The Cybernetics Thought Collective: A History of Science and Technology Portal Project

White Paper

National Endowment for the Humanities
Division of Preservation and Access, Humanities Collections and Reference Resources

Grant No.: NEH PW-253912-17
Project Directors: Bethany G. Anderson and Christopher J. Prom

White paper authors: Bethany G. Anderson, Christopher J. Prom, James A. Hutchinson, Anirudh Chandrashekhar, Brinna Michael, Shreya Udhani, Mark Sammons, Alex Dolski, Kevin Hamilton, Saumye Kaushik, Meghna Shrivastava

Contributors: Charles Greifenstein, American Philosophical Society; Jonathan Pledge, British Library; Thomas Rosko, MIT Institute Archives & Special Collections; and Beverly Turner, MIT Libraries

University of Illinois at Urbana-Champaign
April 24, 2019

The Cybernetics Thought Collective: A History of Science and Technology Portal Project White Paper

Introduction and Summary

The application of computational tools and methods to archival materials has gained traction and prominence over the last few years. The archival community has increasingly sought ways of using computational tools on digitized and born-digital records to extract metadata and find connections between documents and patterns within and among texts. Computational tools have proven utility for both users of archival materials and archivists seeking to find new ways of enhancing access to materials.

Computational tools can be useful for collaborative projects that involve related materials from different institutions, especially for facilitating textual analysis of correspondence networks. The significance of these tools is twofold: 1) they enable us to better understand the ways in which archival materials—particularly correspondence—relate to each other, especially in terms of metadata that reveals shared correspondents, concepts, places, and other named entities; and 2) they afford users different ways to engage with archival materials and understand the nature of intellectual correspondence networks. *The Cybernetics Thought Collective: A History of Science and Technology Portal Project* is a collaborative project that seeks to apply computational science to related archival materials that shed light on the history of cybernetics—a bold and foundational twentieth-century scientific movement. The participating institutions in this project—the University of Illinois Archives, British Library, American Philosophical Society, and MIT Institute Archives and Special Collections— sought to not only apply computational methods to the personal archives of four founding members of the cybernetics movement, but to also develop a digital resource that creates access to digitized materials and machine-extracted data about those materials. Fundamentally, we sought to explore the ways in which we could use these approaches to reveal the cybernetics "thought collective"—the scientific community of individuals exchanging thoughts and ideas—as reflected in the archival records created by the individuals who compose that network.[1]

The institutions participating in this project hold the personal archives of four founding members of the cybernetics movement—Heinz von Foerster and the records of his Biological Computer Laboratory (University of Illinois Archives); W. Ross Ashby (British Library); Warren S.

[1] For a discussion of "thought collective," see Ludwik Fleck, *Genesis and Development of a Scientific Fact*, edited by Thaddeus J. Trenn and Robert K. Merton. Translated by Fred Bradley. Repr. 11. Aufl. Sociology of Science (Chicago: Univ. of Chicago Press, 2008).

McCulloch (American Philosophical Society); and Norbert Wiener (MIT). In an effort to explore ways of uniting the people and concepts that constituted cybernetics, our institutions sought a Foundations project to build a collaborative partnership to digitize and provide basic access to a select portion of these records and personal archives, and to assess the potential of advanced machine learning methods to enhance their access and use. Specific work undertaken included: (1) selective digitization of archival materials that exposes social research networks of communication, thought, and idea exchange; (2) creation and remediation of metadata about these materials; (3) preservation and basic access through established systems; and (4) initial testing and assessment of named entity recognition, natural language processing, and machine learning tools in a prototype "thought collective" platform. The team's main goal was to enable users to explore the cybernetics movement in ways that they would not be able to with traditional systems, via interfaces that use machine learning techniques to expose latent relationships between people, topics, and locations. We also sought to create an "analysis engine" pipeline for normalizing the documents after they had been processed with OCR (optical character recognition) software, extracting named entities, and classifying related documents. Ultimately, we believe these approaches have the potential to enhance access not only to cybernetics materials, but to any large corpus of unstructured textual documents, but that additional development and testing must be completed before the full potential of such affordances can be fully realized.

This white paper discusses the Cybernetics Thought Collective (CTC) team's specific work to digitize a select portion of archival materials; investigate and experiment with natural language processing, named entity extraction, and machine learning software; begin investigating access interfaces for the portal; and ingest the digitized materials and machine-extracted metadata into the University of Illinois Library's preservation repository and Digital Library. The pilot grant project enabled us to explore emerging methods for creating access to archival materials, which resulted in promising outcomes. However, we encountered several unexpected challenges which made it difficult to complete the project within the one-year timeframe. These challenges revolved around completing the digitization of the materials, finalizing access to test content resulting from the annotation, entity extraction, and network analysis tools, and creating access to that content in the prototype portal. We thus requested a one-year no-cost extension to address some of these pending tasks.

Overall, the project has been extremely successful, and in May 2018, the CTC team launched the prototype portal: https://archives.library.illinois.edu/thought-collective/. We also completed digitization, testing software, and investigating access interfaces; ingested the digitized materials into the University of Illinois Library's preservation system; and uploaded all code for the "analysis engine" to our GitHub repository. The prototype portal provides access to

628 digital objects, which are accessible through the aforementioned "thought collective" portal. The portal links to University of Illinois Library's digital collections platform: https://digital.library.illinois.edu/collections/38ec6eb0-18c3-0135-242c-0050569601ca-1. In addition to access to the digital objects, users have access to machine-extracted metadata generated by the named entity recognition, natural language processing, and machine learning software. This metadata also is available for bulk download in addition to be accessible through the metadata profile for the digital objects.

Based on feedback from our advisory board members and usability assessment participants, the prototype portal will require additional development to make it more user-friendly, easier to navigate, and allow users to refine search results for entity maps and machine-extracted metadata that form the basis of the data visualizations. Accordingly, after the main work of the project was completed, we experimented with some additional metadata enhancement and visualization tools, to complement those tested and implemented during the project.

