assignment:
53 minutes 39 seconds
Effective Hourly Rat
Batch Progress:
83% submitted
100% published
Batches ready for review (37)
'Proofread Template' @ 41-45,47
Created:
April 15, 2011
Assignments Compl
Done

Figure 3. Managing batches in Amazon Mechanical Turk.

As aforementioned, the batch will typically be transcribed in three or four days. To access the submissions, Requesters need only click the "Manage" link on the toolbar, then select the batch we want to view. The next step is to copy-and-paste the submissions to Google Docs, a free cloud-based document service, where they are saved and stored. Each transcription is saved as an independent document, resulting in a unique URL for each document. Since Google Docs has privacy settings, these must be set to "public" so that anyone with the URL can view them—they will be used again in the second transcription process.

The second stage of transcription has proven the most troublesome. The purpose of having a second transcription at all is quality control. Rather than proofreading the original submissions ourselves, which would be tantamount to undertaking the transcriptions ourselves, we decided to crowdsource this responsibility as well. Providing the original manuscript image alongside the first transcription, however, did not guarantee that any proofreading would be done—users could simply copy and paste the first transcription's content and submit it as their own. Yet again, having to manually read through each second submission and compare it to the first submission is little improvement upon transcribing the documents ourselves. We therefore considered another means of ensuring that the documents were at least being thoroughly scanned a second time for errors.

Once a first transcription is submitted and saved in Google Docs, a second, corresponding document is made from that transcription with errors inserted throughout the document. Some errors are unusual sequences of letters, such as "vxz," interpolated into words—for instance: "interpvxzolated." To avoid mere "spellcheck" proofreading submissions, we also insert entire words, such as "bark," into the pages—for instance: "we also insert bark entire words." The insertions of each page are logged and the edited manuscripts are saved with their own URLs. Once second transcriptions are submitted, we simply search the submissions for the key words or sequences of letters. If they have been edited out by the worker, then the submission is assumed to have been edited and proofread thoroughly, and therefore accepted. If they

have not been edited, the page is assumed to have not been proofread, the worker is not compensated, and the HIT is published again.

The process for publishing the proofreading HIT is similar to that of the first transcription. The primary difference is that two documents—the modified first transcription and the original manuscript image—are uploaded as links in the HIT rather than one. Using two columns rather than one in a CSV file allows us to enter the corresponding manuscript and image for each HIT. Workers are asked to open both the image and the transcription side-by-side, then to check the work for congruity.

Compensation for the proofread transcription is a dime—slightly more than for the first transcription due to the request to scan for congruity between two documents rather than simply type what one reads in one document. As with the first transcription, the chosen compensation usually yields a fully transcribed batch of seven manuscripts in three to four days. If multiple batches of seven manuscripts are published simultaneously, this projected completion time usually holds constant.

Results

The Written Rummage project has shown promise for providing a viable alternative for private manuscript collectors to transcribe their documents to digital, searchable text. Our test collection of documents, the aforementioned diary from the Frederick Douglass Papers, is completely transcribed and proofread. To make the 72-page transcription viewable, we created a free wiki for anyone interested to visit: frederickdouglassdiary.wikispaces.com. Each manuscript page has its own web page on the wiki, and that page contains the transcribed text as well as a link to the original source manuscript corresponding to that text. As an example, using the search box on the wiki, we find five pages (10, 29, 34, 42, 58) in the diary where the word “slave” or “slaves” appears. (Figure 4)

