2-2017

Tandem 2.0: Image and Text Data Generation Application

Christopher J. Vitale
The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Digital Humanities Commons, Interdisciplinary Arts and Media Commons, Other Computer Sciences Commons, and the Visual Studies Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/1870

This Capstone Project is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
TANDEM 2.0: IMAGE AND TEXT DATA GENERATION APPLICATION
by
CHRISTOPHER VITALE

A capstone project submitted to the Graduate Faculty in Liberal Studies in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York.
2017
This manuscript has been read and accepted for the Graduate Faculty in Liberal Studies in satisfaction of the thesis requirement for the degree of Master of Arts.

____________________ ______________________________
Date  Dr. Matthew K. Gold
Thesis Advisor

____________________ ______________________________
Date  Dr. Elizabeth Macaulay-Lewis
Acting Executive Officer

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

TANDEM 2.0: Image and Text Data Generation Application: A Case Study

by

Christopher Vitale

Advisor: Dr. Matthew Gold

First created as part of the Digital Humanities Praxis course in the spring of 2012 at the CUNY Graduate Center, Tandem explores the generation of datasets comprised of text and image data by leveraging Optical Character Recognition (OCR), Natural Language Processing (NLP) and Computer Vision (CV). This project builds upon that earlier work in a new programming framework. While other developers and digital humanities scholars have created similar tools specifically geared toward NLP (e.g. Voyant-Tools), as well as algorithms for image processing and feature extraction on the CV side, Tandem explores the process of developing a more robust and user-friendly web-based multimodal data generator using modern development processes with the intention of expanding the use of the tool among interested academics. Tandem functions as a full-stack JavaScript in-browser web application that allows a user to login, upload a corpus of image files for OCR, NLP, and CV based image processing to facilitate data generation. The corpora intended for this tool includes picture books, comics, and other types of image and text based manuscripts and is discussed in detail. Once images are processed, the application provides some key initial insights and data lightly visualized in a dashboard view for the user. As a research question, this project explores the viability of full-stack JavaScript application development for academic end products by looking at a variety of courses and literature that inspired the work alongside the documented process of development of the application and proposed future enhancements for the tool. For those interested in further research or development, the full codebase for this project is available for download.
Table of Contents

ABSTRACT ............................................................................................................................................................ iv
I. Project & Build Process Description ........................................................................................................... 1
II. Relationship to Track and Previous Course of Study .................................................................................... 9
III. Project Evaluation & Plans for Continuation ............................................................................................. 22
Appendix A. .......................................................................................................................................................... 25
Appendix B. .......................................................................................................................................................... 26
Appendix C. .......................................................................................................................................................... 27
Bibliography .......................................................................................................................................................... 29
I. Project & Build Process Description

The goal of this project is to expand access to combined language and visual based datasets generated from image based corpora. These image and text based corpora can include, but are not limited to, picture books, comics, newspapers, magazines, and other digitized manuscripts. For those interested in studying these kinds of media, there are limitations as to how to read and conduct research on them. These limitations point to a delta that exists between the methodologies of scholars studying textual literature, visual art, and those who want to study media containing both. As digital research methods, namely quantitative and data driven, are enabling new and interesting ways of dissecting the text or image based media, Tandem seeks to enter the conversation as an advocate for the application of quantitative methods for material containing both. In building Tandem, I am conscious of the increasing creation rate for multimodal media and I understand that the research methods associated with them are required to evolve and mature as a result. In an effort to encourage quantitative and technical scholarship in this area, Tandem marries existing technologies to create multimodal datasets.

This was first addressed when Tandem was built as part of the Digital Humanities Praxis Seminar in the spring of 2015 at the CUNY Graduate Center. The problem that both the original and this project continue to address is the inability to efficiently generate multimodal data for these types image and text based corpora in a single user interface and unified workflow on the web. Because of this, the target audience remains those interested in multimodal data extraction for the application of computational methods on images containing text. In the initial iteration of this project, I lead three fellow CUNY Graduate Center students, Stephen Real, Jojo Karlin, and Kelly Blanchat, to build a minimum viable product (MVP) version of the application. We defined the target users as generally having an inability or insufficient amount of time to develop
their own applications or data scrapers for computer vision based data extraction. The target users also want to expedite the process of retrieving usable data from the image and text based corpora they are researching. One such example is a researcher interested in studying the changing nature of visual and narrative language across a series of children’s picture books from the mid to late 20th century. Questions regarding the relationship between color choice and vocabulary could be explored in tandem to one another. Is there a relationship between the colors and themes within this particular corpus? The impact of this research would be cross-disciplinary and effect children’s literature, media studies, art history, etc. With the growth of computational methodologies being applied in the humanities, a need emerges for quick and easy access to data beneath the surface of such source materials.

Tandem’s first iteration employs a Python-based Django web application; a framework that was chosen because it contained the necessary pieces to quickly build a functional web application. Django allows one to manipulate the user interface that appears on the browser all the way through to the scripts where the data is generated and manipulated in Python. Using Django as the main infrastructure, we made use of the NLTK library for natural language processing and the OpenCV for computer vision and image processing. In order to generate the text data from the image, pytesseract was used to apply pretrained optical character recognition (OCR). Outside of these core functionalities, other components such as PDF file handling, CSV export, and project referencing are also written into the framework in Python. The decision to use Python comes from its status as common language for scholars interested in programming and because it is often used as the entryway for scholars outside of the computer science department to foray into programming as a methodology. The original Tandem remains fully functional at the time of this paper at the live URL http://www.dhtandem.com. Although it is still
live, the first iteration did have a variety of shortcomings that are addressed in this new version including complaints about the user interface, the inability to make accounts to house information for later access, and lack of in-browser visualization of the data. These are a few of the feature requests that I set out to improve upon and address with this project.

