Pamella R. Lach. Thinking Historically about Data: Improving Automation Processes for Harvesting North Carolina City Directories. A Master's Paper for the M.S. in I.S degree. July, 2012. 74 pages. Advisor: Richard Marciano

Scaling up to handle big data can be challenging for the Digital Humanities given the often diverse and unpredictable nature of such data. This project proposes a system for automatically harvesting North Carolina city directories by developing a historically-minded parser. City directories vary significantly in format and structure, making the use of a one-size-fits-all approach impossible. Building a smarter parser requires that historical variances be taken into account from the outset. Such an understanding may come from an analysis of directory attributes, including the presence of a header or the connotation of a resident's racial classification. When taken together, these attributes reveal patterns across directory publishers, city locations, and publication years that form the basis of parameters for adjusting the parser to improve overall automatic data extraction. This project demonstrates how applying historical thinking to computational solutions contributes to more effective tools for handling big humanities data.

Headings:

Data mining

N.C. -- Directories

Digital humanities

THINKING HISTORICALLY ABOUT DATA:
IMPROVING AUTOMATION PROCESSES FOR HARVESTING NORTH
CAROLINA CITY DIRECTORIES

by
Pamella R. Lach

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

July 2012

Approved by

_____

Richard Marciano

Contents

List of Figures

List of Tables

The idea behind the Digging into Data Challenge is to address how "big data" changes the research landscape for the humanities and social sciences. Now that we have massive databases of materials used by scholars in the humanities and social sciences -- ranging from digitized books, newspapers, and music to transactional data like web searches, sensor data or cell phone records -- what new, computationally-based research methods might we apply? As the world becomes increasingly digital, new techniques will be needed to search, analyze, and understand these everyday materials. – Digging into Data Challenge[1]

Introduction: Big Historical Data

Big data is garnering increasing attention from humanities scholars, particularly digital humanists, for its potential to transform the nature and practice of humanistic scholarship. In its June 2012 report on the first Digging into Data Challenge, the Council on Library and Information Resources celebrates a "new era—one with the promise of revelatory explorations of our cultural heritage that will lead us to new insights and knowledge, and to a more nuanced and expansive understanding of the human condition." They foretell a "new paradigm: a digital ecology of data, algorithms, metadata, analytical and visualization tools, and new forms of scholarly expression that result from this research" emerging from the sudden explosion and expansion of digital datasets (Williford, Henry and Friedlander, p. 1-2).

Big data in the humanities combines massive digital collections with cutting-edge computing tools for efficient processing. "The ultimate goal is rapid, online, and on-

---

[1] Welcome to the challenge. Retrieved 22 June 2012 from http://www.diggingintodata.com/.

demand analysis of texts (and other resources) at the corpora scale or across distributed repositories. A scholar should be able to turn on a computer anywhere and not only access, but perform sophisticated processing on, all the world's information, or at least all that resides in digital collections" (Liu, 2012, p. 19). Liu raises a critical concern about scaling up in the digital humanities: either there is too much of a "human bottleneck" to prevent working at scale, or scholarly quality control is sacrificed for "some combination of algorithmic means and crowd sourcing … Crossing this barrier between expert knowledge and algorithmic/crowd knowledge will require … the 'scaling up' of information" (p. 20). It is not simply processes and workflows, then, that must scale up, but the humanist's approach to information and documentation must scale up to accommodate and handle big data.

The Digital Innovation Lab (DIL), housed in the College of Arts and Sciences at the University of North Carolina at Chapel Hill, launched in October 2011 in response to the rising computational and methodological challenges of big data in the humanities.[2] The DIL is predicated on a new model of the humanities and humanistic social sciences, signaled by a shift away from data scarcity towards data hyper-abundance. Its work is informed by and directed towards the digital public humanities. The DIL seeks to create tools, platforms, and processes for handling big data, focusing on automation and crowdsourcing approaches. By making digital resources more readily available, as well as reducing the cost and technical complexity of the tools required to create digital humanities projects, the DIL hopes to lower the barrier to entry for humanists, including students, faculty and cultural heritage organizations, as well as for the public at large. The

---

[2] http://digitalinnovation.unc.edu/.

DIL works collaboratively across the disciplines; its team includes computer scientists, information scientists, and scholars from American Studies, English, and History, to name a few.

One of the DIL's key projects is *P³: Connecting People, Place Past*.[3] P³ explores the intersection of people with places in the past by harnessing big historical data sets. This ongoing project is interested in putting seemingly disparate data sets into conversation with each other: maps of urban spaces, newspapers, census enumerations, and city directories. The ultimate goal of P³ is to use digital humanities processing and visualization approaches, namely spatialization, to uncover and reveal connections across data points that would otherwise be nearly impossible to discover.