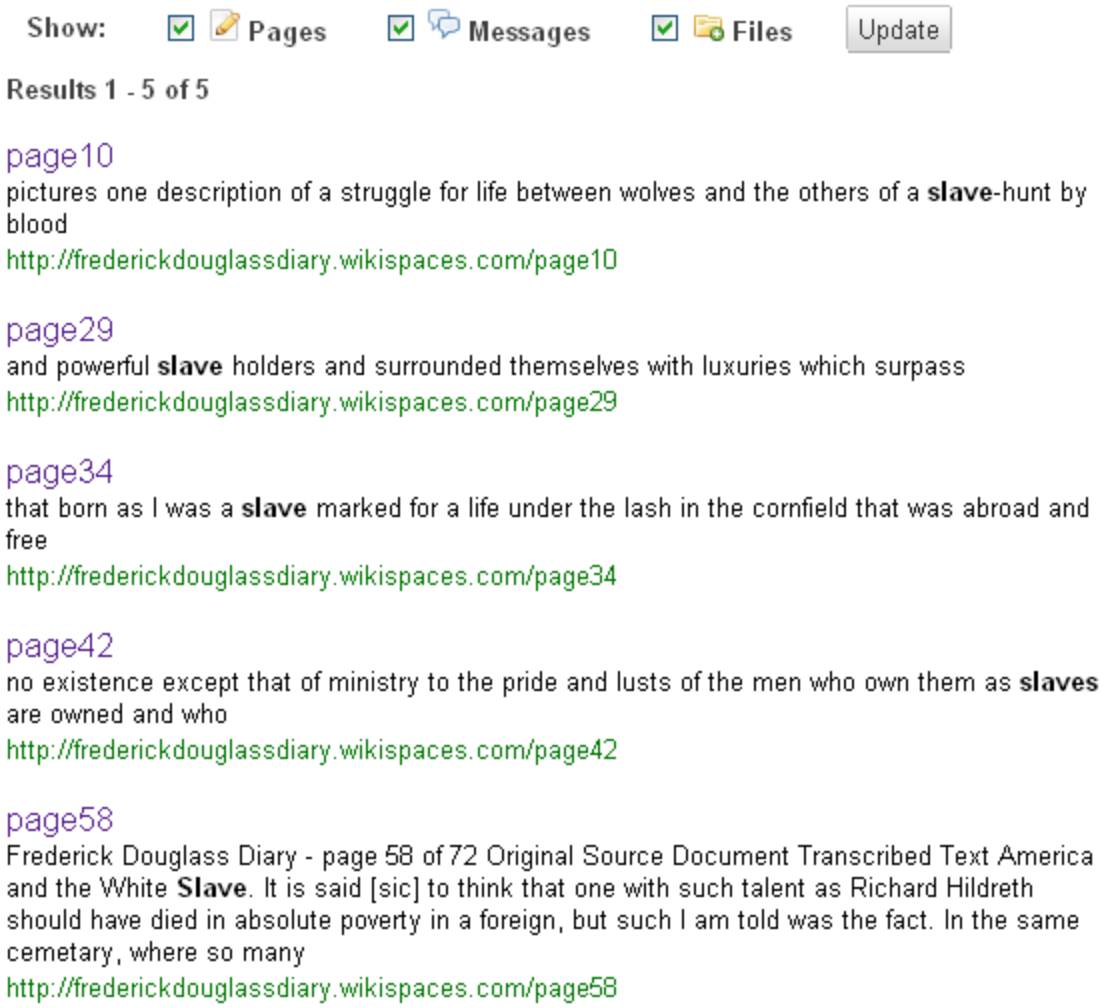


Figure 4. Searching for the word “slave” in Frederick Douglass’ diary.

By clicking on a page, for example page 42, we can see the entire transcribed text of that page. (Figure 5)

Frederick Douglass Diary - page 42 of 72
Original Source Document

Transcribed Text
infidel shoes should not touch their sacred Courts. We saw several washing their feet and afterward kneeling and kissing or touching the floor with their foreheads. In one respect these Mosques are to be commended. They have no images ~~of~~ or pictures of Saints or God. Make no effort to personify Deity. Visited the tombs of the Mamelukes and on our way saw various forms of squalor disease and deformity all manner of importunate begging. It was truly pitiful to see a people thus groveling in filth and utter wretchedness. We also visited the Bazaars where all manner of fabrics are manufactured and sold. Here men were smoking their long pipes drawing the smoke through water and selling or rather offering their wares for sale. The most painful feature met with in the streets are the hooded and veiled women. It is sad to think of that one half of the human family should be thus cramped, kept in ignorance and degraded, having no existence except that of ministry to the pride and lusts of the men who own them as slaves are owned ~~and who like~~ and worst is they seem to like to have it so.

Figure 5. Fully transcribed and proofread text from page 42 of Frederick Douglass'diary.