Before a single line of code was written for the new version of the application, it was necessary to define what the application would look and feel like. To define a new end product, a series of scoping processes were followed beginning with the creation of a user stories document. Building off of the broad definition of what the user personas are we are able to define more specific user stories. These user stories are succinct actions that each user persona could be expected to carry out on the application, and can vary in descriptiveness and priority. User stories are generally plain language feature descriptions and requests. As they are written, each story should follow a similar format such as: “I as a user do X.” One then compiles these individual user stories into a document to collect, prioritize, and track the progress of the individual experiences. The goal of this process is to clearly identify an action that translates into a specific functionality that contributes to the larger components of the application that being built. Also, one could use this process to identify areas that require more than one development action item. Once these stories are written, one should review them to check whether they can then be broken into more granular user stories or left to embody a specific user flow needed for the site. One can find the user stories defined for Tandem can be found in Appendix C.

The second process that drove progress for this project was an agile development strategy. More than a buzzword, agile development workflows allow for a constantly evolving end product that is driven by features rather than timing or resources. This is different than a traditional waterfall build process that requires building one piece after another in a specific
consecutive order that is heavily reliant on timing and resources. By adopting an agile development mentality, a minimum viable product was spun up in a shorter amount of time. Once the minimum is created, iterative developments of additional functionality are layered over the application. Over time the application becomes more robust and complex. For each week of the project build, a set of user stories is chosen and the associated features are added to the development queue. Among this queue, blocks of feature requests, referred to as sprints, are completed and each sprint is followed up with a sprint review to identify if too much or too little was included in the sprint. One would then ensure that the sprint following the review takes previous findings into consideration. While this works best with a team, being diligent about this process as a solo developer helps to keep progress consistent and the project moving forward.

These processes are technology and project agnostic considerations that differentiate this project from its predecessor. Build process aside, the main differentiator is the technology stack that Tandem is built on. JavaScript offers an alternative approach to data extraction and more flexibility for the user interface of the application. For this reason, Tandem is built in a JavaScript framework called Meteor.js (Meteor), a full-stack framework that allows developers to quickly build upon preexisting client and server side JavaScript to execute the full range of functional components of the application. The Meteor Guide, accessible at www.meteor.com, summarizes some of Meteor’s key benefits as the following:

“Meteor allows you to develop in one language, JavaScript, in all environments: application server, web browser, and mobile devices. Meteor uses data on the wire, meaning the server sends data, not HTML, and the client renders it. Meteor embraces the ecosystem, bringing the best parts of the extremely active JavaScript community to you in a careful and considered way. Meteor provides full stack reactivity, allowing your UI to seamlessly reflect the true state of the world with minimal development effort” (Meteor.com).
Meteor’s file structure as well as its unique way of handling dynamic and static information eliminates the need to manually create key components of the application infrastructure from scratch. Further, the boilerplate solutions for user registration, account authorization, session handling comprise a starter pack for multi-user applications.

The application was initiated locally on a MacBook Pro (Retina, 13-inch, Mid 2014) with a 2.6 GHz Intel Core i5 processor and 8GB of RAM. This standard consumer model laptop was chosen to ensure that end-users or future developers would have the ability to replicate the project. It was also chosen to prove that these types of projects do not require a significant amount of expensive hardware. On this laptop, Tandem was initiated as a basic Meteor build using terminal to install the necessary initialization items which took a few hours over a two-day period. This basic application was up and running without data generation or manipulation scripts on the local machine host. To track changes in the build process from this starting point, as well as to identify milestones of the build, a GitHub repository (also referred to as a repo) was set up under the name tandem2. GitHub remains the one of the largest code hosts in the world with millions of active users and an evolving ecosystem of both commercial and open source developers. The GitHub repo can be accessed and forked here: https://github.com/CVDH4/Tandem2. Finally, the application is tested across multiple device sizes including desktop, tablet, and mobile devices running Chrome for functional and flow related issues such as uploading an image, generating data, as well as the login flow. The application has been proven to work on Chrome, Safari, and Firefox. Internet Explorer and Microsoft Edge are understood as unsupported browsers and ones that would not be consistently tested in due to known issues relating to the handling of stylesheets and JavaScript as a language.
In order to make Tandem fully functional cross-browser, additional development knowledge is required.

One key improvement to Tandem is the redesign of the user experience and interface. The Meteor framework is equipped with multiple front-end user interface integrations that take advantage of existing libraries of preset styles, which allow developers to quickly prototype the front-end without the need to write custom code. Most, if not all, general components of a site including headers, footers, buttons, containers, etc., are all stylistically predefined. This saves a significant amount of time that would be spent on creating these elements from scratch. There are limitations as to how things are styled out of the box, but customization of these elements is also possible if necessary. For this project, Semantic-UI, a tool that uses a series of class-based tags to call a larger set of stylesheets within the application, was used to streamline the front-end development portion of the application build process. In HTML/CSS a class is a way of identifying similar elements on a web page. The class identifier tells the system to call a specific set of styles in the predefined Semantic-UI library to render in the browser on the screen. Semantic-UI, similar to its more widely known equal in the space, Bootstrap, improves the aesthetic experience on the screen and makes for a more intuitive user experience.