The inspiration for this work largely comes from the "People of 1911 Charlotte," a visualization project that maps over 4,000 people and businesses of downtown Charlotte, North Carolina.[4] This project was undertaken as part of "Main Street, Carolina," a larger project funded by a Digital Startup Grant from the National Endowment for the Humanities (Grant number: HD-50809-09) and the C. Felix Harvey Award to Advance Institutional Priorities at UNC-Chapel Hill.[5] It created a custom-built platform for displaying historical content over historical maps. The project utilized stitched and georeferenced Sanborn® Fire Insurance Maps in conjunction with the 1911 City Directory, available through the Internet Archive.[6] This mapping project visualizes the "sorting out" process that transformed the city from a relatively integrated

---

[3] http://digitalinnovation.unc.edu/projects/p3/.
[4] http://mainstreet.lib.unc.edu/projects/charlotte/.
[5] The White Paper for this project can be downloaded at securegrants.neh.gov/publicquery/main.aspx.
[6] The City Directory can be accessed at http://www.archive.org/details/charlottenorthca1911pied. Unfortunately, the Carolina Digital Library and Archives (CDLA) at UNC did not publish or otherwise release documentation about the process of stitching and georeferencing Sanborn maps, according to a personal email to the author from Natalia Smith, 25 June 2012.

environment in the late nineteenth century to one of the most segregated spaces by 1970 (Hanchett, 1998). To complete this project, three undergraduate students, a graduate project manager, and a programmer spent months manually extracting the 4,000 data points from the 1911 City Directory, and then manually derived each data point's corresponding latitude and longitude. The resulting project is a set of markers that have been placed over a historical Sanborn map of Charlotte. The project has been used as both a teaching tool in undergraduate classrooms and a tool of public outreach in the Charlotte community. The impressive map facilitates exploration and discovery that can quickly reveal patterns, such as a white vice district in an African-American neighborhood, which might otherwise remain invisible. Similarly, the project dramatically highlights changes in the build environment by enabling users to toggle between the past and the present. This allows users to ponder the legacy of things such as Urban Renewal, which destroyed without replacing nearly 3,000 African Americans dwellings in Charlotte (Hanchett, p. 250).
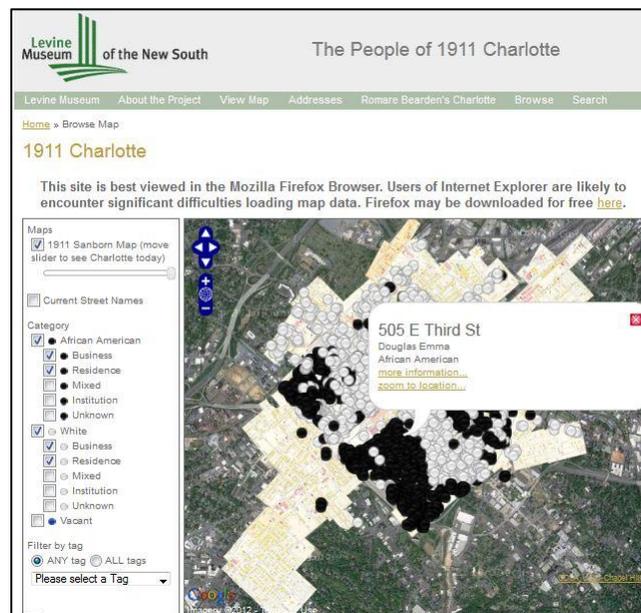
Figure 1. Screenshot of "The People of 1911 Charlotte."

Though this Charlotte mapping project has been successful, and has demonstrated the need for similar work, the process of its creation revealed limitations in the handling of the data. Namely, manually extracting individual listings from a city directory for a single year proved burdensome and un-scalable. Automation is necessary if one hopes to create a similar project for a comparable or larger city in a timely fashion. A response to this challenge, P^3 intends to create a process for automatically extracting data from digitized city directories.

City directories are a rich, yet often underutilized, historical resource which can help historians, demographers, geographers, and genealogists reconstitute places in the past, and the people who moved through those spaces.[7] These sources informed the so-called "new urban history" which emerged out of the quantitative-inflected social history movement of the 1960s and 1970s, as computing and data processing methods began to make their way into historical methodology (Thernstrom, 1971; Hershberg, 1978). But prior to mass digitization, computational analyses of city directory data were limited. In his study of the rise of segregated spaces in Charlotte, which formed the scholarly basis for the Charlotte 1911 project, Tom Hanchett manually created his own database, culled from paper or microfilmed copies of directories. The advent of digitization only just begins to suggest the full potential of city directories.

