9-12-2022

Optical Character Recognition (OCR) Approaches to Cursive Handwriting Transcription: Lessons from the Blythe Owen Letters Project

Jazmyne Lavalas
Andrews University, jazmyne@andrews.edu

Marianne Kordas
Andrews University, kordas@andrews.edu

Rodney Summerscales
Andrews University, summersc@andrews.edu

Journal of Adventist Archives, #2 (2022): 53—71

Optical Character Recognition (OCR) Approaches to Cursive Handwriting Transcription: Lessons from the Blythe Owen Letters Project

by
Jazmyne Lavalas, Marianne Kordas, and Rodney Summerscales

Introduction

Making the contents of an archival collection easily accessible to scholars and a wider public can be a daunting task. There has been a long-standing interest in both the computing and archival communities, however, in using computer-assisted means to facilitate description, access, and innovative analysis of archival collections. This scholarly movement grew out of the field of humanities computing and coalesced under the umbrella term of digital humanities in the early 2000s (Fitzpatrick 2011).

This paper is a case study describing our efforts to use Optical Character Recognition (OCR) technology to transcribe a select, offline sample of PDF scans of handwritten letters by Dr. Blythe Owen—a prominent Seventh-day Adventist musician, composer, and pedagogue—into text-based documents in order to make the letters more accessible to researchers.

Our study is interdisciplinary in nature, drawing on musicology, archival science, software engineering, and the possibilities presented by Artificial Intelligence (AI) programming for its theoretical frameworks. Our goal is to give practical examples of the state of OCR capabilities for archival transcription, and to suggest areas for application and improvement. We offer an overview of why Owen and her correspondence are important to American musicology as well as to Seventh-day Adventist history; a brief history of OCR

applications for offline handwriting recognition problems; and a comparative review of four OCR programs (*Google Cloud Vision, Pen to Print APP, SimpleOCR,* and *Transkribus*) applied to our selected Owen letter dataset. Our project augments, but does not duplicate, the comparative discussion published last year (2021) by Jain, Taneja, and Taneja of sixteen other OCR toolset softwares.

Blythe Owen: Letters of Importance

Born in 1898 in Long Prairie, Minnesota, American pianist and composer Blythe Owen was an accomplished teacher and musician for over seventy years (FamilySearch). Her love of music began early, and she started teaching piano when still a teenager. In 1917 she graduated from the Pacific College Conservatory in Newberg, Oregon; she was subsequently hired to teach piano and other music classes at Walla Walla College, a Seventh-day Adventist school located in College Place, Washington. After marriage and several relocations throughout the United States, Owen moved to Chicago—first in 1926, and then permanently in the late 1930s. (Schultz, Mack, and Kordas 2022). Here she not only studied, taught, performed, and composed, but also connected with other well-known classical musicians of the mid-twentieth century such as composer Florence Price, pianist Rudolf Ganz, and conductor Percy Grainger.

In 1941 Owen received her undergraduate degree in piano performance from Chicago Musical College; a year later, in 1942, she completed a master's degree in composition at Northwestern University. During this period, she taught piano and music theory at several different schools, including Northwestern University, Cosmopolitan House Conservatory, Roosevelt University School of Music, and Chicago Teachers College.

Owen was one of the first women to complete a PhD in composition from the prestigious Eastman School of Music, graduating in 1953. After three decades living in Chicago and surviving via part-time adjunct work, she accepted an offer in 1961 to return to Walla Walla College for full-time employment. She then joined the music faculty of at Andrews University in 1964. She continued to compose and teach privately into her late 90s. She died in Berrien Springs, Michigan in 2000 at the age of 101, leaving behind a legacy of over 150 compositions and numerous accomplished students (Ibid).

Between 1919 and 1963, Owen wrote over 2,000 letters to her mother describing her life as a student, teacher, performer, and composer (Kordas 2021). These items of correspondence illuminate not only Owen's own experiences as a professional musician and a Seventh-day Adventist, but also document the life and works of other artists with whom she interacted. The surviving letters are a valuable source of information for scholars of women in music, American music, Chicago in the twentieth century, and Seventh-day Adventist history in general. Owen's letters are housed along with her other papers—including photographs and manuscripts of her compositions—in the archives of the Center for Adventist Research at Andrews University (Collection 186).

Transcribing: The Research Problem

The process of transcribing Owen's original letters into text-based documents began in 2016 with two student research assistants, under the direction of Andrews University's music librarian. The letters are written in Owen's distinctive cursive script. While effective, transcribing a single letter in person into a typed document is time consuming and can take up to two hours per letter (Kordas 2021).