Behind the clean exterior of the application, some of the core functionalities of Tandem are authored by other active developers in the JavaScript community and are governed by MIT, Apache 2.0, and Creative Commons software licenses. This project builds upon those works by connecting these independent pieces that allow for the creation of natural language and computer vision based datasets of images that contain text. While a full list of the libraries, packages, and pre-existing code used for the application can be located and tracked in the GitHub repo, a selection of some of the most important and influential dependencies can be found below. This
includes Tomi Rescak’s *meteor-uploads*, Jos de Jong’s *Math.js*, and *tesseract.js* authored by Project Naptha.

Tomi Rescak’s *meteor-uploads* is the package used to handle the file upload process for Tandem that is designed specifically for Meteor as the name suggests. Using the code written by Tomi Rescak with minimal modifications, file upload is successfully functional on the application and a user can watch the upload progress in the user interface and will be prompted if there is a failure or issue. The file upload process takes single or multiple files as input and places them in a specified directory and creates that directory if for some reason it has been removed from the code or file structure to ensure that it always exists. Currently, there is no limitation to the file type or size being uploaded, but it is recommended that a jpeg, png, single page pdf, or gif file type be used. An earlier adaptation of this process is currently live on the Meteor production environment, but the more robust feature described here is part of a future deployment scheduled in the next month. In this version, the image is simply pushed to the canvas view for data generation increasing speed and efficiency.

The images themselves are processed and the data is generated once they render on the browser screen. When a user sees their image appear, they also see the results of the images visual characteristics being extracted. This includes the size and shape of the image, its title, and some color related statistics including the color that appears most often. That statistics is generated using the standard Canvas APIs calls that are compatible with modern web browsers. It is recommended that Mozilla Developer’s Network (MDN) be consulted for further explanation of this and other JavaScript APIs. MDN was used to explore the proper implementation of image related web APIs including Canvas and ImageData. To augment these web APIs, *Math.js* is leveraged once the visual data is generated from the image for basic
mathematical calculations. Tandem’s mathematical logic is improved in this version of the application by using the predefined math functions from this package, which reduced the need to create complex original algorithms to solve basic equations. One such example is the ability to quickly select the mode from a large array of arrays containing the RGB values for the uploaded images. This array of arrays contains data for each and every pixel of the image. A high-resolution image could contain millions of pixels. Calculating the mode from the millions of sets of RGB values equates to the quick and efficient retrieval of the most common color from the image. The original Tandem calculated each RGB value and its mode and standard deviation individually rather than as sets. Taking into account the importance of these RGB values in relation to one another, we calculate the mode as an array of arrays, or more simply as a group of sets. Which set appears most among the larger group is pulled into a table for the user to reference.

The text data is generated via Project Naptha’s tesseract.js, a Meteor package that wraps the popular Tesseract OCR technology that was created by Google and is used in the original Tandem. This package is one that required some customization to properly extract text data from the images in a local environment. When the image is uploaded, it initializes a call to an external content delivery network (CDN) containing the Tesseract code. The call retrieves the necessary trained OCR data for the language selected from a dropdown menu. By using the Tesseract API, the application is able to support a large range of languages. Once completed, the text appears in a selectable text area ensuring easy copying and pasting. The accuracy of the Tesseract API is reliant on the clarity of the image. Some areas of the image may be misread as text causing extra characters to appear in the results. This is an area that could be improved in future iterations via a
text area selection step for the application to mitigate the risk of looking at areas where there is no text.

II. Relationship to Track and Previous Course of Study

Tandem’s roots stem from a more general academic interest in picture book study. This interest was seeded while taking the Digital Humanities Praxis course, which introduced the methodologies that Tandem unifies. Some of the readings from that course are discussed below in more detail and parts of the below discussion were previously published as part of Picture books, a CUNY Commons digital annotated bibliography project that sought to explore the literature in the field of picture book study under a digital humanities lens. The work was originally completed as part of Professor Carrie Hintz children’s literature course and has been adapted for the purpose of discussion. In order to understand the impetus for this project, a selection of courses is also discussed below. To answer a broader question regarding motivation, the picture book itself is in a moment of evolution much like the methodologies Tandem seeks to augment. At a more fundamental level, the purpose of this type of data extraction is driven by an interest in the application of computational methods for distant reading bodies of literature that are comprised of digitized images with overlaying text. The picture book is an example of this type of literature.

This project is also spurred by the work of scholars working in the field of digital humanities and pays primary focus toward the methodology of distant reading. This is a contrast to the traditional methodology of close reading that focuses in on certain words and phrases of a text and expounds meaning from them. Distant reading takes a large step back to look at a larger corpus in its entirety to identify trends. Tandem’s intention is to take distant reading of text and combine it with distant reading of visual elements of the aforementioned corpora types. This
methodology is one that could be tracked to canonical digital humanities literature like *Graphs, Maps, Trees* by Italian literary scholar Franco Moretti. This text outlines the efforts of Franco Moretti and his team in their journey to distant read multiple centuries worth of literary work. The graphs, the maps and the trees are explained as valuable tools for not only natural and social scientists, but also literary scholars (Moretti, 2007). The concept of distant reading is defined as the, then groundbreaking, application of an analytic and quantitative approach to reading wide ranges of texts (Moretti, 2007). Moretti’s historiography of the literary canon illuminates the merit and value of distant reading as a legitimate methodology. Moretti sets out to draw a roadmap for how quantification and visualization can augment, complement, and in certain cases completely topple traditional literary research methodologies. Tandem is built with this mentality in mind.