While the project has a number of successes and promising outcomes, to make *The Cybernetics Thought Collective* portal a more robust resource, we expect to undertake future work to address the following areas:

- While the vast majority of the materials are typewritten, handwritten correspondence can be found throughout the personal archives of the four cyberneticians; these handwritten materials will need to be transcribed in order to be processed for the analysis engine.
- The results and the utility of the natural language processing and machine learning analysis will also need to be analyzed for "quality control."
- We will need to add more personal archives of cyberneticians and scholars involved in the cybernetics movement.
- Finally, additional work would need to be completed to provide direct access to documents from some of the visualizations, particularly the node and network diagrams linking people and topics.

Project Background

Between 1946 and 1953, the Josiah Macy, Jr. Foundation hosted ten conferences—all but one in New York City—that brought together a diverse group of scholars representing fields from mathematics to the physical, life, social, and information sciences. Known as the Macy Conferences on "Circular Causal and Feedback Mechanisms in Biological and Social Systems," the postwar meetings initially sought to bridge disciplinary divides by using applied science to synthesize ideas and knowledge around questions about behavior and information-feedback, for both organisms and machines. After MIT mathematician Norbert Wiener published

Cybernetics: Or the Control and Communication in Animal and the Machine in 1948, conference participants adopted the term “cybernetics” as the umbrella under which their interconnected and interdisciplinary web of ideas gained meaning.

Cybernetics has been defined in many ways: as the science of communication and control, or “steersmanship,” in organisms and machines (Wiener, 1948); as the study of form and pattern (Bateson, 1972); or, as quite simply the study of behavior—“[cybernetics] treats, not things but ways of behaving. It does not ask ‘what is this thing?’ but ‘what does it do?’” (Ashby, 1956). Cybernetics thus provided participants with a common language to articulate and discuss similar questions about behavior, or ways of behaving, across disciplines, regardless of whether the subject of study was animal, machine, or social phenomena. This project aims to enable scholars to understand how those questions evolved over time and ask new questions themselves.

Project Overview

The project team digitized a select portion of archival materials between May 2017 and April 2018; investigated and experimented with natural language processing, named entity extraction, and machine learning software; began investigating access interfaces for the portal; ingested the digitized materials into the University of Illinois Library’s preservation system; and uploaded all code created for the project to a GitHub repository. The digitized content needed to be normalized, processed, and extracted before the work with the software could begin. In addition, the CTC team needed to define what entities (mainly cybernetic subjects and people) the software would need to look for and extract, as well as classify a set of digitized documents in order to create a training set for the machine learning algorithm. The team also began exploring interfaces for creating access to the data (this work is described below.) The team also created a website for the project that provides information about the project, which is also serving as the prototype portal (https://archives.library.illinois.edu/thought-collective/).

The CTC team at the University of Illinois primarily communicated via email and through regularly-scheduled one-hour meetings every Friday. We also scheduled several conference call meetings with project partners from the American Philosophical Society, the British Library, and MIT (May 17, 2017; June 8, 2017; August 16, 2017; October 16, 2017; and April 20, 2018). We scheduled three conference calls with the project’s advisory board (August 16, 2017, November 27, 2017, and January 18, 2018) to provide them with an update on the project’s progress and gather advice and feedback. The project team scheduled an onsite meeting at the University of Illinois on January 17-18, 2018, which included the main project partners from the American Philosophical Society, the British Library, and MIT.

The development of *The Cybernetics Thought Collective* prototype portal focused on the main activities: 1) Digitization and Transcription; 2) Text Normalization and Extraction; 3) Software Testing and Implementation; and 4) Portal Development and Access.

Digitization and Transcription

Digitization

The four participating institutions digitized a select portion of content from the papers of Heinz von Foerster, W. Ross Ashby, Norbert Wiener, and Warren McCulloch between May 2017 and April 2018. While digitization of the majority of the materials was to be completed by November 2017, various factors slowed down this work at several of the participating institutions. This subsequently meant that timeframe in which the transcription of the W. Ross Ashby journals and the overall processing of the digital content by the project programmers had to be shifted by a few months.

In the end, the institutions produced 628 digital objects (a total of 61,067 files, approximately 3.6 terabytes total in size), which includes access PDFs and preservation master TIFF files. All access PDFs were OCRed to facilitate text extraction and analysis. The W. Ross Ashby journals, as well as some correspondence throughout the four personal archives, are handwritten; thus, select materials were transcribed to prepare as many items as possible for computational analysis. However, the vast majority of the handwritten materials will need to be transcribed to make the materials ready for computational analysis during the next phase of the project.

Text Normalization and Extraction

Defining Cybernetic Entities

An important part of preparing the digitized texts for computational analysis entailed defining the vocabulary to “train” the software to use as a reference to process the texts, extract cybernetic terms, and classify the texts into distinct categories. Before testing software, the team needed to define the conceptual entities of the cybernetics corpus—what are “cybernetic concepts” and how will we recognize them? Entities such as people and locations are fairly straightforward for named entity recognition and natural language processing software to recognize and parse. The project aimed to map the exchange of cybernetic ideas between the documents and the agents in the documents, and thus understand the genesis and evolution of these ideas over time, as well as developing the procedures and blueprint for generating these entity maps.

In order to create a list of terms that the natural language processing and named entity extraction tools could use as a reference, the team used the anthology *Cybernetics of Cybernetics: Or, the Control of Control and the Communication of Communication* (1974) compiled by Heinz von Foerster et al., as an initial source for cybernetic concepts. The team decided to use *Cybernetics of Cybernetics* since it contains discussions of fundamental ideas in cybernetics and particularly the work of Ashby, McCulloch, von Foerster, Wiener, and their contemporaries.