City directories, similar to today's phone books, recorded a listing of residents and businesses in a town or grouping of towns. But these directories provided more than a simple list of names. Occupations and often familial status were included. Many directories also contained separate street directories, where residents were listed a second

---

[7] For instance, Goldstein (1954) triangulated city directories in Norristown, PA with birth and death records to identify community outsiders and illustrate mobility within and beyond the community.

time by street (but occupations are usually not included here). These street directories offered a block by block listing of residents, enabling the modern reader to imagine what it might be like to walk around the town's neighborhoods and business districts (Figure 3). According to Rose-Redwood (2008), directory publishers, at least the earlier ones, were "either real estate brokers or book publishers. In some instances, those publishing a city directory had previously been employed by the government to conduct a census and then compiled a directory with the information collected" (p. 293).

Directory compilers and publishers usually "hired a team of men to 'canvass' the city door-to-door, or did so themselves. In short, they conducted a privately financed census of the city, often on an annual basis" (p. 296-297). The directories most often accounted for male heads of households, though their wives (and sometimes their children) were listed as well. Widows and other single women were normally included. Rose-Redwood notes how much of the content for the directories, though focused on male heads of households, was reported by women and servants, and thus publishers "cautioned their subscribers that this was a potential source of error" (p. 298).

While northern directories may have excluded non-white residents, southern publishers decidedly did not. These publishers recreated and reaffirmed Jim Crow segregation in their directories by demarcating race, either by segregating non-whites into separate listings, or by marking them as non-white with some sort of symbol, a * or "c" for instance (Figure 2) . While this practice is incredibly cruel by today's standards, the resulting directories nonetheless provide historians and demographers with a wealth of information, as the Charlotte 1911 project's mapping component demonstrates.

9

WATCH :: CHARLOTTE :: GROW

Robinson's Book Store
30 West Trade

SCHOOL AND COLLEGE
—BOOKS—
GENERAL SCHOOL SUPPLIES

306   CHARLOTTE [1911] DIRECTORY

*Mobley Jas (Cora), driver R H Field Co, h 514 e 1st
*Mobley Jno (Mary), lab, h 616 s Myers
Mobley Jno W (Clara N), pres Doggett-Mobley Co, h 200 e Boulevard
*Mobley Johnson Rev (Sylvia), h 811 e 7th
MODEL STEAM LAUNDRY, 28-40 w 5th—phones 110 and 160; P M Cave pres and mngr, R A Lee v-pres (see gold line back cover)
Moffatt Machinery Mnfg Co, 506-512 w 3d; W E Moffatt pres, Z V Kendrick v-pres, E J Matthews sec-treas
Moffatt Wm E (Martha), pres Moffatt Machinery Mnfg Co, h 509 s Tryon
Moffett Enoch A (Henrietta), gro 810 n Graham, h same
*Moffett Jas (Gore & Co), h Ross Town
Moffett Lucy Mrs, stengr Char Power Co, h s Myers
*Moffett Saml, emp Oil Mills, h 203 (801½) s Caldwell
*Moffett Saml Jr, hackman, h 801½ s Caldwell
*Moffett Wm, waiter, h 801½ s Caldwell
MOHN E L COAL & IRON CO, 318-320 Amer Bldg; Jno R Purser agt (see p opp back paster)
Monaghan Jos (Catherine), (W M Milligan & C), h 7 e Stonewall
MONONGAHELA INS CO, 508 Trust Bldg—phone 655; J F Flowers, genl agt
*Monroe Ester, nurse 304 n College
*Monroe Ida, h 211 Pine
*Monroe Wm, waiter Stonewall Hotel, h 507 w 5th
*Monroe Wm (Susan), lab, h Latta Row, D'worth
Montague Jas J (Elizabeth), bkkpr Jos Sykes Bros, h 911 Pine
MONTAGUE MNFG CO (Inc), lumber and building material, 400 s 9th, Richmond Va (see bottom lines and bet pages 188-189)
Monteith Dickson S (Edith), draftsman, h 212 Central av
Monteith Maggie, wid Jno, boarding 313 n Church, h same
*Monteith Otis (Mary), fireman Sou Ry, h Dulstown
Monteith R Henry, driver, h 313 e 10th
*Monteith Robt (Zemora), wks Liddell Co, h Greenville
*Monteith Thos (Maggie), driver, h 420 w 6th
*Montgomery Alex (Mary), lab, h Shorte av nr Fox av
*Montgomery Alice, h 804 e 10th
Montgomery C M, h Belmont rd
*Montgomery Caesar (Sallie), lab, h 611 e Stonewall
Montgomery Clyde P (Gladys), bkkpr R H Jordan & Co, h 507 w 8th
Montgomery Earle R, student, h 1012 w 1st
Montgomery Gracie L Miss, h 1012 w 1st
Montgomery Hugh (Grace), agent, h 212 e Boulevard
Montgomery J Boyd (Goldie), loomfixer, h 303 e 11th
*Montgomery J Lewis (Lizzie), driver C Valaer Bottlg Wks, h 410 e Stonewall
Montgomery Jas, tinner, rms 214 n Church
Montgomery Jas T (Della C), opr, h 504 Elizabeth av
Montgomery Jno C (Ada), phys 26 s Tryon, h 912 s Tryon
Montgomery L Sidney (Carrie), ins agt, h Seversville
Montgomery M Virginia, wid Winslow A, h 415 n Graham
Montgomery Margaret E, wid Jas T, h 508 e 9th
*Montgomery Mary, h 217 Pine