Therefore, we sought other, computer-based solutions for transforming handwritten sources into typed documents. The most commonly used technological approach to handwriting recognition is Optical Character Recognition (OCR) software (Plötz and Fink 2009; Peng et al. 2013; Patel et al. 2012). This type of software is fairly easy to acquire and usually inexpensive. In their 2003 monograph surveying the advancement of handwriting recognition, data scientists Liu, Cai, and Buse note that:

> There are two major problem domains in handwriting recognition: online and offline. In online handwriting recognition, data are collected while they are being generated on a digitizing surface [...] However, in offline handwriting recognition, all that is available to the recognition system is the digitized spatial information, e.g., the image of the address scanned from an envelope or an amount shown on a cheque. As a consequence, online handwriting recognition has a much higher recognition rate as compared to that for the offline case (1).

Thus, with an offline cursive handwriting recognition task such as found in the Owen corpus of letters, common OCR software may not always prove to be successful (Plamondon and Srihari 2000). The challenge is finding OCR software that leverages machine learning technology and was trained on similar datasets (Fitrianingsih et al. 2017; Hämäläine and Hengchen 2019; Drobac and Lindén 2020). Doing so was the main focus of the project documented in this paper: finding cost-effective, widely available OCR software that sufficiently reads and transcribes sources inscribed with handwritten cursive script.

Literature Review: A History of OCR Applications for Handwriting Recognition Problems

The possibilities and challenges of OCR applications for handwriting recognition have both tantalized and frustrated researchers in the various branches of computer and information sciences for over seventy years (Lorette 1999, 4–5; Cheriet et al. 2007, 2–3; Memon et al. 2020).

It is commonly accepted that the field of digital humanities may be seen as starting in 1949, when Father Roberto Busa approached IBM for computing help with his index to the complete works of medieval theologian Thomas Aquinas. Punch cards led to magnetic reels, then to micro-computers, and finally to an internet-based server and website, which can still be searched today at https://www.corpusthomisticum.org/it/index.age (Busa 2004; Priego 2011; Breure 1993, 1–3). Some twenty years later, in 1968, another pioneering computing scholar, Richard Morgan, envisioned being able to batch-load books and other documents into the "maw" of a computer machine enabled with OCR capabilities.

Great strides have been accomplished since then in performing optical character recognition of both text-based and handwritten documents, but many challenges still remain. Indeed, according to computer scientist Ching Y. Suen, "handwriting recognition is known to be one of the most challenging subjects in the field of pattern recognition" (1994, 71). Entire conferences, across many languages and countries, are devoted to finding solutions towards having computers read the intricacies of human handwriting. The most prominent conference in this area of research is the biennial International Conference on Frontiers in Handwriting Recognition, which started in 1990 in Montreal,

Quebec, as the International Workshop on Frontiers in Handwriting Recognition (ICFHR 2022).

The literature on computerized handwriting recognition describes a broad range of applications for OCR techniques in handwriting recognition across a spectrum of areas such as banking, medical records, and historical research (Allen 1987, Agrawal et al. 2020, Komkov 2022). Our project, while drawing from this broader technical knowledge, focused on testing the practical application of OCR software products for handwriting recognition presently available to the average consumer or archivist who may not have a rigorous background in computer and data science. Advanced technical skills, however, are advantageous when undertaking a digital humanities project of the type we discuss here, as even the simplest software we tested did require some data manipulation.

