Textension: Digitally Augmenting Analog Texts Using Mobile Devices

Adam James Bradley
University of Ontario Institute of Technology
Oshawa, ON L1H 7K4, Canada
adam.bradley@uoit.ca

Christopher Collins
University of Ontario Institute of Technology
Oshawa, ON L1H 7K4, Canada
christopher.collins@uoit.ca

Victor Sawal
University of Ontario Institute of Technology
Oshawa, ON L1H 7K4, Canada
victor.sawal@uoit.ca

Sheelagh Carpendale
University of Calgary
Calgary, AB T2N 1N4, Canada
sheelagh@ucalgary.ca

MobileVis '18 Workshop at CHI 2018, April 21, 2018, Montreal, QC, Canada.
https://mobilevis.github.io/

Abstract
In this paper, we present a framework that allows people who work with analog texts to leverage the affordances of digital technology, such as data visualization, computational linguistics, and search, using any web-based mobile device with a camera. After taking a picture of a particular page or set of pages from a text or uploading an existing image, our prototype system builds an interactive digital object that automatically inserts visualizations and interactive elements into the document. Leveraging the findings of previous studies, our framework augments the reading of analog texts with digital tools, making it possible to work with texts in both a digital and analog environment.

Author Keywords
Text Visualization; Mobile Visualization; Digital Humanities

ACM Classification Keywords
H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

Introduction
Analog text as a medium still remains persistent in the workflow of scholars even though there is a plethora of digital options available that afford great power and flexibility to the user. Word processors and other applications are ubiquitous and have started to replace pen and paper as a

modality for interacting with the written word; when it comes to books, the affordances offered by digital platforms such as search and copy are considered paradigm shifting additions to the act of reading. But, even though these tools exist, scholars still write on paper and still have books on their bookshelves. There is a tension that exists between these new digital formats and our history. We often create digital tools to mimic the affordances of books, but while they improve steadily, the weight, smell, and sounds of a book are still unique to bound paper and ink. Also, the ability to quickly digitize a document for augmentation but retain the look and feel of the original is important to many scholars [6].

Beyond the affordances of physical texts, many older documents used for research often do not have reliable digital versions, and many corpora are still digitized as images such as Early English Books Online [1]. The solution we offer is a combination of techniques that bring together the paper and ink history of our past, with the digital affordances of our present. To demonstrate the framework, we present a mobile interface, geared toward researchers, that allows for the quick digitization and augmentation of paper documents. By using any web-based device with a camera or uploading an existing image, our framework allows the user to create interactive digital objects from analog texts that retain the look and feel of the originals.

While there are great efforts to digitize the world's books, such as the Google books projects [2], large-scale OCR projects are difficult to implement. With the growth of digitized text repositories that leverage these technologies, two problems are still outstanding: digitally supporting books that have not yet been digitized, and enabling better use of books that have been digitized as images and are not currently interactive. The flexibility and freedom of digital writing and reading are leading to increasing pressure to digitize texts. However, most of these solutions are costly, time-consuming, and never seem to reach the document of current interest. There is a need for a quick, direct and simple way to gain these freedoms with the document you currently have in hand - whether it is a hand-written letter, an old book, or the newspaper.

In this paper, we present a framework that extends the power of the digital to physical books in near real-time. Our contribution is bringing together ideas studied in digital document spaces and existing word-scale visualizations to demonstrate how these known quantities can be leveraged to bridge analog and digital reading and writing. Our framework is informed by previous results describing text visualization and the different ways they are used with analog documents. Our prototype system offers quick access to an integrated digital/analog environment, only requiring common equipment such as the camera in a phone and a preferred web-enabled device. Simply photograph or upload an existing picture of a document, display it in the system on a web-browser and start interacting. By making paper documents interactive on mobile devices we allow for a smooth transition between our history and our present by allowing users a quick way to digitize documents while working on-site in places like libraries. Our system produces in-line visualizations and interactive elements directly on the newly built digital document, allowing for work to continue while having an augmented digital document at the ready.