Figure 2. A page from the Charlotte 1911 Directory. Note the * denoting non-white individuals.

CALDWELL   498   CALDWELL

American Machine & Mnfg. Co.
Textile and Oil Mill Specialties.
Engine & Machine Repairing
2000-2010 s Boulevard

307 *Nicholson Effie
308 *Butler E W
309 *Ross Jno
309½ Parks Chas
310 *Clinton James
311 *Turner Hattie
312 *Stewart Rufus
313 *Fundburg Hampton
315 *Owens Henry
316 *Plummer Hattie
    (7th intersects)
408 *Hawkins Frank
400 *Eagles J L
412 *Cardwell Martha
413 *Gillard Jas
414 *Collins Eliza
415 *Crawford G H
416 *Williams J W
417 *Gillard Cyrus
418 *Broadway Dock
419 *Davis Margaret
420 *Glenn Elsie
    (8th intersects)
502 *Williams David
504-506 *Vacant
505 *McCombs Edith
507 *Williams Henry
510 *Rice Bailey
512 *Robinson Jas
514 *Brewer Wm
516 *Hannon Susan
518 *Vacant
    (9th intersects)
600 Ninth av Bapt Ch
602 Pruett L R Rev
606 under constr
607 Brown J W
609 Martin E C
610 Stroup C T
    Wilcox Walter
611 Rogers S R
612 Alexander J F
613 Wearn W H
614 Hall F N
615 Vacant
616 Cochrane R L

618 McLaughlin J M
    (10th intersects)
700 Earnhardt M M
701 Glenn J C
702 Fink D W
    Harrell J P
704 Marshall Margaret A Miss
706 Sing Jane E Mrs
707 Davis James
708 Polk T F
    Polk Buena V Mrs, boarding
709 Renn E W
710 Goos Gottlieb
711 Wodside J Mc
712 Kennerley A F
713 Pyron T N
714 Cook C W
715 Vacant
    (11th intersects)
802 Vacant (store)
804 Lockman C R
805 Howell E F
806 Barrett J J
807 Barrington J A
    Barefoot M M
808 Queen City Dyeing & C Wks
 — DuBose J H
809 Bradshaw Chas
    Bradshaw J L
    Bradshaw Emma Mrs, dressmkr
810 Schachner G S
811 Dempster R J
812 Blackwelder C T
813 Ramsey C W
814 Lawing R F
816 Graham Mary Mrs
817 St Paul's Presby Ch
    (12th intersects)
900 Goldstein David, gro
901 Pangle Sarah A Miss
902 Goldstein David
903 Parker W A
904 Holtzclaw D H
905 Rimmer J W
906 Mills J H
907 White J T
908 Presson J B
909 Lawson Julius
910 Hill Arthur
911 Weddington W T
    Thompson J H
912 Stewart W E
913 Walker T N
914 McCall D H

Figure 3. Street Directory for Charlotte 1911.

City directories were touted as an important community and business resource. Free copies were made available in "public spaces, such as post offices, hotels, railroad depots, libraries, banks, and on steamships, thereby drawing attention to a city from those at a distance" (Rose-Redwood, p. 297). Directories were also available by subscription and were marketed heavily to local business owners. Hill's 1933 Raleigh Directory contains a two-page spread: "City Directory Uses for Every Business" (p. 46-47). The uses are organized into five categories: sales promotion, credit, delivery and shipping, purchasing, and general information (Figure 4).

Figure 4. City Directory Uses for Every Businesses, Raleigh City Directory 1933.

There are many challenges to extracting the content from digitized city directories into usable output. For one, the optical character recognition (OCR) performed on the

directories is limited; often the software cannot detect * or "c", which are both critical indicators of race. Secondly, the OCR text output is typically a single text or XML file, which is not easily searchable. Finally, there is a lot of "noise" on a typical directory page; namely irrelevant information such as advertisements and headers. Advertisements can appear as print ads, or as text running vertically along the side of a page (see Figures 2 and 3). There is rarely any consistency in their placement even within a single directory. On its own, a machine cannot discern relevant from "noisy" information, nor can it detect subtle nuances found in an inconsistent and unpredictable historical collection.