The team analyzed *Cybernetics of Cybernetics* in Voyant Tools and generated an initial list based on word frequencies.[2] The text yielded a list of terms that could be ordered by broad philosophical and scientific categories (e.g., autonomy, memory, system); more specifically cybernetic categories (e.g., autopoiesis, eigenvalue, feedback); technical, sensorimotor, and concrete categories (e.g., brain, machine, nerve); and proper names (e.g., Biological Computer Laboratory, McCulloch, and Turing). We shared the list with the Advisory Board members for review before finalizing. As we later realized in the project, the list turned out to be a good "guess" and many of the terms from the list were extracted by the software.

Text Normalization

The team hired two student programmers in the summer of 2017 to begin testing software on a small sample of digitized content prepared by the four institutions. After we defined the entities for the software to recognize and extract, our first task was to create a Python pipeline to process the OCRed PDFs into a format that we could use to work with strings or lists of words. We used a Python library called PDFMiner,[3] which extracts data from documents into a TXT format. We then wrote the plaintext to a file in order to perform various statistical analyses later in the workflow. Our first iteration of the Python script ran our program on a set of PDFs and then converted the PDFs to text files with equivalent filenames. This script, however, produced a large number of unreadable files, due to inconsistencies and errors from the OCR.

Extracting data from the digitized content was especially challenging due to differing levels of accuracy of different OCR software—some produced more errors with the types of handwritten/typeset characters that are common throughout the digitized corpus. Many of the OCRed documents resulted in long sections of garbled data or data that consisted primarily of non-ASCII symbols and non-English characters. Other documents' formatting created spaces in

[2] For the *Cybernetics of Cybernetics* workbook in Voyant Tools, see https://voyant-tools.org/?corpus=ac8dba063f032caa44260e32c1e71ba6.
[3] PDFMiner, https://github.com/jaepil/pdfminer3k.

between the letters of a word; for example, "c y b e r n e t i c s i s a s c i e n c e." Additionally, many documents are in languages other than English. If we were to be able to construct a machine-learning tool with our data, we had to normalize the OCR results and try to ensure a certain amount of uniformity in our data, so that we could reliably extract words and phrases related to cybernetic thought.

In machine learning training sets, as in all of data science, the algorithm is only as good as the data set. We thus adopted the following approach to normalize the texts:

- First, we had to address the issue of garbled data/non-ASCII symbols. The simplest way to normalize the OCR results was to assign two separate scores to each extracted PDF file: a word length threshold score and a symbol threshold score. The former tested how long the average word was in a document and marked extreme outliers (words with a greater length than 14 characters, for example, are extreme outliers in most languages). The latter scoring algorithm found the ratio between letters and non-letters and non-ASCII characters (as some of the garbled files often had a mixture of different characters, with the symbols outnumbering the actual letters). The first check tested and removed documents where the extractor detected few spaces (for examples, "cyberneticsisascience"). Or the converse, this data would not be useful, and in fact probably add statistical noise/confusion to the final results. The average English word is about 4.5 letters long, with most other European languages around the same number, so we could perform a check across a document to see how usable it might be. Please note that there *are* reasons to keep symbols in a file; for example, many of the cyberneticians included mathematical equations in their correspondence and journals. However, when a document is all or mostly symbols, we can be relatively sure that this is an extraction error.

- Next, we had to convert files with excessive spaces (like the aforementioned "c y b e r n e t i c s i s a s c i e n c e"), with an average word length of about 1.0 that *could* potentially be made into usable data.

We also briefly investigated the idea of using English dictionaries to detect and resolve these issues using a default Natural Language Toolkit (NLTK) library, but it should be noted that this task is extremely time intensive, with a separate disk-read required for every search on every word in every document. A computer has no idea (without time-intensively searching every single letter and every combination of length) of where and how to break up longer words. Word segmentation and (fast and reliable) text extraction, while a very interesting problem to solve, proved to be far beyond the scope of this project; the above step, instead, aimed to extract as much data as easily as possible. The next section discusses the approaches to another issue in text extraction—translation.

N-Grams and Language Identification

Ascertaining the language of a document is a common task in computational linguistics, and is the first step to translating the documents. By determining the statistical probability of a certain word in a corpus given its context, as well as the context of the entire document, one can create a reasonably useful language identification algorithm. One of the project's programmers, Anirudh Chandrashekhar, created a test set of approximately 200 documents in English, German, French, and Italian, and then used and compared different N-gram approaches to language identification.[4] The accuracy for each of these models vary considerably in terms of accuracy. However, some models are more accurate than other models, described in more detail in the table below:

N-gram Model	Accuracy	Approach
Unsmoothed Letter Bigrams	48.66%	This model uses the letters found at the n and n+1 word of every line (with n=0 being the first word of the document, and n+1 being the second), and finds the probability of that sequence occurring in each individual language given the training. This model is not very accurate, because it has no way to handle bigrams outside the vocabulary, which is very small compared to the entire corpus of English letter bigrams. For example, the bigram *tz*, occurs in many English words of German, Yiddish, or Greek origin, but is not represented in the sample bigram: *The man I was talking to, was now so beautifully waltzing across the floor* The moment we hit an unknown bigram, *tz*, we multiply our probability by zero because the bigram has not occurred before. By disregarding all other probabilities in the sentence, we might erroneously conclude that a zero percent chance exists that the sentence is English. This leads us to assign another

[4] For a description of N-gram, see: https://en.wikipedia.org/wiki/N-gram.