Textension Framework

Previous studies on digital document interaction, annotations, e-readers, and marginal interactions [3, 5, 6] have identified five spaces that are interacted with on a digital document. Mehta discussed word space, line space, and margin space as places of interaction for analog annotation

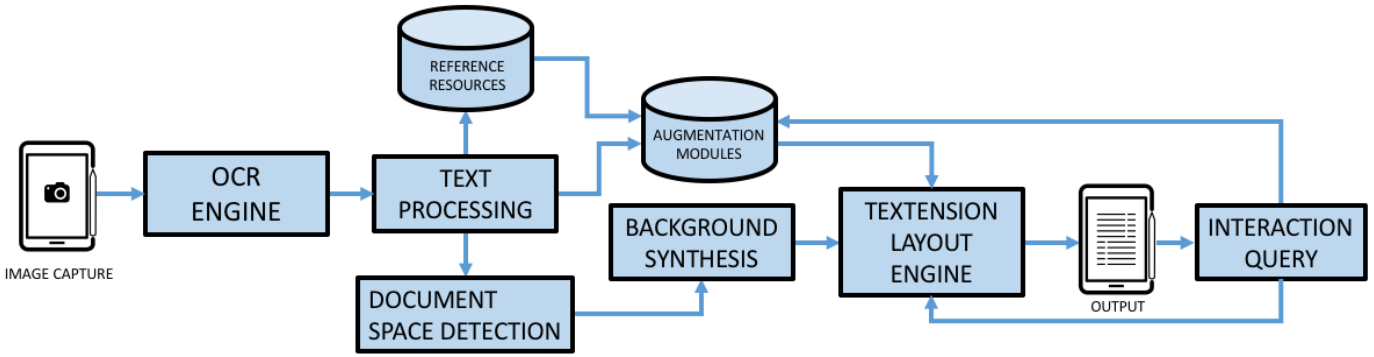

Figure 1: The Textension framework architecture: images are captured using a mobile device or web cam. The text is extracted by the OCR engine and the digitized text is processed using NLP techniques. The background of the document is detected and synthesized for the insertion of interactive elements. External resources are brought in as needed to create augmentations which combine with the document image to create the final layout on screen. Finally any interaction queries are fed back to the layout engine and augmentations and an updated output is generated.

[6] after studying literary critics working on poetry with pen and paper. Goffin et al. use word space and line space as alternatives for the placement of word-scale visualizations [5]. Occlusion space or the space above the document has long been an accepted interaction modality, the most famous incarnation being the everyday tooltip. Goffin et al. use this space to provide enlarged maps on text documents [5]. Canvas space was discussed by Cheema et al. in a paper outlining how to extend documents using external resources such as drag and drop images [3].

Framework Architecture

For this project, we imagined a tool that scholars of the humanities could use to do their required work on printed manuscripts, edited collections, and books, while still having access to digital affordances. One of the main problems that is ever present within the emerging field of the digital humanities is the roadblock of technical knowledge needed to produce tools. We set out to build an extensible framework that would allow a humanities scholar with limited technical knowledge the ability to process, augment, and export digital versions of analog texts. To achieve this, we bring together multiple technologies including OCR, machine translation, and information visualization.

The Textension framework starts with a document image, processes the image to discern both the content as well as the use of space on the page, adds space to the page as needed, and creates augmentations, both static and interactive, to insert into the image. The resulting processed image is presented to the user for further exploration, annotation, and interaction. The framework architecture is illustrated in 1.

Prototype

We provide a specific implementation of this framework in a web-based system which offers a selection of document augmentations and interactive tools, which we will describe in this section.

Image Capture and Processing

When a user comes to the opening screen of Textension they are presented with two input options. They can either use the camera that is built into their device (webcam, phone camera, front facing tablet camera) and take a snapshot of the document they wish to process, or they can upload an image file that has been previously prepared. Document images can be single or multiple pages and are uploaded with a drag and drop interface. The next stage of document processing begins immediately after the upload completes. The system uses image processing from the Python Image Library and image manipulation from OpenCV. We have found that a combination of binarization, grey-scaling, and image sharpening have had a noticeable effect on the results of the OCR, which is the next stage of processing.

OCR Engine

We used the open source Tesseract OCR engine in the Textension prototype. Smith provides an overview and a history of the development of the engine [8, 9], and Patel et al. provide a case study approach for its use [7]. Tesseract can be trained with many different languages and also with handwriting, making it a robust choice for an implementation such as this.

Background Synthesis

In order to augment documents with helpful annotations, or to provide space for users to make pen-based annotations, document spaces often need to be enlarged. This is not possible when working with an analog document. However, in the digital version, we can manipulate the image to provide the needed space. For example, to place a translation of text between lines, the inter-line spacing first needs to be increased. Document backgrounds can be complicated, with changing lighting conditions almost guaranteed using mobile phone and tablet cameras. To retain the original look of the document image, we created a method for inserting space by synthesizing sections of the document background which seamlessly integrate with the original.