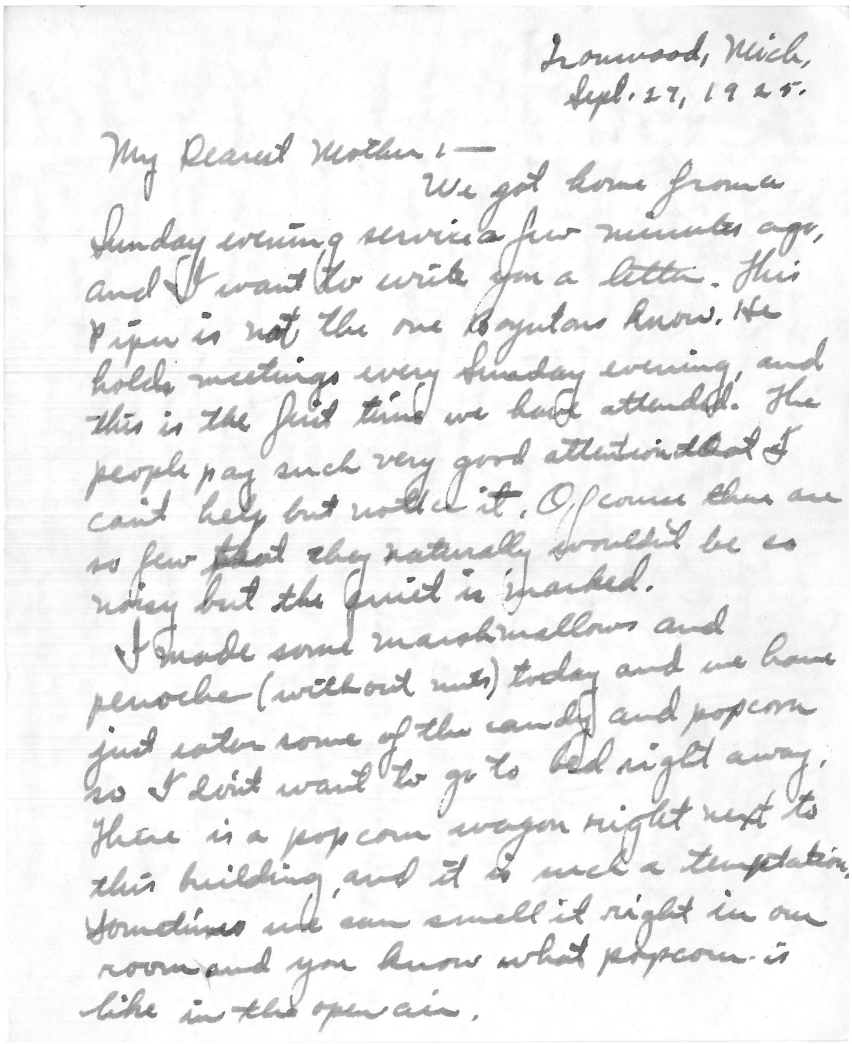
Project Methodology

The first step in our case study of applying OCR techniques to an offline handwriting recognition problem was to conduct both general searches online as well as more targeted searches in journal databases for recommendations and references to currently available OCR software products. Approximately ten to fifteen OCR software products were pre-screened with a scan of one Owen letter. We examined a wide variety of OCR software options, trying to find good results from both free and paid OCRs. When pre-screening which OCRs to test, we looked at the ease of use, accuracy, compatibility with our dataset, and accessibility. The four products that did well in pre-screening— *Google Cloud Vision, Pen to Print APP, SimpleOCR, and Transkribus*—were subsequently fully tested with a larger subset of Owen's complete letters.

Our testing subset consisted of ten randomly selected letters written by Owen across several decades. [1] These letters range from between three to seven pages in length. We used letters that had been already scanned as PDFs, as well as human-transcribed into text documents, in order to generate a clear validation set.

[1] Dates of the ten letters: 1919-10-01, 1919-10-29, 1925-09-27, 1928-05-09, 1933-08-06, 1935-03-17, 1947-09-23, 1944-03-08, 1951-02-11, 1952-01-12. Two letters per decade were used for the section of letters that had been hand-transcribed and could form the validation set.

Throughout this paper we will compare figures that highlight the results between the various OCR applications. Below in **Figure 1** is an image of one of Owen's original letters, dating from September 27, 1925, paired with its human-transcribed text in **Figure 2:**



Ironwood, Mich,
Sept. 27, 1925.

My Dearest Mother:—
We got home from a Sunday evening service a few minutes ago, and I want to write you a letter. This piper is not the one Boyntons know. He holds meetings every Sunday evening and this is the first time we have attended. The people pay such very good attention that I can't help but notice it. Of course there are so few that they naturally wouldn't be as noisy but the quiet is thanked.

I made some marshmallows and penoche (without nuts) today and we have just eaten some of the candy and popcorn so I don't want to go to bed right away. There is a popcorn wagon right next to this building, and it is such a temptation sometimes we can smell it right in our room and you know what popcorn is like in the open air.

FIGURE 1: A scan of Owen's original letter

Ironwood, Mich.
Sept. 27, 1925.

My Dearest Mother: —
We got home from a Sunday evening service a few minutes ago, and I want to write you a letter. This [Piper?] is not the one Boyntons know. He holds meetings every Sunday evening, and this is the first time we have attended. The people pay such very good attention that I can't help but notice it. Of course there are so few that they naturally wouldn't be so noisy but the quiet is marked.
I made some marshmallows and penuche (without nuts) today and we have just eaten some of the candy and popcorn so I don't want to go to bed right away. There is a popcorn wagon right next to this building, and it is such a temptation. Sometimes we can smell it right in our room and you know what popcorn is like in the open air.

FIGURE 2: OCR result of Owen's letter (The "penuche" mentioned in this letter is a fudge-like candy).

Depending on the OCR tested, the requirements for the input file varied. For example, for several OCR options tested, images were the only accepted document type. To solve any specific data manipulation or preparing, some Python coding and packages were used. Once the data was cleaned, it was fed into the four OCR software programs we had selected for full testing.

The produced OCR transcription results were then evaluated against the human-transcribed versions of the ten letters in our dataset to determine the accuracy of the OCR. After this, the OCR was retested on the dataset to see if results could be improved using different capabilities and settings. Once we were satisfied that we had received the best results using a particular OCR, we compared the results with those produced by the other three OCR software product examined in this study. Our accuracy performance metric judged the rate of accuracy based on correct words rather than just alphabetic characters. In all, we fully tested four products, two of which were free, and two of which require payment for long-term, complete functionality.