		language, even when these languages have a statistically small chance of being correct because max (0, 0.0000000000000001) would still be the latter. However, the next model we employed addressed this issue.
Smoothed Letter Bigrams	99.66%	This model gives us significantly better results, because we are using add once, or “Laplace Smoothing,” which adds one to every count (and compensates in the denominator) when finding probability. This means that even if we have an unseen combination, this will likely cause fewer issues identifying language, because we can get a small (albeit non-zero) probability of it occurring in the language, and then continue multiplying it across the sentence to get the maximum number probabilities from the same training set. Its 99.66% accuracy sufficed to convince us that this smoothed model vastly outperforms the unsmoothed bigrams model. However, the model needed to be compared to the smoothed word bigram model.
Smoothed Word Bigrams	96.66% Accuracy	Smoothed models are absolutely essential when making word bigram models, even more so than letter bigrams. This is because the odds of having an unseen word in the language is so much higher, especially because we draw from a limited corpus. The reason our accuracy is slightly lower here than it is for the letter bigrams (96.66% vs. 99.66%) is due to the fact that our letter bigrams learn what sequences of letters (phonologically) are most common for a language, and so the algorithm does not necessarily need to have encountered the word before in order to make a judgment. We used the following sentence as a test: *Papa a construit le moulin.* (“(my) father has constructed a windmill”)

		Running this sentence through the corpus, the only words we find in the corpus are *papa* and *a*, which exist in all the cyberneticians' personal archives, and *le*, which exists in French and Italian. Now we have an issue—we have never seen *construit* or *moulin*, and so we assign it the smoothed score of 1/(size of corpus). But now we have the most likely estimate being that this sentence is in Italian (which is incorrect, it's in French). With the letter bigram, we can identify an unknown word, as long as its letter bigrams are present in the bigrams. When we run the smoothed letter bigram model, the sentence is correctly identified in French, not Italian. Therefore, the smoothed word bigram appeared to be the more accurate model, though not quite as accurate as the smoothed letter bigram model.
Good-Turing Word Bigram	38.66%	Using the Good-Turing equation, Chandrashekhar applied the equation for a smoothed count to be: c*= (c+1) * (Nc+1)/ (Nc). Where, if a word was unseen, we simply assigned the following probability: (N1)/N. And then applied it to the probability as before. "Nc" is derived from the content through our dictionary-of-dictionary, unrolling it to get every value (count located inside the dictionary of dictionaries), and finding out how many bigrams have this count. And "Nc+1" is found by taking the count, adding one to it, and seeing how many bigrams then have this count. However, the accuracy percentage for this method is fairly low, and lower than an unsmoothed letter bigram approach. Chandrashekar was not quite not sure why this model's output was less accurate than other models, but this may be due to issues with the model itself. For example, in this model, if we have a (Nc+1), this turns out to be zero and the entire model

		is unsure how to handle it. We see these zeros in a significant (approximately 40%) amount of the sentences, meaning that in these cases the count becomes zero, and we have the same issue as the unsmoothed corpus. After this analysis, we decided that our best option is the *smoothed letter bigrams*, though a Turing model, with some smoothing, might produce better results. This may be worth exploring in a future phase of the project.

Once we identified many of the main the languages in the documents, we then used Google's automated translation pipeline to translate the documents,[5] passing the normalized texts as well as the identified languages from our N-gram model through the translation tool, and then wrote the output to TXT files. This pipelining tool has recently been blocked by Google's translation service, but the mechanism can essentially be run by any translation-server tool that Python can access.

Software Testing and Implementation

Natural Language Processing

After all the texts had been normalized and "cleaned," the programmers were ready to begin testing the University of Illinois Cognitive Computation Group's Natural Language Processing (NLP) Pipeline software in order to extract feature data (entities) about the data points;[6] the idea being to extract feature data that contain entities of words in common with those on the prepared list of cybernetic concepts, or in which cases those concepts from the list appear the most frequently. The NLP Pipeline is a package that bundles various software components required to run specific NLP applications and tools. We then would use machine learning to look at the category of the entry and then look at its features and be able to predict with which categories data from the corpus will be associated. The results of this testing yielded a great deal of data on word frequency and counts, but we realized that much of the data still needed to be normalized and refined to eliminate prepositions and pronouns, and other words that are not significant to, or as, cybernetic concepts or ideas.

[5] Googletrans, https://pypi.org/project/googletrans/.

[6] Now managed by the University of Pennsylvania: https://cogcomp.org/page/software/.

In parallel, the programmers tested Wolfram Text Analysis tools (part of Wolfram Language),[7] including “WordCounts” and “TextWords” functions, in addition to others, to facilitate text recognition, NLP, and semantic analysis. Stephen Wolfram, founder and CEO of Wolfram Research, is a member of the Advisory Board, and generously donated technology resources toward the project. Wolfram Language is a *very* high-level programming language that allows one to automatically perform certain tasks, such as produce word statistics and generate graphics. It also has a built-in machine-learning toolkit in its *classify* function.

Another useful built-in Wolfram Language tool is the StopWords function, which helps reduce statistical noise for when we eventually transition to the machine learning part of the project. By removing insignificant words like “the”, “and”, “he”, etc., we could focus on the core information in any given document. Wolfram also allows us to find statistical features like 3-grams, and even allows us to produce feature-space-plots which plot three grams on a 2-D surface, with the closer three grams representing three grams with at least one shared feature (“neuron”, “impulse”, “measure”) or (“standard”, “neuron”, “behavior”).

Once we removed many of the stop-words and made an analysis of the most important terms using Wolfram’s built in N-Gram analysis, we could transition to the classification of our texts, which could be done using Wolfram’s highly intuitive classification function that allows us to tweak many parameters, and even allows us to create neural nets at the base level.