To improve image capture quality, which can affect background synthesis, we provide the user with a frame to set their image in. While it is possible to adjust skew correction and automatically crop text from images, we found from our internal testing that forcing the user to frame the image themselves resulted in much better OCR and therefore a much better experience. There is precedence for this type of interaction in commercial settings such as remote cheque deposits for online banking, where a user is forced to frame and focus the cheque before the system will accept the image. Once we have the image we use the bounding boxes provided by the OCR engine to rebuild the document in image fragments within the web platform. Each space and word is modeled separately to allow us to manipulate those elements within the browser.

Layout Engine

After creating an interactive, expandable document from the captured image, augmentations can be added to provide supportive features as required for the specific task and context. For example, a student may require word definitions, while a literary scholar may be interested in the contemporary use of the words in the document. Augmentations can take the form of inserted glyphs, images, overlays, and annotations in the document spaces, or they may replace or change the words in the document. Augmenta-

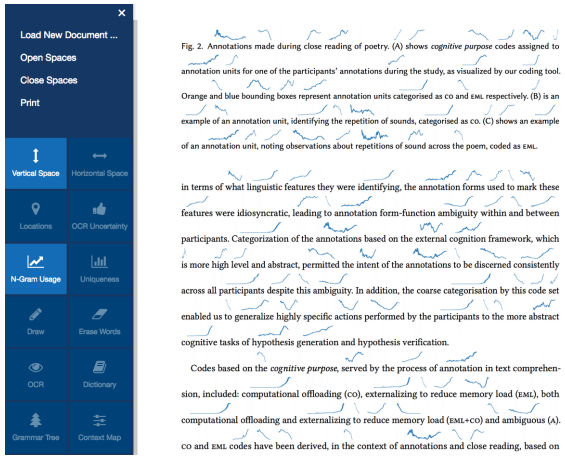

Fig. 2. Annotations made during close reading of poetry. (A) shows *cognitive purpose* codes assigned to annotation units for one of the participants' annotations during the study, as visualized by our coding tool. Orange and blue bounding boxes represent annotation units categorised as CO and EML respectively. (B) is an example of an annotation unit, identifying the repetition of sounds, categorised as CO. (C) shows an example of an annotation unit, noting observations about repetitions of sound across the poem, coded as EML.

in terms of what linguistic features they were identifying, the annotation forms used to mark these features were idiosyncratic, leading to annotation form-function ambiguity within and between participants. Categorization of the annotations based on the external cognition framework, which is more high level and abstract, permitted the intent of the annotations to be discerned consistently across all participants despite this ambiguity. In addition, the coarse categorisation by this code set enabled us to generalize highly specific actions performed by the participants to the more abstract cognitive tasks of hypothesis generation and hypothesis verification.

Codes based on the *cognitive purpose*, served by the process of annotation in text comprehension, included: computational offloading (CO), externalizing to reduce memory load (EML), both computational offloading and externalizing to reduce memory load (EML+CO) and ambiguous (A). CO and EML codes have been derived, in the context of annotations and close reading, based on

Figure 2: Sparklines showing lexical usage from the Google books corpus from 1800–2012. Words that the OCR did not recognize do not have a sparkline.

tions can be temporary or permanent, as appropriate for their purpose and the document space in which they appear. The insertion and placement of augmentations and the provision of interactivity on the document and its augmentations is provided by the layout engine. The images after upload are broken into individual word and space objects that are then recompiled in order onto an HTML canvas to reproduce the original image with the added flexibility of moving, inserting, and changing elements. Augmentations are placed on the canvas as a layer on top of the image objects. Textension has been developed to support the creation of new augmentations, which can draw on custom data processing, local datasets, or public APIs and data. Textension was built using flask, a python server back-end; bootstrap, for UI elements; jinja, a template engine for python; and jquery, for data handling.

Document Augmentations

What we present in this section are a series of concrete implementations of document augmentations that demonstrate a subset of the possibilities of the Textension framework. We explore insertion augmentations, as well as temporary and permanent overlays. The availability of the plain text allows easy integration of natural language processing, and the fact that the digital document is built in pieces allows for easy insertion of space to accommodate for the adding of new features. In this way, we envision Textension as both a sandbox for designing interactive elements for digital documents and a way to use both digital and analog affordances simultaneously when working with texts.

OCR Confidence

Often when digitizing analog texts OCR confidence is very important. Textension provides a feature where users can see an overlay of how uncertain the OCR algorithm was for each word. This augmentation is displayed as an overlay in the word space. The darker the color the less confident the score. This mapping was designed to draw attention to and 'obscure' those words that the system had difficulty recognizing. With the current trend in machine learning and the relative anxiety that is brought with black box algorithms, showing the inner workings of the OCR engine is a way to both help the user understand how the system is working but also where exactly work may need to be done to make for a better user experience and digital document.