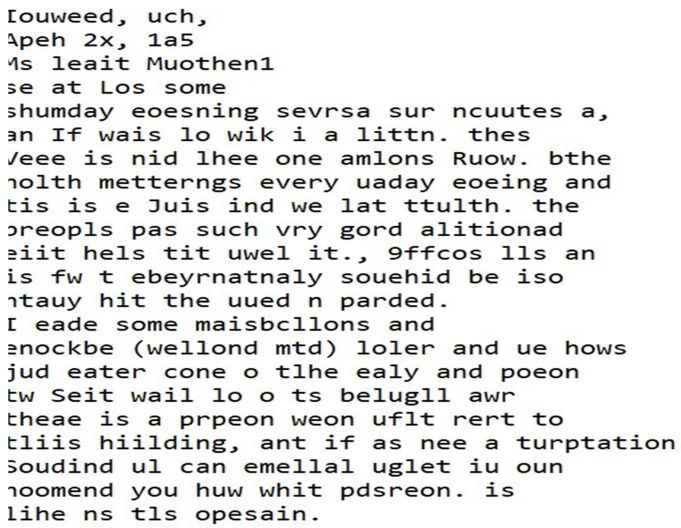
Results: Free OCRs

SimpleOCR

The first OCR we tested was *SimpleOCR*. This is a free software used for transcribing handwritten as well as machine text. It is offered in both online and downloadable versions.[2] The downloadable version allows for more customizations, including text recognition of several languages, and this was the version used for this project. *SimpleOCR* is primarily designed for segmented scripts, so it was not expected to work very well with our cursive dataset. However, this OCR does have some features that could be used for cursive datasets. The software is user friendly and is built from another common open-source OCR called *Tesseract*.

SimpleOCR offers direct connection to a scanner; it also allows file uploads. However, the uploaded files can only be in the form of images. This caused an extra step of work for our dataset. It is recommended for handwriting recognition that documents consisting of 300–500 words be used for training the OCR. While this OCR does mention having features to help transcribe cursive writing, these features did very little for the performance of this OCR with Owen's letters. Even though this was one of the highest rated free OCRs used in this study, the results produced were still very poor. The accuracy of this OCR was only 13% across all ten documents. While individual characters could be recognized accurately, whole words—and thus the whole document—did not transcribe well, as can be seen in **Figure 3a and Figure 3b**:

[2] https://www.simpleocr.com/



```
Iouweed, uch,
Apeh 2x, 1a5
Ms leait Muothen1
se at Los some
shumday eoesning sevrsa sur ncuutes a,
an If wais lo wik i a littn. thes
Veee is nid lhee one amlons Ruow. bthe
nolth metterngs every uaday eoeing and
tis is e Juis ind we lat ttulth. the
preopls pas such vry gord alitionad
eiit hels tit uwel it., 9ffcos lls an
is fw t ebeyrnatnaly souehid be iso
ntauy hit the uued n parded.
I eade some maisbcllons and
enockbe (wellond mtd) loler and ue hows
jud eater cone o tlhe ealy and poeon
tw Seit wail lo o ts belugll awr
theae is a prpeon weon uflt rert to
tliis hiilding, ant if as nee a turptation
Soudind ul can emellal uglet iu oun
noomend you huw whit pdsreon. is
lihe ns tls opesain.
```

FIGURE 3a: Results of *SimpleOCR* transcription

```
word, i
tepl. 27, 1928.
My Dent Water
We got home from
Sunday woning verrich fer mal age,
and want to with lette.
Vign is not the one Boyulme know. He
holde meetings I did the
anche very good attentioned of
but with it. Oferme ha
mit bes
• for that they naturally
way hut the fundin hed
Sinds and march mellom and
today and i hove
che (with and popcon CATE
jester some of the Alright
I don't want to go?
there is a popcom wagen night next
this heilding, and it a week a templatin
Souch mell it right in o
od you know, popcom
like in pain.
```

FIGURE 3b: Results of *SimpleOCR* transcription retry

Pen to Print App

The second OCR tested was the *Pen to Print App*. This OCR is an application available for both iOS and Android phones, and was designed as the first phone app to be able to convert images of handwritten notes into texts that can be saved and shared.[3] This product is used mainly for offline handwriting recognition. While it is not available for laptops or desktops from the original developer, combined with other software this can be accommodated.

Its main function is to convert images rather than larger documents in PDF. Separating documents into smaller units and saving them as images (JPEG, PNG, etc.) before processing is a possible solution. This app is free with the option of upgrading to a paid subscription and enables users to process a large amount of data even if that data is in small segments. Overall, the phone app is very user-friendly, allowing for quick transcriptions to take place with very few steps required. This mobile app performed better than expected with our complex dataset. With a sample of the Owen letters broken up into single pages, this OCR produced acceptable results, as illustrated in **Figure 4**.