Unfortunately, due to some roadblocks with Wolfram Language and the NLP Pipeline software, we realized that we needed to test out a different programming language. Wolfram Language, as intuitive as it is, was unfortunately very slow with the text corpus that we were using, even on powerful servers and machines. Once we tried to load larger amounts of classification data in as text files, the Wolfram system repeatedly crashed and wasn’t able to handle or convert the files. Though the Wolfram notebook system was portable and easy to document, it did not enable us to convert PDFs into a TXT format, and any sort of conversion would have to involve a pipeline in multiple programming languages. We decided to use Python, which would enable us to perform many of the same tasks as we could with Wolfram Language. Python is a well-documented language, has a large number of libraries, and is much faster than the Wolfram Language. Therefore, we used Python for the rest of the project.

[7] Wolfram Text Analysis, http://reference.wolfram.com/language/guide/TextAnalysis.html.

TF-IDF Feature Recognition

When examining such a large body of texts—varying over the many topics that are germane to cybernetics, such as psychology, medicine, and mathematics—it is important to try to find features or entities that may be important to the corpus (and compare these results with the aforementioned list of cybernetic terms we created). The first approach we employed to identify which terms are most important to us used the Term-Frequency Inverse Document Frequency (TF-IDF) measure. The first part of this term-term frequency finds two scores of a given word, or term- firstly, how often the term is found in a particular document, and then how many times this word has occurred, at least once, in the entire corpus. We used TF-IDF to find the words that are most common in the corpus. Then we used the *inverse* document frequency—that is, words that occur frequently in some documents but not in all documents (in order to filter out the most common words that are stop-words, like “the” and “and”, etc.

One challenge that emerged during this analysis was that the document length varied so much across the corpus that the TF-IDF results were not as useful and yielded many features that were unimportant while not including cybernetic terms. In a future phase of the project, it may make sense to weight and normalize the document length and use this to find the more important features in the text. Instead, we used raw word counts and manually analyzed the top searches to find terms that, from our perspective, found to be important. While TF-IDF analysis can be an important way to isolate important words, and should be further analyzed, unless more document-length normalization is done at some point, we wouldn’t be able to reliably use this type of analysis.

Raw Counts and Manual Feature selection

Rather than using TF-IDF or a stop-word removal library, we decided on instead using a raw count of the most common words across the corpus and then manually going in and finding all of the cybernetic terms that corresponded with the most common words. Once we got a list of the most common words, we simply ignored pure stop-words, but also removed non-topical words like “dear” and “university,” which occur frequently in the correspondence, but have very little value in terms of cybernetics content. Before doing this, we already had a list of cybernetic terms to use as a reference, and when we compared this old list to our new list of most common features, we found that there were only two or three words that needed to be added to the original list. We were thus pleased to see that our original list of cybernetic terms was fairly accurate.

These features (both cybernetic concepts and people mentioned in the correspondence and journals) were extracted at the file-level and appear in the metadata profile for the ingested content.[8] An illustration of the overall workflow we developed can be seen below (Figure 1).

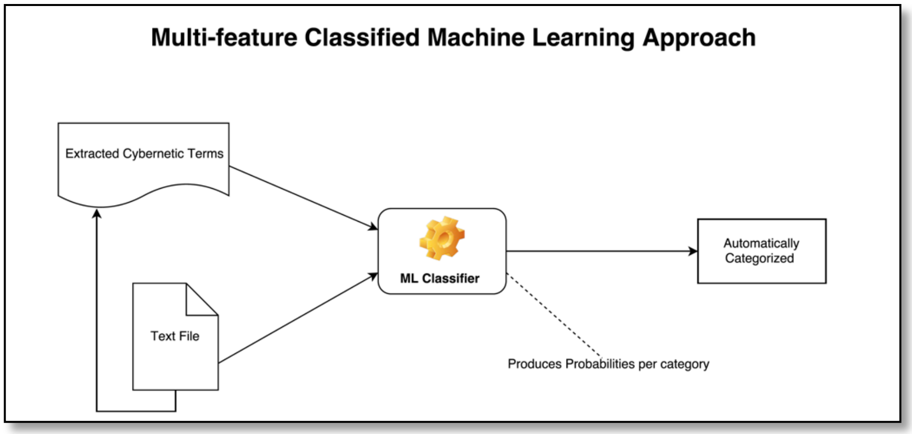

Figure 1: Illustration for the supervised machine learning pipeline we employed

Supervised vs. Unsupervised Machine Learning

Following text normalization and the extraction of cybernetic entities and prior to beginning software testing, the project team discussed next steps and whether to employ a supervised or an unsupervised approach to machine learning (or a combination of the two). Employing machine learning, a subset of artificial intelligence (AI), would help us find connections between documents by classifying related documents into distinct categories. There are two main approaches to machine learning—supervised and unsupervised.

A supervised learning approach entails preparing a training set for the machine-learning algorithm (either manually or with the help of tools via generated word statistics). We would create a list of general categories to sort documents into (e.g., personal correspondence, scientific correspondence); these categories train the system to classify (unseen) patterns. We can do this by creating *classes*, or categories, to look at the *features* of a certain document, and teach the machine to view hidden connections between the documents and features so that it can assign the documents to the right *class*.

An unsupervised machine learning approach entails using a "clustering" method. We wouldn't create categories for any of the documents into which they would be classified; rather, we

[8] The content for the project has been ingested into the University of Illinois Library's digital preservation service (Medusa), which has a front-facing access interface digital library. This metadata can be viewed in the "machine-generated feature" fields: https://digital.library.illinois.edu/collections/38ec6eb0-18c3-0135-242c-0050569601ca-1.

would use the features of each text (for example, the cybernetic entities), and ask the program to create an unnamed category, or a cluster, which in some way has an underlying similarity to all other entries in that cluster. We could use the features of each text (a combination of words used, for example) and ask the algorithm to create a cluster, which has an underlying similarity to other entities in that cluster. However, an unsupervised approach could be unpredictable.

The project team leaned toward a supervised approach—tagging a training set of the documents with concepts revealed from the word statistics tool, and then using that training set to classify the rest of the texts in the corpus. A supervised approach, on the other hand, could be more labor intensive since we would need to create the training set. The initial results of this supervised approach will be discussed below. However, we think it would be useful to test out an unsupervised approach during a future phase of the project.