Computer scientist David Mimno follows in Franco Moretti’s footsteps as he data mines a massive archive of classics journals in his 2012 paper "Computational Historiography: Data Mining in a Century of Classics Journals.” Distant reading is, again, the preferred methodology for Mimno who has identified the ability close read such a large corpus as an unrealistic endeavor that would take far too long and would be far too complex. We can think about this type of data mining as the step that Tandem seeks to simplify for scholars. In his paper, Mimno discusses the use of computational tools that allow for the statistical analysis of his corpus. He makes sure to identify that this work is explicitly complimentary to his traditional scholarship. Tandem seeks to simplify is the time that researchers will save for the manipulation, analysis, and exploration of their corpora. The collection that Mimno is working with is the result of the application of optical character recognition to more than twenty classical philology and archaeology journals (Mimno, 2012). Whether or not these journals contained images along with
their text is unimportant. It is understood that this project was primarily interested in the text objects of these journals. Outlining the tools used in statistically driven mining of texts, Mimno discusses tokenization, removal of stopwords, word distance and divergence, and topic modeling (Mimno, 2012). These natural language processes are core areas where Tandem can alleviate pressure. The algorithmic representations of these computational methods are given as well as an introductory discussion of the ways they work and are used (Mimno, 2012). Just a few years later, and much of what Mimno discusses are standard practices found in a variety of libraries like the one Tandem is calling upon. Finally, Mimno presents his findings in the forms of graphics, topic models, and observations (Mimno, 2012). The data that Mimno extracts and manipulates is the same in nature that Tandem produces.

Distant reading began taking a stronger foothold in my own academic interest as I read of researchers like Matthew Jockers and David Mimno who began teaming up and applying the methodology in interesting ways. In 2013 they published “Significant Themes in 19th-Century Literature” where they discuss their distant reading project that sought to map and explore relationships between the themes of 19th century literature. Jockers and Mimno mined and modeled topics from over 3,300 works of literature (Jockers & Mimno, 2013). The corpus includes British, American, and Irish texts (Jockers & Mimno, 2013). Taking a variety of other factors into consideration the two seek to find trends in the themes of novels from that century. Their process incorporates the counting of words, tokenization of the text as well as more complex topic modeling using MALLET as the main engine for mining the text (Jockers & Mimno, 2013). In the article, Jockers and Mimno discuss the preprocessing elements (stopword removal, bag of words segmenting of the text, part of speech tagging of nouns, and modeling of the topics) (Jockers & Mimno, 2013). These are natural language processes that clean the data
and prepare it for analysis. The analysis and observations that are made are the result of robust data generation methods like the ones that Tandem seeks to achieve. One of the key assertions of this text is that topic modeling is shown to be a scalable solution for those interested in reading massive selections of hundreds and thousands of books (Jockers & Mimno, 2013). Topic modeling is a feature that remains in the queue for development for future iterations of Tandem.

Having first read of the work in Art History as part of the Praxis course and later as part of Computational Problems in Digital Humanities taught by Professor Scott Dexter of the CUNY Graduate Center, it became evident that the humanities scholars would need to borrow from other disciplines to execute successful research projects. In a 2014 paper titled “Toward Automated Discovery of Artistic Influence,” Babak Saleh, Kanako Abe, Ravneet Singh Arora, and Ahmed Elgammal outline their findings while working in the Department of Computer Science at Rutgers. They experimented with corpora of art and applications of computer vision. Their addition to the field was their use of advanced computer driven recognition of images to explore the influence of other artists on particular pieces of art (Saleh et al, 2014). The study keyed in on two primary types of computational inquiry: “discriminative vs. generative models” as well as feature extraction and comparison (Saleh et al, 2014). By doing so, they positioned an argument that computers had the ability to recognize the influence of artists in multiple works of art (Saleh et al, 2014). An automated process involving trained computer vision algorithms and a system of classification is argued to be able to do the work of an art historian. Tandem builds on a comparable type of inquiry. The computer algorithms in this article are paired with art history analysis of the paintings similarity, and traditional and novel methodologies are compared to one another (Saleh et al, 2014). This paper is a valuable source of tools and methods for computer
vision in applications for illustrations and it also makes a parallel case for the legitimacy of this type of methodology in this academic setting.