Translation

Once the user has tuned the OCR to their liking, they can toggle the auto-translate menu button which will then use the Google translate API and automatically insert a translation of the text in the line space. This method works for all of the languages currently supported by Google and its one limitation is typographical, in that books often split words on

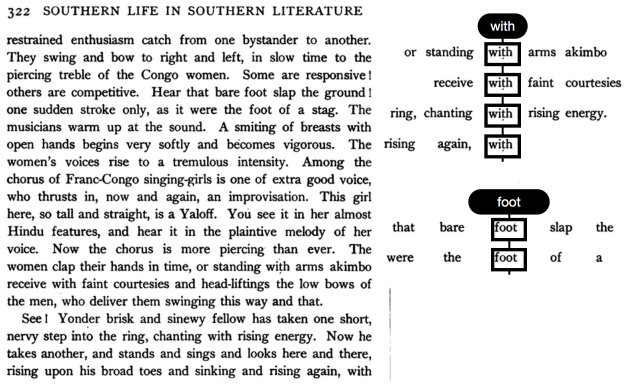

Figure 3: Context maps generated within the **canvas space** around a document image.

the end of lines. Future work will address this limitation.

Location Based Maps

With named entity recognition, we demonstrate the power of digital affordances with photographed texts by inserting maps. During pre-processing, we detect and store place names within the text. When the map feature is toggled Textension highlights the place name in the document, and automatically inserts a map from the Google maps API directly into the document in the margin space. This feature is a demonstration of the power of combining existing technology, such as the Google maps API with automatic document space expansion. Because the document is built in pieces we can freely move interactive elements into different document spaces to see which works best for the specific implementation.

Sparklines

Word space visualizations have been showing promise as ways to augment digital texts [4] Textension offers the ability to automatically insert these visualizations into images of analog texts. We have chosen to implement lexical usage sparklines [10] directly above each word showing the usage within the Google Books corpus from 1800-2012 2. This technique could be used for many different types of visualizations limited only by the power of the OCR and NLP techniques available.

Context Maps

A context map lists all of the ways that a particular word or phrase has been used within a document. This digital affordance uses canvas space to build interactive concordance lines that highlight the four words before and after the word in question. The maps are built using the images patches of words in the document to maintain the document aesthetics and reduce the impact of OCR errors. This is an example of the types of things that can be done with ready access to linguistic information and expandable canvas space 3.

Conclusion

The tension that exists between our analog pasts and our digital present can be addressed using our mobile framework. Our prototype leverages the power of OCR and digitally manipulates documents in near real-time. The system we present is an implementation of previous studies brought together in a way that can be extended easily for domain-specific analysis tasks. The web-based platform allows for easy integration with mobile technology and makes it possible to use the framework in a variety of locations and scenarios.

REFERENCES

1. 2003. EEBO - Early English Books Online. http://eebo.chadwyck.com/home. (2003). Accessed: 2017-09-18.

2. 2017. Google Books. https://books.google.ca/. (2017). Accessed: 2017-09-18.

3. Muhammad Faisal Cheema, Stefan Jänicke, and Gerik Scheuermann. 2016. AnnotateVis: Combining Traditional Close Reading with Visual Text Analysis. In *IEEE VIS Workshop on Visualization for the Digital Humanities*. 4.

4. Pascal Goffin, Wesley Willett, Anastasia Bezerianos, and Petra Isenberg. 2015. Exploring the Effect of Word-Scale Visualizations on Reading Behavior. In *Proc. ACM Conf. Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1827–1832. DOI: http://dx.doi.org/10.1145/2702613.2732778

5. Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. 2014. Exploring the placement and design of word-scale visualizations. *IEEE Trans. on Visualization and Computer Graphics* 20, 12 (2014), 2291–2300.

6. Hrim Mehta. 2015. Augmenting Free-form Annotations with Digital Metadata for Close Reading of Poetry. (2015).

7. Chirag Patel, Atul Patel, Dharmendra Patel, Archana A. Shinde, Hui Wu, and Jian Liang. 2012. Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study. *Int. Journal of Computer Applications* 40, 10 (2012).

8. R. Smith. 2007. An Overview of the Tesseract OCR Engine. In *Proc. Int. Conf. on Document Analysis and Recognition (ICDAR)*, Vol. 2. 629–633. DOI: http://dx.doi.org/10.1109/ICDAR.2007.4376991

9. Ray W. Smith. 2013. History of the Tesseract OCR engine: What worked and what didn't. *Proc. SPIE* 8658 (2013), 865802–865802–12. DOI: http://dx.doi.org/10.1117/12.2010051

10. Edward Tufte. 2004. Sparklines: Intense, simple, word-sized graphics. In *Beautiful Evidence*. 46–63.