Naïve Bayes and Machine Learning

After we had extracted features (cybernetic entities and people), we could use this data to extract only the most important parts of the text. Again, the reason for this had to do with practical considerations; if we wanted to have a usable classification set based on raw text of varying lengths per document, we would need at least 10,000 (at the very minimum) human-classed examples, for which we would not have the resources to use on the project. Instead, if we just use a bag-of-words-technique (i.e., we do not care about order, as in an SVM analysis, just frequency) on the most important cybernetic concepts, and cyberneticians using Naïve Bayes.[9] From this, we could narrow down the size of the training set. This has some advantages:

- Low latency runtime/resource requirements
- Very fast (comparatively) to set up and adjust; but one major disadvantage to consider: Assumes independence in every word (not looking at position of word near other words or phrases)

We decided on a list of four classes (Mathematics/Logic; Computers/Machines; Psychology/Neuroscience; and Personal Correspondence), which we determined were important enough to both be useful to users of the portal as well as unique and distinct from the other categories to prevent as much overlap as possible. Because this was our first attempt, we only attempted classification, for its ease in creating a training set as well as testing to see if it works properly. However, it could be posited that a *regression* model could be useful, with discrete values (from 0.0 to 1.0) to represent, for example, the nuances in cybernetics concepts

[9] Naïve Bayes, https://github.com/codebox/bayesian-classifier.

(and how, really, all of these topics are connected); however, we would need to explore this in future phase of the project.

Our classification model was at the very end of a pipeline that automatically converted PDFs, cleaned them, discarded ones that weren't viable, extracted features and language, and then found the predicted class and percent certainty for every class. Overall, we achieved a classification accuracy of 71.1%, which can be significantly improved upon if we had more than 30 files. In fact, if we achieve such accuracy with so few files, then we know that the features we chose were important to the classification system as a whole.

Weka Feature Analysis

It is important to note that the features we selected are somewhat sufficient indicators for our classifier to use. Though our accuracy is relatively high (and certainly much higher than can be attributed to chance), we also have a small test set and a small training set. To help us better understand the significance of these features and their relationships to the document, we would need to see a full ranking of all of our features so that we could reduce noise.

To do this we used a GUI-based machine-learning toolkit called Weka,[10] from the University of Wakaito, New Zealand. Among other things, Weka allows us to perform chi-squared analysis (relevance of feature in classification) and forty different types of classification (provided the data is in a CSV and built in cross validation and numerical result generation). After converting our data into an acceptable CSV, we analyzed some of the results of our machine learning output, as below, before checking the feature list.

Results from Naïve Bayes:

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
Weighted average	0.711	0.04	0.863	0.711	0.71	0.934

[10] Weka, https://www.cs.waikato.ac.nz/ml/weka/.

Explanation for all the metrics:

- TP (True Positive) Rate (Recall): How many times the model correctly classified a model as being correct divided by how many it should have gotten correct. We see that, for the reason described above, this yielded a high “True Positive” result
- FP (False Positive) Rate: How many times our model incorrectly classed our data to be positive, when it should have been assigned negative. Because Naïve Bayes determines that all features have equal weight, it most likely assumed that certain features were more important than they actually were (and made assumptions of independence that a decision tree would not) and marked more false positives than other models that we may implement in the future.
- Precision: How many true positives we had, out of how many positives our classifier says are correct (including false positives). Because we have a higher TP and a lower FP rate in our decision tree, our Precision is naturally better.
- F-Measure: the harmonic mean of *both* precision and recall, which consider the True Positive rate, the False Positive rate, and the False Negative rate.
- ROC (Receiver Operating Characteristic): How good our model is at being able to classify (in a binary system) what is correct and what isn’t. In other words, can our model determine what differentiates a True Positive and a True Negative. The number, usually plotted as a curve, compares True Positives to True Negatives over a performance threshold. Both of these seem to be capable of this over different thresholds.

Based on these accuracies, we believe that we chose features that align with the data typical of the corpus; in other words, the cybernetic concepts and people mentioned are good hints as to what class into which a certain document is classified. Using chi-squared analysis, we can actually put a number to and rank how good our features are. We found that the names of the people, followed by *most* of the cybernetics terms, did actually play a significant role and have merit as far as classification is concerned. This enabled our programmer to determine which features to delete at the very bottom of the list that only added to statistical noise. All the code developed for the project has been made accessible in a GitHub repository.[11]

Portal Development and Access

The project team investigated ways of creating access to the machine-extracted data and machine learning results and the digitized materials. We decided on two different modes of

[11] See: https://github.com/cybernetics-thought-collective.

access: 1) access to the digital content through the University of Illinois Library's digital collections platform and through select, and 2) access to different visualization interfaces.

Metadata Profile and Digital Library Access

As noted above, a cybernetics specific metadata profile was created for this collection, using native functionality in our Digital Library application. While based on *Describing Archives: A Content Standard* (DACS), the profile includes additional elements that extend beyond DACS. The table below lists and describes each profile element. All metadata that specifically resulted from the machine learning or natural language processing software pipeline is noted as "machine generated":

Metadata element	**Description of element**
Level of Description	Level at which the materials are described, such as item-level, file-level, or collection-level
Title	Title of materials
Scope and Contents	Collection-level description
Creator	Creator(s) of materials
Subject	Subjects drawn from local subject headings
Location	Geographic locations noted in the correspondence
Format of Material	Genre, such as correspondence, journal, etc.
Language	Primary language(s) of materials
Cybernetic Classification (Machine Generated)	Classification into which the materials were placed based on highest percentage (i.e., Mathematics/Logic; Computers/Machines; Psychology/Neuroscience; and Personal Correspondence)
Classification Certainty (Machine Generated)	Percentages of certainty for all four categories of classification
Machine-extracted Feature (Machine Generated)	Cybernetic entities extracted from materials
Date (Machine Generated)	Calendar dates, including months, days, and years
Geo-Political Entity (Machine Generated)	States, countries, city-states, etc.