Much like this paper, the Computational Problems in Digital Humanities course attempted to unify the humanities and computer science scholars in a collaborative environment. Unfortunately, the course did more to unearth an attitude of disdain between the scholars. Many people quickly dropped the class under the pretense that the two mentalities could not collaborate with one another. On one hand, the digital humanities students brought a fresh thought process and strategy to a new body of source material. On the other hand, computer science students were baffled as to why a literary scholar was interested in algorithm creation and applied computer vision techniques. Rather than focusing on the computational problems that digital humanities scholars faced, the students on both sides centered the discussion on the apparent disconnect between how and why computational methods were being applied in the first place. The legitimacy of methodology is an area that digital humanities scholars often grapple with in the academic setting. In 2014, Emily Spratt of Princeton University self-published a paper entitled "The Digital Humanities Unveiled: Perceptions Held by Art Historians and Computer Scientists about Computer Vision Technology" that recounts a survey completed by both art historians and computer scientists in relation to a computer's ability to interpret aesthetic and beauty. The work's significance is contained in the responses to the survey that she conducted. It was already known at this point that computer vision was evolving into a more accessible method for examining art. The question of whether it was acceptable remained. For art historians and computer scientists, the implications of automation are obvious and draw parallels from the fears of automation in the private sector. Fear of obsolescence is one that is not hard come across in academia. Spratt’s digital humanities project used “twenty-one questions for art historians and
sixteen for computer scientists that were intended to shed light on field members’ knowledge of the capabilities and applications of computer vision technology, attitudes and perceptions about the use of it, and reactions to the meaning of this type of digitization in the humanities” (Spratt, 2014). Spratt discusses the positive and negative reactions to computer vision’s ability to detect and automatically recognize aesthetic experiences of beauty. Channeling philosophy, Spratt defines what these variables mean for her survey (Spratt, 2014). The results are less than staggering. The emotional response is mixed and the legitimacy of the methodologies is consistently questioned. It becomes less of a question of whether it will impact both fields and more of a question of how it will impact both fields. Tandem seeks to enter that space for those who are interested in testing this kind of research for themselves. Whether the computer science students abandon their bias and become involved in the digital humanities’ projects is unimportant.

While studying with Professor Carrie Hintz in the Children’s Literature department at the CUNY Graduate center, I unearthed additional literature regarding the picture book as an object of study. Tandem had already been built and this second project was in its beginning stages. It was during this class with Professor Hintz that questions regarding the picture book began to inspire more interest and discussion. That discussion was often cross discipline as students came from English, Psychology, and Children’s & Youth studies backgrounds. We discussed the relationship between two types of data within picture books, text and image, as one that is explicitly cross discipline. These discussions come with questions of necessity and value, a problem that Christina M. Desai’s “Weaving Words and Pictures: Allen Say and the Art of Illustration” makes an effort to solve. Her positioning of the study of illustrations found in children’s literature as legitimate enough for inquiry emphasizes the subtleties of illustration.
Desai examines the relationship between text and illustration read through its work as visual narrative, illustrated cues, as well as perspective and meaning in three books illustrated by Allen Say: Dianne Snyder’s *The Boy of the Three-Year Nap* (1988), *El Chino* (1990) and *Emma’s Rug* (1996) (Desai 2004). Graphic art is shown to be a central driving force and fundamentally necessary element of the story in illustrated picture books (Desai, 2004). Children’s literary studies should be credited for the various ways of to reading and interpreting the information and relationships between text and image in these combined objects.

In “A Multimodal Analysis of Image-text Relations in Picture Books” by scholar of linguistics Shuxuan Wu, a Systemic Functional Multimodal analysis is used in an effort to illuminate image-text relationships. Wu completes a thorough multimodal classification and construction of meaning that is followed by a brief overview of Systemic Functional Linguistics (Wu, 2014). Wu explains the choice to use picture books is because the genre contains a wide array of image-text relation examples (Wu, 2014). Wu defines categories for image-text relation types: Elaboration, Extension, Enhancement, and Projection (Wu, 2014). Wu breaks the analysis into a multi-level analysis presents a generic framework of picture books (Wu, 2014). From a culture context, context of situation, to the content stratum, down to the display stratum, a new perspective for non-linguistic scholars is provided (Wu, 2014). Wu concludes that the text and images in picture books contribute autonomously, but that the image-text relationship is complex and not as simple as complementary, extensible, etc. (Wu, 2014). Wu’s work inquires into the complexities of not only picture books, but these types of multimodal objects at large. The technical sophistication here is an important component to address as well. Wu approaches the problem as a computer scientist interested in identifying relationships that would otherwise be of interest to humanities scholars.
The evolution of media into digital objects and more broadly digitization is a topic that is discussed regularly at the CUNY Graduate Center. Digitized and born digital media are entering into the syllabi at CUNY at all levels. We look to the evolution of the media itself as an opening for how we read and research that media. In 2014, authors Frank Serafini, Dani Kachorsky, and Earl Aguillera of “Picture books 2.0: Transmedial features across narrative platforms” write of their concerns with the transmedial features of the text as well as the work being done by visual images, sound fx and music, textual and paratextual elements of the story, as well as a range of interactivity points within the text such as navigational elements, transitions, and animations. The printed picture book is very different than the experience the same narrative on a digital device. For Serafini and the team, the content of the book is something that can now be experienced in more places and more ways (Serafini et al., 2015). The digital platform alters the experience that one can have with a children’s narrative (Serafini et al., 2015). As a result, the meaning extrapolated from stories changes as well (Serafini et al., 2015). For picture books published in print format, digital formats present a new range of entry points for multimodal inquiry. The main primary source for the researchers in the above paper is The Fantastic Flying Books of Mr. Morris Lessmore, a text that was published in a print and digital form (Serafini et al., 2015). The print version of this text is a traditional static object that can be digitized and pushed through data generation processes, whereas, the digitized version of this is more akin to an application with interactivity and motion graphics. Tandem is not yet equipped to handle such objects as it remains interested in the more traditional version of the object. That said, this more complex media offers interesting new problems for multimodal research and data generation like how to handle motion graphics and sound.
Beyond that evolution, some researchers like Lisa M. Schons go so far as to claim the picture book is on its way out. This adds a level of urgency to the field where researchers must explore the value, or lack of value, of traditional picture-book media. In 2011, Schons published “Is the Picture Book Dead? The Rise of the iPad as a Turning Point in Children’s Literature.” She observed that from 2010’s release of the Apple iPad, digital publishing saw a drastic change. More so, the picture book genre saw a wave of new possibilities (Schons, 2011). Schons explores the market effect as well as the radical reconceptualization of the genre in this article (Schons, 2011). Using *The Heart and the Bottle* as an example, Schons notes the differences between the traditionally printed picture book and the new experience driven interactive iPad picture book (Schons, 2011). Multimodality is positioned as the new normal for digitally published picture-book apps with an emphasis on the blurring of lines between movies, games, and books (Schons, 2011). Schons also presents the genre of born digital e-literature as a revolutionary tool for improving the literacy and digital literacy of a child (Schons, 2011). One important conclusion to note is the statement that the two different formats, traditional and digital, have a place in the complex and rapidly changing market. One could argue that a digitally focused picture-book market will influence higher levels of adoption for projects like Tandem as the born digital media gives way to digital research methodologies.