In order to parse the digitized collection of city directories for North Carolina, a thorough understanding of the range of historical variance across the collection is required. This project seeks to identify a common set of attributes for creating a set of parsers that can be applied to the collection to automate (or semi-automate) the data extraction process. Combining historical and computational methodologies will enable data harvesting at scale.

Literature Review

There is very little precedence for parsing methodologies that take historical variance into account. Indeed, much of the humanistic work on text mining falls into literary, rather than historical, domains. Beyond TEI-based work, humanistic data harvesting projects often focus on parsing parts of speech (POS) and grammatical clauses (Hundt, Denison & Schneider, 2012).

There are several digital humanities projects that address data mining in historical contexts. Dan Cohen, of the Roy Rosenzweig Center for History and New Media at George Mason University, has been exploring and developing application programming interfaces (APIs) for humanities data mining. While working on his Syllabus Finder he determined that, rather than build his own internal database of humanities-related syllabi, it would be far easier to develop a tool to crawl the web. In order for this to succeed, he conducted an attribute analysis to determine a set of elements contained in all or most syllabi. These elements became parameters for improving precision and recall in his database. Cohen's second project, H-BOT, is "an automated historical fact finder" that relies on the construction of a lookup table of irregular verbs to translate and interpret search term queries (for instance, to allow the system to handle a range of tenses). Cohen's work blends humanistic thinking with computational approaches. As he argues, "these computational methods, which allow us to find patterns, determine relationships, categorize documents, and extract information from massive corpuses, will form the basis for new tools for research in the humanities and other disciplines in the coming decade" (Cohen, 2006). His approaches serve as important models for applying computational power to historical data, as well as for thinking historically about computational challenges.

The closest public digital humanities project to the DIL's P[3] undertaking is "AddressingHistory: People, Places, Professions."[8] This project is run by Edina, in partnership with the National Library of Scotland and funded by JISC's Developing Community Content program. It is a Web 2.0 mash-up of over seven hundred Edinburgh

---

[8] http://addressinghistory.edina.ac.uk/. Accompanying blog available at http://addressinghistory.blogs.edina.ac.uk/.

post office directories and historical maps for 1784-1785, 1865 and 1905-1906. The project team automatically populates their database by parsing XML output from Internet Archives digital files of directories. They are experimenting with auto-generating geocodes for the directory entries, with crowdsourcing techniques forming a second pass to clean the data (Macdonald web log post, 2011). This mixed approach, they maintain, will "lead to a fully geo-coded version of the digitised directories thus providing significant added-value to the general public, local historians and specialist researchers across multiple disciplines" (About page, AddressingHistory blog). Their work is an excellent roadmap for P^3 both in scope and technical approaches.

Likewise, much can be gleaned from text mining efforts, particularly natural language processing (NLP), information extraction using named entities, and automated metadata creation. Though very little current research focuses on the challenges of applying such processes and methodologies to varied historical documents, many researchers are interested in extracting content from diverse collections of digital documents. Pekar and Evans (2007) have developed an approach to information extraction (IE) over a diverse set of documents. They define IE as "an area of research that aims to perform … intelligent analysis of the contents of documents. The goal of an IE system is to extract text fragments instantiating predefined semantic entities that can be mapped to fields in a database and later easily manipulated using database queries" (p. 330). Theirs is a "two-step machine learning approach that first aims to determine segments of a page that are likely to contain relevant facts and then delimits specific natural language expressions with which to fill template fields" (abstract, p. 329).

Though their work focuses on crawling a variety of web pages, rather than harvesting data from a single dataset, their approach is nonetheless instructive. Historical data can be treated in an analogous way to a set of wide ranging web pages, since historical documents can exhibit a high degree of unpredictable variance even among seemingly similar document types such as city directories. As Pekar and Evans explain, "the layout of a document can constitute very valuable evidence for information extraction. However, not only is there hardly any consistency in the layout of web pages retrieved by a domain crawler, but also one can seldom rely on even general formatting cues occurring on the page" (p. 331). By contrast, the general formatting of a city directory, such as the presence of a header or footer, *can* provide the appropriate cues for a parser, as this project seeks to demonstrate.

Similar lessons can be learned from attempts to automatically generate annotations within digital collections. Bontcheva, Maynard, Cunningham, and Saggion (2002) faced many challenges in annotating and indexing the OldBailey digital collection. Those difficulties resonate strongly with the problems of historical data harvesting. Relying on named entity recognition in an IE approach, they found traditional IE systems required modifications unique to the eighteenth-century collection of documents. For instance, they could not rely solely on an existing gazetteer, or lookup list, for automatic annotation creation; the list needed to be expanded to account for social status, historically peculiar occupations, and abbreviations for given names. Indeed, many researchers dealing with humanistic-like content find that traditional gazetteers, whether grammatical or informational, are not sufficient on their own (Pekar and Evans, p. 332-333). Bontcheva et al.'s approach allowed the team "to identify the

parameters relevant to the creation of a name recognition system robust across these types of variability," thereby resulting in a system that can hold up against a range of domains and genres (p. 143).