[3] https://www.pen-to-print.com/

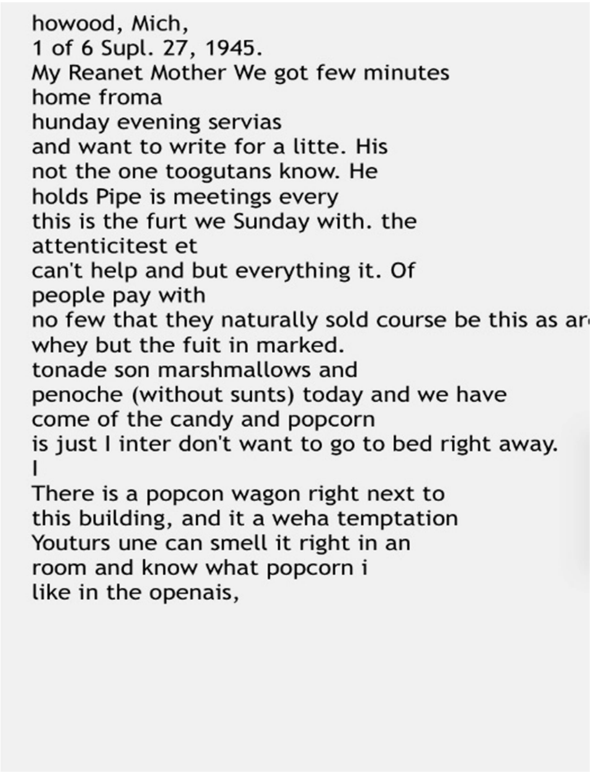

howood, Mich,
1 of 6 Supl. 27, 1945.
My Reanet Mother We got few minutes
home froma
hunday evening servias
and want to write for a litte. His
not the one toogutans know. He
holds Pipe is meetings every
this is the furt we Sunday with. the
attenticitest et
can't help and but everything it. Of
people pay with
no few that they naturally sold course be this as ar
whey but the fuit in marked.
tonade son marshmallows and
penoche (without sunts) today and we have
come of the candy and popcorn
is just I inter don't want to go to bed right away.
I
There is a popcon wagon right next to
this building, and it a weha temptation
Youturs une can smell it right in an
room and know what popcorn i
like in the openais,

FIGURE 4: *Pen to Print* app results

For most of the data, the app's OCR was able to read the documents with few errors. However, these results were not consistent across all the data. In fact, when this OCR did not perform well, the results were quite poor, leading to documents with many errors. Another issue with this application is that users are unable to edit the results of the text file in this application. While using this platform the correction of errors would need to happen post-transcription. Overall, the accuracy for this OCR was 31.7%. While this is still relatively low, it performed twice as well as *SimpleOCR*.

Results: Paid OCRs

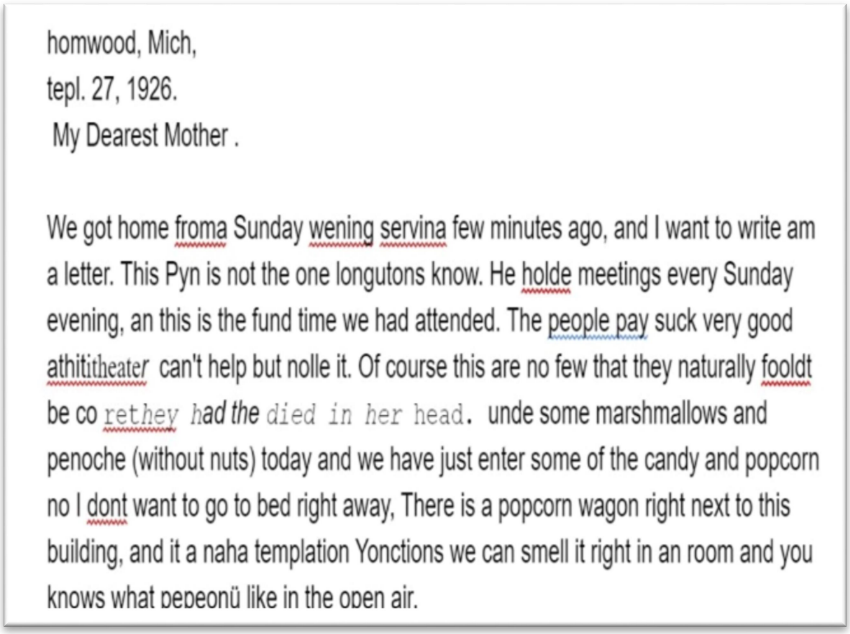
Google Cloud Vision

The third OCR tested is not actually an OCR but an Application Programming Interface (API) with OCR capabilities. This API was

created by Google and is called *Google Cloud Vision AI*.[4] The use of this API is not free long-term, but it does offer a free trial with a set number of credits that can be used monthly. This API offers a wider range of capabilities than all other OCR software evaluated in this paper. It allows for searchable output files and displays more properties of the documents being transcribed. One can search for common themes in literature and the program gives a likeliness score to show what a given document is about. It also gives specific characteristics of the document based on set parameters. For example, with our dataset, the program identified that the documents were letters. It assessed the original quality of the scanned document and recognized that ink was used to write the Owen letters.