The digital age has altered the way we read and study the picture book. Ghada Al-Yaqout of the College of Basic Education in Kuwait and Maria Nikolajeva of the University of Cambridge maintain the need to expand the theory beyond the multimodal text and image understanding of visual and verbal understandings to include the new elements of multimodal picture books: “auditory, tactile, and performative dimensions” (Al-Yaqout & Nikolajeva, 2015). The article looks at both digital stories as well as apps that are shaping the way visual texts are
delivered and interacted with (Al-Yaqout & Nikolajeva, 2015). The immateriality of this new way of digital texts is brought forth for consideration (Al-Yaqout & Nikolajeva, 2015). The researchers explore the impact of the exploitation of interactivity, the destabilization of the confines of layout and space found in printed texts, and the performance of reading, being read to, or following along (Al-Yaqout & Nikolajeva, 2015). What affect does this type of corpora have on projects like Tandem? Books that are now available in both print and augmented digital formats, like *The Cat in the Hat* and *Goodnight Moon*, are juxtaposed and discussed (Al-Yaqout & Nikolajeva, 2015). Digital children’s literature commands a need for a reevaluation of picture-book theory and the research methods associated with the genre. This need to rewrite picture-book theory to accommodate digital texts offers the perfect moment for a project like Tandem to enable researchers to explore the traditional media in new ways.

Professor Eliza T. Dresang of the University of Washington and associate professor Bowie Kotrla of Florida State University begin their 2009 paper “Radical Change Theory and Synergistic Reading for Digital Age Youth” by discussing radical change theory in relation to the award-winning extended length picture book, *The Invention of Hugo Cabret*. Dresang notes that this text is an indication of the changing nature of picture books in response to “digital age youth” (2009). The changing reader is one that is now concerned with interactivity, connectivity, and access (Dresang & Kotrla, 2009). Radical Change Theory dictates that books must become more sophisticated in order to keep up with the demand of its changing audience (Dresang & Kotrla, 2009). As books become more sophisticated, so must the ways we research them. An important part of this article is the discussion of “Changing Children” (Dresang & Kotrla, 2009). The preconceptions we have about children and childhood are rapidly changing into more complex and ambiguously defined identifications (Dresang & Kotrla, 2009). By acknowledging
Radical Change Theory, we can more accurately assess books created in the digital age or for the generation of people consuming this type of media. We are forced to think about the way we read in new terms. That shift in production toward digital first may offer opportunity for more distant reading projects in the future through more expedited means.

The close reading of children’s literature is one that should also be challenged. We should use tools like Tandem to attempt to alter the way we previously understood this type of literature. In a 2010 article titled “Focalization in Children’s Picture Books: Who Sees in Words and Pictures”, Angela Yannicopoulou discusses the importance and implications of our predominantly visual culture on children’s literature. Yannicopoulou outlines the differences between types of focalization starting with nonfocalization where the main point of focus is the characters, moving to internal focalization where the narrator is seeing the story through his or her own eyes, onto fixed internal focalization where the story is given to the reader through the restricted view of a single character (2010). Moving beyond these, variable internal focalization, multiple internal focalization, external focalization, and discrepant modes of focalization are discussed in depth (Yannicopoulou, 2010). In doing this kind of defining of focal points and perspectives, Yannicopoulou considers the assumptions that must be made from an ideology standpoint for those reading picture and non-picture books. The role of pictures for each of the aforementioned forms of focalization is varied, but overall facilitates a type of assumptive understanding of sociological, cultural, and even historical focalization of the text, its characters, and its world. The narrative text works in conjunction with the images in order to push forward verbal and visual story information. This work is done at a close range to the picture-book media. It would be interesting to take this type of observation under a distant reading lens to see if it changes.
Children’s picture books are more than a 20th century creation, but rather an integrated aesthetic experience consisting of concepts and images that has been around for much longer. Barbara Kiefer assigns the art in these texts as having the ability of creating meaning (2008). The core of Kiefer’s work explores the form and formats of the picture book as it evolved from sequential illustrated picture books, comic strips, graphic novels, to non-sequential fiction and nonfiction in the picture book (2008). All of these formats have widened the audience for the form of the picture book (Kiefer, 2008). Technological developments, such as woodblock versus moveable type, computers, laser scanners, and color separation are given their due credit for creating the possibilities of mass-produced picture books (Kiefer, 2008). Media developments like black-and-white to full color, computer generated to mixed media, have also had a profound effect on the picture book (Kiefer, 2008). Most compelling is Kiefer’s discussion of the effect that new topics have had on the form. In this review of visual images in children’s literature, Kiefer does a thorough job of providing work that exemplifies the points she is making. This text offers a variety of other areas and types of objects that could be explored using Tandem.