Ultimately, these studies show that pure computational approaches will likely fall short when faced with the peculiarities of historical collections. Sculley and Pasanek (2008), for instance, caution that too heavy a reliance on pure computational methods and outputs could be dangerous in the humanities. They call for a balance between computation and humanistic analysis and interpretation. Their work focuses on data mining and machine learning, which they pit against textual analysis. They argue for a leveraging of tools in combination with human capacity, recognizing that such tools play an important role in facilitating humanists' work. But those tools, they remind us, cannot fully replace the work of humanists. To take one example, they demonstrate the ways in which the digital humanities consistently disprove the learning theory of data, which "assume[s] that the data is produced by some process with constant probabilistic qualities … The key is that the distribution's probabilistic behavior *does not change over time,* and that it will continue to produce as many examples as requested" (p. 411). Changeability is the very nature of "historical data," which varies over time and location (p. 411). Pure computational approaches to data will inevitably fail in these kinds of situations because they do not take data variation into account. Rather, in the digital humanities, the "No Free Lunch Theorem" holds—"there is no single best learning algorithm, and we may have to employ a good deal of ingenuity to learn from difficult data" (p. 413). Historical mindfulness, then, must accompany machine learning for data harvesting.

Methodology

During the 2011-2012 academic year, the DIL created a parser based on a bundle of three Python scripts that transform an Internet Archive dejavu.xml file into a CSV text file, which can then be processed through Google Refine into tuples with unique attributes. One tuple is the equivalent of a single entry in a city directory (last name, first name, race, marital status, occupation, address). Producing a set of distinct values for each directory entry will facilitate and strengthen search capabilities, enable mapping and other visualization activities similar to the Charlotte 1911 project, and support other analytical approaches.

By the end of the spring semester, the combined scripts successfully extracted all of the entries from the Charlotte 1911 city directory (the test bed directory). The script largely ignored the advertisements intermixed within the general alphabetical listing of residents. That is, the script proved smart enough to pick up relevant information while skipping over irrelevant items.

Though these scripts effectively parsed the Charlotte 1911 City Directory using pattern matching, they will undoubtedly fail when applied to another location, time period, and/or directory publisher. City directories, as a historical data set, range significantly over publisher, location, and publication date. To take one example, different publishers conveyed race in different ways and at different times. Many used * to denote non-white status, while others use (c) or (col). Hill Directory Co. of Richmond, VA and publisher of roughly 45% of NC's digitized directories,[9] employed several different designations: * from the 1890s through 1926, then (c) up through 1952, when

---

[9] This number is based on the collection of directories hosted at Internet Archive as of 2 May 2012: http://archive.org/search.php?query=subject%3A%28N.C.%20directories%29. Many more have since been digitized, but are not included in this analysis.

the practice of racial designations appears to have ceased. Still other directory publishers did not use any designation, choosing instead to segregate the races into separate directory listings. All told, there are eight major ways that race is handled across this collection, which ranges from 1860 to 1963 (Table 1).

Table 1. Depiction of race across publishers. N=415

Racial Designation for non-whites	Total Instances	No. Publishers	Years
*	216	13	1875-1947
(c)	130	5	1871-1951
No designation (typically separate listings for white and "colored")	32	16	1860-1933; 1953-1963
©	11	1	1936-1940
c	9	8	1866-1948
(col)/col	7	6	1887-1897
None in general directory but * used in street directory	8	2	1911-1922
None in general directory but (c) used in street directory	2	1	1935-1936

To support automated parsing of as many North Carolina directories as possible, the range of difference across publishers, locations and/or time periods must be assessed to determine a common set of groupings based on similar structural attributes. These groupings will then form the basis for a variation of bundled Python parsers that could be applied to some subset of the directory collection. This project is designed to discover that set of attributes to facilitate building a workable parser.

Deriving a complete set of attributes required the creation of a complete list of digitized directories, including a listing of publishers, cities, and dates. Though the paper directories are housed in various repositories across the state, their digital counterparts

are hosted centrally at the Internet Archive (IA).[10] Given the use of Python as a scripting

language for other pieces of P^3, two additional Python scripts were written for this phase.