Location (Machine Generated)	Geographic locations noted in the correspondence, including mountain ranges, bodies of water, etc.
Facility (Machine Generated)	Entities such as buildings, highways, bridges, etc.
Nationalities, Religious, or Political Groups (Machine Generated)	Any nationalities or religious or political groups
Organization (Machine Generated)	Entities such as companies, agencies, institutions
Person (Machine Generated)	Persons mentioned in the materials (could be correspondents or subjects)
Product (Machine Generated)	Branded products
Overall Sentiment (Machine Generated)	Percentage by which sentiment could be identified in materials
Percent Positive Sentiment (Machine Generated)	Percentage of positive sentiment within the materials
Percent Negative Sentiment (Machine Generated)	Percentage of negative sentiment within the materials
Percent Neutral Sentiment (Machine Generated)	Percentage of neutral sentiment within the materials
Rights	Rights information for use of materials
Parent Collection	Collection from which the digital object derives
Repository	Repository that holds the materials
Collection Identifier	Identifier used for collection, such as record series number
Collection	Collection within the Digital Library that the materials are part of

Data Visualizations

Between December 2017 and July 2018, digital project assistants Shreya Udhani and Brinna Michael worked with the project team to investigate access interfaces for the data generated from the machine-extraction and classification work. We investigated several different types of visualization software, especially for network visualizations. Many of the network visualization tools we tested had difficulty displaying a large amount of the machine-extracted data. We thus decided to use a small sample of data for the visualizations, but we also provide users full access to the data for download.[12] Additionally, users can access the metadata/data via the digital collections' IIIF manifest.[13]

The main data points we decided to use in the visualizations are: Unique Identifier (filename), Title (folder title), Date, Sender (creator), Associated Person (person mentioned in correspondence), Machine-Extracted Features (cybernetic terms), Location, Cybernetics Classification (machine-learning classification), and Certainty (percentage of certainty that the algorithm thinks the file belongs in a specific class). One of the challenges we encountered is that many visualizations could not display well the sheer amount of data and related connections. Thus, we decided to use a small set of the data for testing purposes.

We experimented with the following visualization tools:

- WebVOWL: interactive software for visualizing ontologies. Due to the volume and structure of data, there were too many nodes making it difficult to read and understand[14]
- RAWGraphs: RAWGraphs is open source and provides users the option to create static visualizations using excel data. We had some amount of success with this tool, though the visualizations produced are static and not interactive[15]
- Plot.ly: Plot.ly has a chart studio which can be used to upload data and create interactive visualizations. We had some success with these visualizations and are integrating them into the prototype portal[16]
- Onodo: Onodo is an open-source network visualization and analysis tool

We decided to use three of these four visualization tools to provide access to different types of visualizations about the data. These visualization interfaces, along with information about the project, and a link to the ingested content in the University of Illinois' Digital Library system, are currently accessible in the prototype site.

---

[12] All metadata is available for bulk download via a CSV file in the Digital Library: https://digital.library.illinois.edu/items/6687d2a0-20eb-0138-7070-02d0d7bfd6e4-d.

[13] The IIIF manifest is available here via the Digital Library: https://digital.library.illinois.edu/collections/38ec6eb0-18c3-0135-242c-0050569601ca-1/iiif.

[14] WebVOWL, http://vowl.visualdataweb.org/webvowl.html.

[15] RAWGraphs, https://rawgraphs.io/.

[16] Plot.ly, https://plot.ly/.

Usability Assessment and Focus Group Testing

In order to assess the usability of the portal and the ways in which potential users would access and understand the data, we organized two focus group testing sessions in September 2018. Questions were asked via a webform followed by an in-person meeting for one of the focus groups. The focus group participants were asked questions about their web usage (such as favorite websites and why), and their initial impressions of the thought collective site, including what they can do with the site and what they think it is for. The participants were asked to perform a series of tasks, including gathering information about the site (i.e., what entity funded the project? What are the participating institutions? What is the purpose of the project?); locating digitized content (i.e., Please find the digitized Norbert Wiener Papers. How did you go about finding them?); locating specific materials within the digitized content about persons or subjects (e.g., Please try to find specific documents (correspondence, publications, etc.) authored BY or ABOUT Günther Gotthard. How many items do you find and how did you go about finding them?); and locating data about persons or subjects (e.g., Please try to find specific documents or visualizations that show who was talking about the cybernetic term "control." What do you notice about this term (who was talking about it, when, etc.)?).

The first focus group consisted of three colleagues from the University of Illinois Library who provided initial feedback on usability testing questions and on the portal. This initial testing enabled us to refine the questions slightly so that they could be better deployed to facilitate feedback. Overall, this focus group noted that the site was easy to navigate and the purposes for which the site could be used were understandable. However, some of the content requires more explanation to make information about cybernetics more accessible to all audiences. Furthermore, the focus group participants were able to access and perform the tasks asked of them by accessing the data in the Digital Library, but performing the task that required exploration of the data visualizations was not as clear. The focus group ultimately suggested that the CTC team consider creating a more integrated visualization experience for users of the portal.

The second focus group consisted of six subject experts in cybernetics or the history of science and technology. Like the first focus group, they were able to determine the aim of the site and the purposes for which it could be used. Also, like the first group, the second focus group was able to figure out how to access and explore digitized content and metadata in the Digital Library, but they had more difficulty in finding content and in seeing the utility of the visualizations. However, it was not clear to one participant that they had to double-click on folders in the Digital Library in order to see the files displayed. One participant noted that it was important to access the content through a traditional archival organization. It became apparent that more explanation was needed about how the visualizations were created and how the user could deduce meaningful content from them. Another participant voiced an interest in being able to access digitized archival materials related to specific persons or subjects in the visualizations and to be able to “influence” the visualizations by creating individual queries.