Tandem’s audience should be at least acquainted with the value of understanding image text relationships. The image text relationship is key to building understanding for the readers of these types of objects. Bettina Kummerling-Meibauer argues pictorial and textual elements of ironic narratives, more specifically picture books, are different than that of regular literary examples. Kummerling-Meibauer coins the ability to detect and understand nonliteral language as “metalinguistic awareness” (Kummerling-Meibauer, 1999). Kummerling-Meibauer states that it is an accepted understanding that young children have the ability to foster this type of metalinguistic awareness when reading picture books (1999). According to this argument, there are four key patterns associated with this phenomenon: “semantic gap, contrast in artistic style,
change in point of view, and sequential structure” (1999). In “The Dynamics of Picture book Communication”, authors Maria Nikolajeva and Carole Scott discuss the groundwork for the larger book *How Picture Books Work* in this article. The interplay and interaction between text and images is presented and distilled down to a relationship between words and images that is ambiguous and subjective (Nikolajeva & Scott, 2000). It is acknowledged that in an effort to categorize these kinds of relationships, some are more easily identifiable than others (Nikolajeva & Scott, 2000). The relationship is often complex and reliant on a range of factors that are different for each text. The distinction between illustrated books and picture books is also essential for these researchers. Picture books imply narrative work being accomplished by the illustrated elements of the text (Nikolajeva & Scott, 2000). Illustrated books use illustrations as decoration to complement rather than work alongside the text.

In 1996, Zhihui Fang’s “Illustrations, Text, and the Child Reader: What are the pictures in Children’s Storybooks for?” discusses the relationship between the illustration and the reader. The explicit purpose of the paper is to “delineate the main functions of illustrations in relation to the text in picture books and to examine the significance of illustrations to the child reader” (Fang, 1996). Fang argues that art works as a tool for storytelling as opposed to a decoration alone (Fang, 1996). The work being done includes establishing the setting of the story, definition and development of the characters, and further as an arm to extend and develop the narrative story (Fang, 1996). The art alters and augment the perspective of the reader while adding understanding and reaffirming elements of the narrative (Fang, 1996). After positing the previous, the author goes into a psychological overview of the effect of illustrations for child readers (Fang, 1996). Illustrations are concluded to be an essential narrative device for
developing readers (Fang, 1996). The interplay of text and illustration is also concluded to be supplementary for young readers (Fang, 1996).

Finally, there is a language to image and text based literary experiences. Stretching further back than the coursework at the CUNY Graduate Center, the vocabulary of multimodality was introduced in the final stages of my English degree at CUNY Queens College. Professor Jason Tougaw introduced Steven McCloud’s 1994 comic book and canonical guide to reading comic books, and more generally image and text based media, titled Understanding Comics: The Invisible Art. In depth descriptions regarding visual iconography, the definitions associated with the genre, as well as the way that text and illustration are interdependent and related is the main focus of the text (McCloud, 1994). The book acts as a reference point for reading both comics and other illustrated forms of literature. The dynamics between psychology and color are given fair attention, and McCloud’s text provides an overall analysis of the medium of comic books. By building a reference guide to illustrated literature in the form of illustrated literature, McCloud achieves a referential work that defines and explains the importance of the multimodal narrative. Tandem seeks to offer new translations to that existing language. These narratives can offer very different meanings when read from a distance. Tandem owes its inception to the discussion around McCloud’s work and its value for reading multimodality.

III. Project Evaluation & Plans for Continuation

This project achieves a significant amount of what it set out to do in that a prototype for a large subset of the functionality was developed, tested, and successfully deployed to the web. This prototype proves that JavaScript is a viable programming language for this type of project. With additional expertise and time, a more robust tool is possible. All things considered, Python may still remain the stronger language choice at this time for the data generation, feature
extraction, and corpus manipulation. What was easily achieved in Python proved to be more complex when translated into JavaScript. Strictly speaking in lines of code written, the original application and packages requires less development. Although a simpler project may have been a direct translation of the Python scripts from the original iteration into JavaScript, I decided to avoid this as is may have taken away from JavaScript best practices. What worked in Python might not have directly correlated to what works in JavaScript. A reasonable area to explore next is a tool that marries both languages in a single web application. Much like what Tandem does with NLP and CV, this future work could mash up Python for the heavy lifting data generative behavior and JavaScript for the functional user interface and file handling components.

It is also important to briefly acknowledge that the successes of this project are due in part to the open source community and actively involved community of developers at large that discuss their own issues and share their resolutions on forums like StackOverflow. These discussion forums prove to be a valuable companion for this process. When a considerable setback is experienced these communities have in most cases seen a similar problem and come together to resolve it. This active conversation space lifts the majority of the blockers during each stage of development.

From a timing perspective, a single semester did prove to be an insufficient timeline for full delivery of what was set forth. It would be incorrect to assert that a developer with a stronger skillset and the sole focus of this project could not have done this in a semester, but due to various other responsibilities, this project overall will take longer to complete. In order to finalize the features outlined in the user stories document, some additional development both in code and in skillset is required. One of the most important features that fell short is user specific file handling. While user login and authentication was an easy problem to resolve, the creation of
user specific directories that are triggered and tracked as each user uploads their own corpus is a key area of functionality is not fully functional. It remains in progress at the time of deposition. Outside of file handling, more robust natural language processing data extraction and output are targeted for later versions of this application.