While this API offers a wide range of capabilities, it is not recommended for the average user. For best results, *Google Cloud Vision* requires experience with setting up technical environments and proficiency with common programming languages, such as Python. It required the most preparation and cleaning of the dataset before processing, thus causing it to be the most time-consuming software discussed in this paper. The results of this API were somewhat successful, but the resulting documents would need to be checked thoroughly, as shown in **Figure 5.** There were specific nuances of our dataset that did not work well with this API. The accuracy of this OCR was 56.2% across our dataset.

[4] https://cloud.google.com/vision


homwood, Mich,
tepl. 27, 1926.
My Dearest Mother .

We got home froma Sunday wening servina few minutes ago, and I want to write am a letter. This Pyn is not the one longutons know. He holde meetings every Sunday evening, an this is the fund time we had attended. The people pay suck very good athititheater can't help but nolle it. Of course this are no few that they naturally fooldt be co rethey had the died in her head. unde some marshmallows and penoche (without nuts) today and we have just enter some of the candy and popcorn no I dont want to go to bed right away, There is a popcorn wagon right next to this building, and it a naha templation Yonctions we can smell it right in an room and you knows what pepeonü like in the open air.

FIGURE 5: *Google Cloud Vision API* results

Transkribus

Finally, the last and most successful OCR we tested was *Transkribus*.[5] This software specializes in offline handwriting recognition, with its main focus on using AI for text recognition of historical documents. *Transkribus* is relatively easy to use. This software can be used in a browser, or it can be downloaded onto a PC. *Transkribus* is highly customizable, offering recognition not only of several languages but also different eras. Its platform is built around different transcription models trained on several distinct handwriting datasets; this allows a user to select a model that is similar to their own dataset in order to achive the best results.

Transkribus is a paid software and uses credits for each document transformation. Upon registration, users receive 500 free credits; depending on the model needed for a dataset, a credit will typically process a page of data. This software is one of the

[5] https://readcoop.eu/transkribus/

most straightforward OCR tested in this project. Images or PDFs to be transcribed are uploaded into the program. After the original document is processed, the new, transcribed document can be downloaded in several formats including PDF, docx, image, etc. It also offers a text editor for each page, so that transcriptions errors can be corrected before they are downloaded.

With the Owen dataset, we used the *English Handwriting 18th -19th century (2)* model. This model was trained on the manuscripts of a British philosopher. It trained on a sufficiently large dataset which looked similar to the Owen letters. This specific model performed relatively well with our dataset, as seen in **Figure 6**:

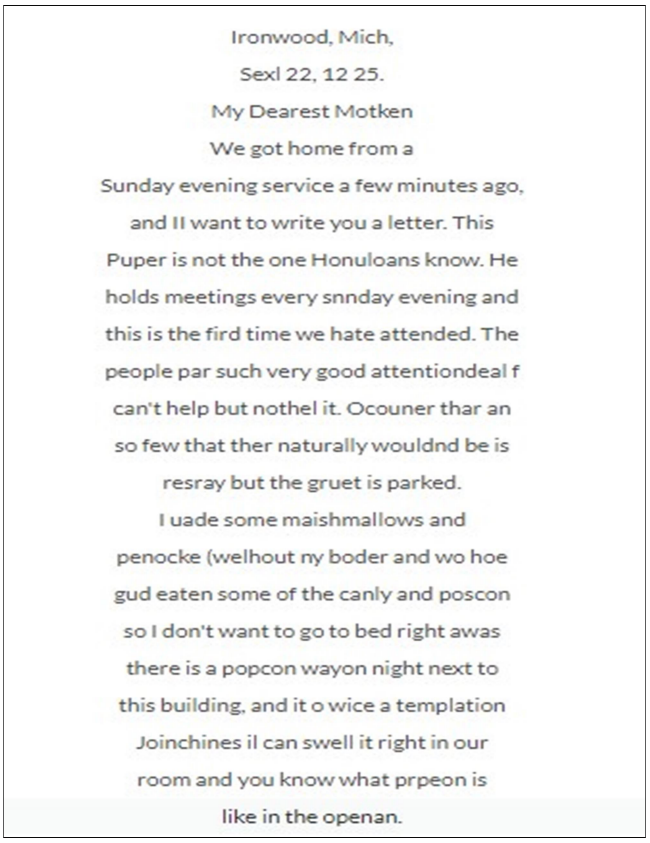

Ironwood, Mich,
Sexl 22, 12 25.
My Dearest Motken
We got home from a
Sunday evening service a few minutes ago,
and II want to write you a letter. This
Puper is not the one Honuloans know. He
holds meetings every snnday evening and
this is the fird time we hate attended. The
people par such very good attentiondeal f
can't help but nothel it. Ocouner thar an
so few that ther naturally wouldnd be is
resray but the gruet is parked.
I uade some maishmallows and
penocke (welhout ny boder and wo hoe
gud eaten some of the canly and poscon
so I don't want to go to bed right awas
there is a popcon wayon night next to
this building, and it o wice a templation
Joinchines il can swell it right in our
room and you know what prpeon is
like in the openan.

FIGURE 6: *Transkribus* results

This OCR had an accuracy rate of 61.8%, performing better than all other OCRs tested in this project. While the results of the transcribed document were not perfect and needed to be checked, few errors were made overall.

Conclusion and Recommendations

Offline cursive handwriting recognition is a complex problem for which solutions are still developing. Solutions to this problem are especially useful for historical documents where transcription would allow more people to have access to valuable information. Given the OCRs tested in this project, *Transkribus* would be the tool most useful for our dataset and similar datasets. This OCR is easy to use and offers several models from which users can select the one best suited for their documents. We should note that while *Transkribus* produced the best results, there were still errors in the documents, so transcriptions would benefit from double checking against original sources and other quality control protocols.

The next question asked could be how one could go about improving the results of different OCRs. One possible approach could be the use of multimodal deep learning (Zhongwei et al. 2019) or feature-based word classification (Kissos and Dershowitz 2016). Some studies have been conducted to expolore the use of multimodal models for multilingual OCRs, but not specifically if these approaches could be leveraged for taking the results of a successful OCR and improving it (Peng et al. 2013). Offline cursive handwriting recognition is an area of study that should continue to be investigated, and one which holds significant promise for improvement and growth. We acknowledge that there may be other software options that exist, but believe that the four OCR softwares discussed in this article are among the best current products in the field. Meanwhile, human-typed transcriptions of historical and archival documents remain the most accurate, if time-consuming, transcription method currently available.

The authors wish to thank Andrews University and the Office of Research for the Undergraduate Research Scholarship that funded this project. We also appreciate the responses from our reviewers.

List of sources

Allen, Marie. 1987. "Optical Character Recognition: Technology with New Relevance for Archival Automation Projects." *The American Archivist* 50(1): 88–99.

Agrawal, Prateek et al. 2021. "Automated Bank Cheque Verification using Image Processing and Deep Learning Methods." *Multimedia Tools and Applications* 80: 5319–5350. https://doi.org/10.1007/s11042-020-09818-1

Breure, Leen. 1993. "Introduction." In *Optical Character Recognition in the Historical Dicipline: Proceedings of an International Workshop*. Göttingen, Germany: Max Plank-Institut.

Busa, Roberto A. 2004. "Foreword: Perspectives on the Digital Humanities." In *A Companion to Digital Humanities*. Edited by Susan Schreibman; Ray Siemens, and John Unsworth. Malden, MA: Blackwell Publishing. xvi–xxii.

Cheriet, Mohamed; Nwwaf Kharama; Cheng-Lin Liu; and Ching Y. Suen. 2007. *Character Recognition Systems: A Guide for Students and Practitioners*. Hoboken, NJ: Wiley-Interscience.

Collection 186. *Blythe Owen [Papers]*. Center for Adventist Research, James White Library, Andrews University, Berrien Springs, MI. http://www.centerforadventistresearch.org/?mdocs-file=5587

Drobac, Senka and Krister Lindén. 2020. "Optical Character Recognition with Neural Networks and Post-Correction with Finite State Methods." *IJDAR* 23: 279–295. https://doi.org/10.1007/s10032-020-00359-9

FamilySearch. "United States Census, 1920." V. Blythe Owen in entry for Minnie B. Owen. Database with images. https://www.familysearch.org/ark:/61903/1:1:M4ZG-D9Q

Fitrianingsih; Sarifuddin Madenda; Ernastuti; Suryarini Widodo; and Rodiah. 2017. "Cursive Handwriting Segmentation Using Ideal Distance Approach." *International Journal of Electrical & Computer Engineering* (2088-8708) 7 (5): 2863–72. doi:10.11591/ijece.v7i5.pp2863-2872