Then clicking the link provided retrieves an image of the original handwriting, as shown in Figure 6.

infedel shoes sh[d] not touch their so
their feet and afterward kneeling
with their foreheads. On one
to be commended. They have no
or God. make no effort to pe

Visited the tombs of the Mamelu
various forms of squalor. deser
manner of importunate beg
to see a people thus grovelling
We also visited the Bazars. w

Figure 6. Original image of page 42 of Frederick Douglass' diary. Copyright Library of Congress.

Time has been the most fickle variable in the nascent stages of Written Rummage. We have sought to develop an efficient model for receiving, transcribing, storing, proofreading, and re-storing documents. Most pressing has been the need to ensure that the HITs through Mechanical Turk were completed as hastily as possible. We began with low compensation per HIT. While documents were eventually transcribed for transcription rates of $0.01 per first transcription and $0.03 per proofread, they typically took two to four weeks to finish a batch of five to seven pages. After several infrastructural changes and changes in transcription rates—now about $0.08 per first transcription and $0.10 per proofread—most batches of six to eight pages are completed in less than or equal to a week. If multiple batches are published simultaneously, this number tends to hold constant rather than proportionally extend; that is, several batches of six to eight HITs can typically be transcribed and proofread in as much time as one batch.

With rates as they are presently set, our method offers transcription and proofreading at a rate of $0.18 per page, plus the 10% service charge to Mechanical Turk. Thus, our services can be implemented by private collectors for roughly $0.20 per page. Should one choose to add another stage of proofreading, this might increase to $0.30 per page. The exact cost will depend on exactly how long you are willing to wait for the transcription and proofreading process. Even then better price and speed performance can be attained. [19]

Our research has found that this is a remarkable improvement upon other professional transcription rates, which range from $2.00 per page to $8.00 per page depending upon the service and service provider. Further, unlike volunteer transcription efforts, which require IT support and have server maintenance issues, all our money went directly to the transcription effort. To provide some perspective on the difference these rates make, we can take, as an instance, the same 72-page diary used for this pilot run of Written Rummage. With rates held constant where they are now, the projected cost of transcribing the whole diary with Written Rummage is between $14.26 and $21.60. Due to complications and test runs and other restructuring expenses, we actually spent $22.86

to complete the diary. In contrast, using the range of rates offered by other services, the same transcription project would cost somewhere between $144.00 and $576.00. Our rates were so low with Mechanical Turk that at one point we had serious ethical discussions as to its use.

Certainly the economy of Mechanical Turk has given critics and users pause—and rightfully so: is it ethical to request tasks for such low compensation? [20] Is work simply being outsourced to people who benefit from the exchange rates, but at a pay rate otherwise unacceptable—for instance, unacceptable in the U.S.? We decided that using MTurk to transcribe handwritten documents was not unethical but we realize that these questions may be potential barriers to some researchers who are considering using our methods.

We in the end felt comfortable using MTurk because the character of crowdsourcing in general has typically been avocational for the crowd rather than vocational; that is, the workers are typically not performing HITs for living wage, but rather either as a hobby or for pocket cash, though this seems to be changing. [21] Many crowdsourcing projects offer no compensation at all but instead only call for volunteers. We use MTurk under the assumption that workers enter the Mechanical Turk marketplace aware of its supply and demand economy and choose their HITs accordingly—no HITs are compulsory. We therefore determined that, so long as workers choose to accept the tasks for the prescribed compensations, the mutual agreement establishes that the HIT's value is acceptable.

Conclusion

The "Information Age" is voracious; libraries and private collectors are looking for means of transcribing their handwritten manuscripts to make them available to the academic community and broader public. As the void they have yet to fill becomes more apparent, collectors are seeking out convenient, inexpensive ways of digitizing their documents. Digitization is the focus of a number of major projects going on in the academic and Internet community; likewise, crowdsourcing is increasingly being used to accomplish otherwise tedious or impossible tasks.

Our project utilized avant-garde technology to transcribe handwritten historical documents, and it did so affordably. The crowdsourcing model that we have presented uses no funds for server maintenance or IT support, nor for any other support personnel. Because of this, given a collection of handwritten documents from a library or other manuscript collector, we can have that collection transcribed and proofread at rates near $0.30 a page–approximately 15% the nearest competition's rates. A 200-page collection that would cost at least $400 to transcribe elsewhere would cost only around $60 to be transcribed with us. Further, due to significant milestones in overcoming past quality control challenges, we can expediently offer accurate transcriptions. Entire collections, depending upon their size, can be returned digital and searchable in 2-4 weeks, allowing collectors to contribute them to shared library resources or other academic spheres. The Written Rummage project set out to develop a means of expediently and accurately transcribing handwritten documents to digital, searchable text; it did so–and thereby broadens the possibilities for sharing knowledge in the digital realm.