This project is one that will benefit from continuous development practices. Open sourcing this project allows others who find it useful to not only benefit from its current state, but to also build on it. Improvements will be made in all areas of the application with additional insight and time. As users come to the application and generate their data, new user stories and user flows will be discovered. New requests will be made for future features. I intend on maintaining and developing new features as those requests emerge. Whether supported financially or not, the general upkeep of the application is projected to be minimal and will be determined based on adoption and community interest. The desire to continue forward with this type of data generation also has to do with personal academic goals. More people at varying levels of academic expertise, from high school to undergraduate, to graduate and post graduate, are beginning to not only tinker with, but are actively applying computational methods in their work. Taking Tandem to a broader interactive audience is the next step for this project. Social media outlets like Twitter and academic forums like the CUNY Commons are useful tools for disseminating the project and introducing it to potential users. Empowering those with similar interests and building community around similar lines of research is a component of digital humanities that makes it particularly valuable and drives me to improve the product.
Appendix A.

Supplementary Code Repository & Live Link

Description:
This code repository contains the necessary source files to replicate the application described in this paper. A full version-tracked history of this applications development can be found at the below. In addition, the application is currently hosted on Galaxy, a service provided specifically to host and manage Meteor applications.

Github Repository URL: https://github.com/CVDH4/Tandem2

Live Production URL: http://tandem.meteorapp.com/
Appendix B.

TANDEM 2.0

The MIT License (MIT)

Copyright (c) 2017 Christopher Vitale

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
Appendix C.

**Tandem 2.0 User Stories**

The below are user stories used to define the functional activities users of Tandem should expect to be able to execute in the application.

<table>
<thead>
<tr>
<th>Component</th>
<th>Priority</th>
<th>User Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCR</td>
<td>High</td>
<td>I as a user select a file/s for Optical Character Recognition.</td>
</tr>
<tr>
<td>OCR</td>
<td>Low</td>
<td>I as a user, after Optical Character Recognition occurs, view the raw text output of my file/s.</td>
</tr>
<tr>
<td>NLP</td>
<td>High</td>
<td>I as a user bulk process my OCR text output.</td>
</tr>
<tr>
<td>NLP</td>
<td>High</td>
<td>I as a user run basic NLP to generate statistics on my text output including Tokenization, Word Count, Unique Word Count, etc.</td>
</tr>
<tr>
<td>NLP</td>
<td>Medium</td>
<td>I as a user run predefined NLP topic modeling on my text output.</td>
</tr>
<tr>
<td>NLP</td>
<td>Medium</td>
<td>I as a user run predefined NLP sentiment analysis on my text output.</td>
</tr>
<tr>
<td>NLP</td>
<td>High</td>
<td>I as a user, after NLP occurs, view the raw text output of my file/s.</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>High</td>
<td>I as a user run basic image processing to generate a dataset of image based statistics including image size, number of pixels, mode RGB values, etc.</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>Medium</td>
<td>I as a user run basic image processing to find the most common colors within the file/s.</td>
</tr>
<tr>
<td>Landing Page</td>
<td>High</td>
<td>I as a user navigate to dhtandem.com to find general information about Tandem 2.0.</td>
</tr>
<tr>
<td>Dashboard View</td>
<td>Low</td>
<td>I as a user have access to key navigation points like an upload view, data view, and information views via a left rail navigation</td>
</tr>
<tr>
<td>Dashboard View</td>
<td>High</td>
<td>I as a user download my raw data output as a .csv.</td>
</tr>
<tr>
<td>Dashboard View</td>
<td>High</td>
<td>I as a user have access to predefined data visualizations for my NLP data.</td>
</tr>
<tr>
<td>Dashboard View</td>
<td>High</td>
<td>I as a user have access to predefined data visualizations for my Image feature data.</td>
</tr>
<tr>
<td>Registration/Login</td>
<td>High</td>
<td>I as a user fill out a registration form including my full name, email, username, password, and affiliated organization to create</td>
</tr>
<tr>
<td>Function</td>
<td>Priority</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Registration/Login</td>
<td>High</td>
<td>I as a user can login using my Github.</td>
</tr>
<tr>
<td>Registration/Login</td>
<td>High</td>
<td>I as a user can login using my Twitter account.</td>
</tr>
<tr>
<td>Registration/Login</td>
<td>High</td>
<td>I as a user can login using my Google Account</td>
</tr>
<tr>
<td>Upload</td>
<td>High</td>
<td>I as a user upload one or multiple images.</td>
</tr>
<tr>
<td>Upload</td>
<td>High</td>
<td>I as a user see the progress, success, or failure of my image upload.</td>
</tr>
</tbody>
</table>

a new account for the application.
Bibliography


Kaplan, Deborah. “Read All Over: Postmodern Resolution in Macaulay’s Black and


Schons, Lisa M. “‘Is the Picture Book Dead? The Rise of the iPad as a Turning Point in


Spratt, Emily L. "The Digital Humanities Unveiled: Perceptions Held by Art Historians and Computer Scientists about Computer Vision Technology" (Self Published).


Wu, Shuxuan. “A Multimodal Analysis of Image-text Relations in Picture Books.”
Theory and Practice in Language Studies 4.7 (Jul. 2014): 1415-1420