Applied in combination, these scripts crawled the IA's collection using "N.C. directories"

as the search term in the Advanced Search. The first script pulled IA metadata (author,

publisher, volume, etc.) out of a single entry using regular expression pattern matching.[11]

The second script iterated through each of the fifty entries on an individual search result

page, and then iterated through each page of the search results, invoking the first script to

pull out all of the appropriate metadata. The metadata pulled out by the script was

formatted with XML-like tags (e.g. print "<TITLE>", matchObj.group(1), "</TITLE>")

to support future transformation of the Python output into valid XML.

The script ran successfully on 2 May 2012, picking up 427 entries in its output.

Three were irrelevant and nine more were duplicate directory entries that had been

digitized by two different institutions, for a final count of 415 unique directories. The text

output was saved as an XML file, validated against a schema created for this purpose,

and then transformed into Excel with the insertion of <?mso-application

progid="Excel.Sheet"?> into the second line of the XML file.[12] Once in Excel, the

content was cleaned and disambiguated. Missing data (such as a URL or year) was filled

in manually. Unique IDs were added to each entry, city names were verified against the

original files, and a state field was added to aid in spatial visualizations.

---

[10] The Internet Archive is not the only hosting service for city directories, though it seems to be the central repository for North Carolina's collection. N.C. directories constitute more than half of the IA's entire collection of directories as of 27 June 2012.

[11] This approach was informed by TutorialsPoint's tutorial, "Python-Regular Expressions" retrieved 12 February 2012 from http://www.tutorialspoint.com/python/python_reg_expressions.htm with additional guidance from Lutz (2009).

[12] See http://en.wikipedia.org/wiki/Microsoft_Office_XML_formats and http://www.brainbell.com/tutorials/ms-office/excel/Save_To_SpreadsheetML_And_Extracting_Data.htm both accessed 8 April 2012. Ampersands in the IA metadata were changed to "and" to comply with XML standards.

Once relatively clean, attribute columns were added to the original Excel
spreadsheet. Each entry was manually coded over the course of about 90 hours; a more
automated process for deriving the set of attributes would be required to scale this
project. Likely some of these fields could be eliminated if repeating this process in the
future. The following attributes were manually coded in Excel:

- **Read Online File URL** – the Python script pulled the URL for each
  entry's main web page, which contains its metadata and digital file in
  various formats. The URL for this field corresponds to the streaming
  image, typically a 2up digitized file (PDF and text files are also available).
  This allows for direct navigation to the digitized file.

- **City** – the city name was collected to compare against the original IA
  metadata. In the event that there was a discrepancy, this field was favored.

- **Publisher** – this was a critical field, and one which varied from the IA's
  metadata. IA metadata often listed publishers and authors separately and at
  times inconsistently. For this analysis, publisher was broken into two
  fields: primary publisher and secondary (often a regional publisher
  working with a local press). These fields were later refined into a more
  standardized list of publishers (see Results).

- **Year** – the year(s) covered by the directory (not the year of copyright), as
  listed either on the title page or, if unavailable or unreadable, from the
  general alphabetical listing's title or header. The year often differed from
  the one provided by the IA. Frequently directories were published

biannually, in which case a two-year span was noted but later refined into a single year (using the first year in the range) for sorting purposes.

- **Price** – if available, typically listed on the cover page. This is not an essential field and could be dropped in future iterations of this project, particularly because price can often not be determined.

- **Directories appearing before the General Alphabetical Directory** – the original Python parser looked for the word "Directory" as the starting point for parsing. However, many directories included a detailed street directory or telephone directory in the front of the book, before the general directory. In cases such as these, the parser would pick up irrelevant data. It is critical to know what the starting point for the parser should be in order to eliminate as much noise as possible.

- **Starting page of General Alphabetical Directory** – Because the main listings were not always at the front of the book, one possible solution for the parser problem would be to tell it to jump to a certain page. Note that the starting page of the actual directory (e.g. page 101) does not always correspond to the actual page in the digitized file; many scans skip over the front matter and advertisements. The page number, then, corresponds to the page number visible on the directory page, often located in the header. In cases where separate white and colored directories existed, this field was treated as the start page of the white directory.

- **Colored Directory Starting Page** – in about 6.75% of cases, whites and non-whites ("colored") were listed in separate directories. This was noted,

in large part because many colored directories provided no other indication of race.

- **Street Directory** – this field tracks the presence (or lack thereof) of a street directory, in the event that the parser were ever expanded to include the street listings. To this end, additional attributes were collected about the street directory (see below).

- **Abbreviations** – typically, a city directory used a series of abbreviations for occupations, directional references, and even marital status (e.g. "wid" for widow). This field indicates the presence of general/occupational abbreviations, given names (Chas for Charles), special abbreviations, street and suburb abbreviations, and abbreviations of local firms. This information may be used to build a lookup table of controlled vocabulary to improve parser performance.