This feedback from the focus groups made it clear that the CTC team should have three priorities moving forward: to provide more explanation about the visualizations and how they can be used; to enable the users to have more control of and influence over the visualizations; and to better integrate the extracted metadata and digitized documents with the visualizations.

Additional Testing

After the work described above was completed, we expended a bit of additional time testing other potential approaches to classifying the body of materials that were digitized. Specifically, we applied alternate visualization, sentiment analysis, and named entity recognition tools to the corpus of digitized texts and to downloadable copies of the metadata found in the Digital Library.

The additional testing work demonstrated a “collections as data” approach that we implemented in the Digital Library. Graduate hourly employees Meghna Shrivastava and Saumye Kaushik downloaded the text and associated metadata files, then pursued visualization and NLP projects, as described below.

Visualizations

Meghna Shrivastava used Tableau Software to construct additional node-based diagrams for five sets of records namely: BCL Publications, Heinz Von Foerster Papers, Norbert Wiener Papers Visualization, W. Ross Ashby Papers and Warren S. McCulloch Papers. Constructing visualization for the mentioned publications using Tableau required the following steps:

- Importing the excel metadata files
- Data Cleaning on the text files, using Jupyter Notebook and python language
- Loading those cleaned excel files over tableau using Tableau ‘Open Data Source Option’
- Manually placing the attributes properly to construct the visualization

Tableau Workbooks are linked from the project site, and these workbooks implement two basic features:

1) Tooltip views: To make the visualizations interactive and readable for the non-technical persons, Shrivastava implemented the Tableau Tooltip features, to provide a

description about each entity of the column. The tooltip provides details about Certainty, Creator, Cybernetic Classification related to each document, Date, Subject, Title, Number of Records associated with each document in the Metadata.
2) Linking to Documents via 1URL: For each publication, we created a path to view the documents on the website extending Tableau's 'URL ToolTip Option'. The Tooltip URL option helps us add a working link over the visualizations. Given the amount of time available, at this stage we link only to folders of documents. Additional work would need to be completed to link each node to the individual documents it represents.

Natural Language Processing
Based on the work with the NLP pipeline discussed above, we experimented with methods to provide machine generated metadata for other text-based collections (i.e. not ones associated with Cybernetics. After identifying three text-based collections stored in the Library's Medusa Repository, Saumye Kaushik developed a method for adding machine created metadata to any collection that uses the "Archives General" metadata profile, a schema that supports descriptions that comply with the DACS rules. Specifically, he undertook the following work:

- Experimenting with different NLP tools to find the best ones for the types of data we are using
- Developed a python script that
    - Provides a simple user interface
    - Runs the NLP Tools against a defined folder and metadata spreadsheet downloaded from our Digital Library
    - Performs a simple cleaning process on the data to remove static and other gibberish content
    - Saves the metadata to a new spreadsheet. The columns of this sheet match the column order for a modified version of the "Archives General" metadata profile, adding in the supplementary machine generated data.

In additional work that stretches beyond the scope of the grant, we will have a student do further manual cleaning on the spreadsheets, then ingest them into the digital repository.

Future Work and Recommendations

The ultimate goal of this testing is to develop interfaces and a portal that allow researchers to identify additional connections in the cybernetics thought collective (the "research network") and testing the software from the Cognitive Computation Group and Wolfram will allow us to begin developing tools that can provide these features for other collections of unstructured text. Future work will also include assessing research on unsupervised learning methods that generate cybernetic ontologies from unstructured or semi-structured texts. Anirudh

Chandrashekhar also created an automated pipeline for the multi-feature machine learning classification system. We hope to explore ways of making this “analysis engine” accessible to users of the portal in a future phase as well.

This white paper describes the preliminary progress of the Cybernetics Thought Collective project. We envision this as a long-term project, eventually digitizing the personal archives of Ashby, McCulloch, von Foerster, and Wiener in entirety (as appropriate) and incorporating texts from other institutions holding the personal archives of individuals who influenced or were influenced by cybernetics. We hope to expand the portal into a comprehensive resource that makes it more feasible for researchers to understand and study this unique and pivotal moment in not just the history of science and technology, but intellectual history. Cybernetics influenced the development of a number of different disciplines, including as artificial intelligence, computing, and anthropology; enabling scholars to trace—at a large scale—through archival sources the exchange of ideas between those individuals for whom cybernetics served as their intellectual home, and the disciplines that the movement influenced. In addition, later phases of the project will more fully develop the software we tested and developed during this pilot phase and an access interface that enables users to search the entity maps in the corpus through different defined queries.

As an inter/transdiscipline, cybernetics continues to excite the imagination and invite new research questions about what the movement meant and its legacy. Making accessible archival materials that document the evolution of cybernetic ideas, collaborations, and networks of communication highlights not only the importance of archives for a subject like cybernetics, but the ways in which archival materials like correspondence help us understand the iterative and collaborative nature of many scientific disciplines.

Despite the uniqueness and significance of the personal archives of Ashby, McCulloch, von Foerster, and Wiener, they have been largely inaccessible to scholars who are unable to travel to the United State or Europe for research. Digitizing and making accessible these geographically dispersed archives in a centralized portal will enable scholars to access the archives of cybernetics in one space as well as reconstruct and explore different aspects of that network. Recreating the cybernetics thought collective through the material substrates of its scholarly activities—i.e., archives—and enabling users to explore connections between correspondents, ideas, and places latent in the materials, will enhance the research value of these personal archives and records. This pilot project is helping us begin revealing and enhancing access to the larger context of the cybernetics phenomenon and the actors at its center.