References

[1] The James Madison Papers -American Memory from the Library of Congress [internet] [cited 2011 October 13] Available from: http://memory.loc.gov/ammem/collections/madison_papers/mjmabout.html

[2] The Thomas Jefferson Papers -American Memory from the Library of Congress [internet] [cited 2011 October 13] Available from: http://memory.loc.gov/ammem/collections/jefferson_papers/mtjabout.html

[3] Cohen P. 2010. Scholars Recruit Public for Project. New York Times. [internet] [cited 2011 October 13] Available from: http://www.nytimes.com/2010/12/28/books/28transcribe.html?nl=books&pagewanted=all

[4] Zooniverse – Real Science Online. [internet] [cited 2011 October 13] Available from: http://www.zooniverse.org/home

[5] Cardamone C. et. al. 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. Monthly Notices of the Royal Astronomical Society, 399: 1191–1205. doi: 10.1111/j.1365-2966.2009.15383.x

[6] Lintott C. et. al. 2009. Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo?. Monthly Notices of the Royal Astronomical Society, 399: 129–140. doi: 10.1111/j.1365-2966.2009.15299.x

[7] Old Weather – Our Weather's Past, the Climate's Future. [internet] [cited 2011 October 13] Available from: http://www.oldweather.org/

[8] The Open Dinosaur Project [internet] [cited 2011 October 13] Available from: http://opendino.wordpress.com/

[9] Old Weather sails on. [internet] [cited 2011 October 13] Available from: http://blogs.zooniverse.org/oldweather/2011/02/old-weather-sails-on/

[10] List of crowdsourcing projects – Wikipedia, the free encyclopedia. [internet] [cited 2011 October 13] Available from: http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects

[11] Dunning A. 2011. Innovative use of crowdsourcing technology presents novel prospects for research to interact with much larger audiences, and much more effectively than ever before. [internet] [cited 2011 October 13] Available from: http://blogs.lse.ac.uk/impactofsocialsciences/2011/08/25/innovative-use-of-crowdsourcing/

[12] Moyle M, Tonra J and Wallace V. 2010. Manuscript transcription by crowdsourcing: Transcribe Bentham. LIBER Quarterly, 20 (3-4)

[13] Transcribe Bentham. [internet] [cited 2011 October 13] Available from: http://www.ucl.ac.uk/transcribe-bentham/

[14] Winners of the AHRC's £4m Digital Programme Announced.[internet] [cited 2011 October 13] Available from: http://www.ahrc.ac.uk/News/Latest/Pages/winnersdigitalprogramme.aspx

[15] Brumfield B. 2011. 2010: The Year of Crowdsourcing Transcription. [internet] [cited 2011 October 13] Available from: http://manuscripttranscription.blogspot.com/2011/02/2010-year-of-crowdsourcing.html

[16] McHenry K. et al. 2011. Toward free and searchable historical census images. Electronic Image and Signal Proscessing. 22 September 2011. http://dx.doi.org/10.1117/2.1201109.003833

[17] Scripto | Crowdsourcing Documentary Transcription. [internet] [cited 2011 October 13] Available from: http://scripto.org/

[18] Amazon Mechanical Turk. [internet] [cited 2011 October 13] Available from: https://www.mturk.com/mturk/welcome

[19] Faridani S, Hartmann B and Ipeirotis P. 2011. What's the Right Price? Pricing Tasks for Finishing on Time. [internet] [cited 2011 October 13] Available from: http://husk.eecs.berkeley.edu/courses/cs298-52-sp11/images/c/c7/Faridani-hcomp11.pdf

[20] Fort K, Adda G and Cohen K. 2011. Amazon Mechanical Turk: Gold Mine or Coal Mine? Computational Linguistics. Vol. 37, No. 2, Pages 413-420 doi:10.1162/COLI_a_00057

[21] Ipeirotis P. The New Demographics of mechanical Turk. [internet] [cited 2011 October 13] Available from: http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html