Fitzpatrick, Kathleen. 2011."The Humanities, Done Digitally." *Chronicle of Higher Education*

https://www.chronicle.com/article/the-humanities-done-digitally/

Hämäläinen, Mika and Simon Hengchen. 2019. "From the Paft to the Fiiture: a Fully Automatic NMT and Word Embeddings Method for OCR Post-Correction." White Paper. https://arxiv.org/pdf/1910.05535v1.pdf

ICFHR 2022. "International Conference on Frontier in Handwriting Recognition." Annual conference website with organization's history. http://icfhr2022.org/

Jain, Pooga; Gavita Taneja; and Harmunish Taneja. 2021. "Which OCR Toolset is Good and Why? A Comparative Study." *Kuwait Journal of Science* 48(2): 1–12. https://journalskuwait.org/kjs/index.php/KJS/article/view/9589/443 and https://doi.org/10.48129/kjs.v48i2

Lorette, G. 1999. "Handwriting Recognition or Reading? Situation at the Dawn of the 3rd Millennium." In *Advances in Handwriting Recognition*. Edited by Seong-Whan Lee. Singapore: World Scientific. 19–28.

Kissos, Ido and Nachum Dershowitz. 2016. "OCR Error Correction Using Character Correction and Feature-Based Word Classification." *Proceedings of the 12th IAPR International Workshop on Document Analysis Systems (DAS 2016), Santorini, Greece*. 198–203. https://doi.org/10.48550/arXiv.1604.06225

Komkov, Artem. et al. 2021. "Pervoye issledovaniye meditsinskoy informatsionnoy sistemy RuPatient po avtomaticheskomu raspoznavaniyu meditsinskoy dokumentatsii na osnove "mashinnogo obucheniya" = "First study of the RuPatient health information system with optical character recognition of medical records based on machine learning." *Kardiovaskulyarnaya Terapiya i Profilaktika* 20(8): 3080 https://doaj.org/article/18c34244808942f79a9bae16437bf8b2

Kordas, Marianne. 2021. "What We've Managed to Transcribe Thus Far: Lessons from the Blythe Owen Letter Transcription Project." *Music Reference Service Quarterly* 23(3-4): 153–170. DOI: 10.1080/10588167.2020.1816039 and https://digitalcommons.andrews.edu/cgi/viewcontent.cgi?article=2436&context=pubs

Liu, Zhi-Qiang; Jinhai Cai; Richard Buse. 2003. *Handwriting Recognition: Soft Computing and Probabilistic Approaches*. Berlin: Springer-Verlag.

Morgan, Richard. 1968. "Computers and the Humanities." *Annual Survey of Recent Developments* 3(1): 61–64.

Memon, Jamshed; Maira Sai; Rizwan Ahmed Khan; and Mueen Uddin. 2020. "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)." *IEEE Access* 8: 142642–142668. https://ieeexplore.ieee.org/document/9151144

Patel, Chirag Indravadanbhai; Atul Patel; and Dharmendra Patel. 2012. "Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study." *International Journal of Computer Applications* 55(10):50–56. DOI:10.5120/8794-2784

Peng, Xuju; Huaigu Cao; Srirangaraj Setlur; Venu Govindaraju; and Prem Natarajan. 2013. "Multilingual OCR Research and Applications: an Overview." *Proceedings of the 4th International Workshop on Multilingual OCR August 2013* 1:1–8 https://doi.org/10.1145/2505377.2509977

Plamondon, Réjean and Sargur N. Srihari. 2000.. "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1):63–84. http://ivizlab.sfu.ca/arya/Papers/IEEE/PAMI/2000/January/Handwriting%20Recognition.pdf

Plötz, Thomas and Gernot A. Fink. 2009. "Markov Models for Offline Handwriting Recognition: a Survey." *IJDAR* 12:269–298. DOI 10.1007/s10032-009-0098-4

Priego, Ernesto. 2011. "Father Roberto Busa: One Academic's Impact on HE and My Career." Blog post. *The Guardian* Friday, August 12, 2011. https://www.theguardian.com/higher-education-network/blog/2011/aug/12/father-roberto-busa-academic-impact

Schultz, Dan; Linda Mack; Marianne Kordas. 2022. "Owen, Vivian Blythe (1898–2000)." *Encyclopedia of Seventh-day Adventists* https://encyclopedia.adventist.org/article?id=CJ9R&highlight=y#fnref9

Suen, Cing Y. 1994. “Automatic Recognition of Handwritten Characters.” In *Fundamentals in Handwriting Recognition*. Edited by Sebastiano Impedovo. Berlin: Springer-Verlag.

Xie, Zhongwei; Lin Li; Xian Zhong; Yang He; and Luo Zhong. 2019. "Enhancing Multimodal Deep Representation Learning by Fixed Model Reuse." Multimedia Tools and Applications 78 (21) (11): 30769–30791. https://doi.org/10.1007/s11042-018-6556-6