- **Title of General Alphabetical Directory** – since the parser is not yet "smart" enough to know where to start, this field can help parameterize where the parser should begin.

- **Race** – denotes how race was represented in the directory, if at all (see Table 1).

- **Married** – many directories not only indicated a person's marital status, but many listed the wives in parenthesis. Conversely, this means that frequently only men and male heads of household, as well as single adult women, were listed.

- **Number of Columns** – the parser will need to know whether the general alphabetical directory consists of one or two columns.

- **Header or Footer** – used to train the parser to detect either a standard header or, in rare cases, a footer, to help eliminate noise such as advertisements appearing above (or below) the header (footer).

- **Header (Footer) Details** – format for the even page, the odd page, and whether the pages are mirror images of each other or identical. For instance, a common format is the mirror images of *page number / city name, N. C. (year) / City Directory* (where the page number is on the left side on even pages and the right side on odd pages).

- **Last name continuations** – this field captures the symbols used to convey a repeating surname in the general alphabetical directory (typically a dash or quotation marks).

- **Notes on racial listings** – provides additional information about how race was denoted, if at all. For instance, some directories that had separate white and colored listings with no racial designation used racial codes in the combined street directory.

- **Other special symbols** – captures other symbols that might trip up the parser, such as an "h" or an "o" in a circle to connote home ownership, or a bell icon to denote that a household has a telephone.

- **Street Index information** – in the event that the parser is ever expanded to process street indexes, it will be critical to know whether street indexes have the same **number of columns** as the general alphabetical directory,

and whether the **header/footer** is the same or different. Additional

symbols (e.g. a dash to connote a missing building number) are also

indicated here.

- **Notes on digitized copy and/or CD contents** – lists any peculiarities that

  might trip the parser, such as missing pages.

These fields formed the basis of the analysis and proposed set of attributes for refining

the parser.

Results

Once duplicate and irrelevant entries were verified and removed, and all attributes

collected, analysis could begin with the remaining set of 415 directories. All told, the

collection represented forty-six unique publishers ranging from as early as 1860 to as late

as 1963 (Table 2).

Table 2. Count of publishers. N=415

Publisher	Count	% of Collection	Years	Mean Year	Median Year	Mode Year
Hill Directory Co.	189	45.54%	1897-1963	1908	1908	1934
Miller-Commercial Directory Co.	60	14.46%	1916-1937	1933	1939	1930
Miller (Chas)-Southern Directory Co.	43	10.36%	1931-1947	1924	1937	1939
Miller-Piedmont Directory Co.	40	9.64%	1908-1947	1922	1914	1913
Baldwin Directory Co.	14	3.37%	1935-1940	1937	1937	1937
Chas S. Gardiner	6	1.45%	1915-1918	1916	1916	1916
Edwards, Broughton and Co.	6	1.45%	1880-1893	1884	1892	1880
Maloney Directory Co.	6	1.45%	1899-1901	1900	1900	1899

Miller-Piedmont Directory Co.-Hackney and Moale	5	1.20%	1907-1915	1911	1911	1915
Walsh Directory Co. (Charleston, SC)	5	1.20%	1905-1910	1908	1908	N/A
Carolina Directory Co.	2	0.48%	1932-1933			
E. F. Turner and Co.	2	0.48%	1889-1890			
Miller-Commercial Directory Co. (Piedmont)	2	0.48%	1916			
Miller-Southern Directory Co.	2	0.48%	1943-1945			
P. Heinsberger (Wilmington, NC)	2	0.48%	1865-1871			
Baughman Brothers	1	0.24%	1883			
Beasley and Emerson	1	0.24%	1875			
Benj. R. Sherriff	1	0.24%	1877			
Brady Printing Co.	1	0.24%	1920			
Chase Brenizer	1	0.24%	1896			
Frank D. Smaw	1	0.24%	1866			
Franklin Printing and Publishing Co., J.S. McIlwaine, publisher	1	0.24%	1896			
Geo. H. Kelley	1	0.24%	1860			
Hackney and Moale	1	0.24%	1906			
Home Directory Co. (Hickory, NC)	1	0.24%	1935			
I. E. Maxwell (Hendersonville, NC)	1	0.24%	1915			
Interstate Directory Co. (Atlanta, GA)	1	0.24%	1884			
Interstate Directory Co. (Charlotte, NC)	1	0.24%	1902			
J. Edwin Carter and A. Kyle Sydnor	1	0.24%	1913			
J. H. Chataigne (Raleigh, NC)	1	0.24%	1875			
Levi Branson (Raleigh, NC)	1	0.24%	1887			
Miller Press	1	0.24%	1939			
Miller-Commercial Directory Co.-Hackney and Moale	1	0.24%	1918			
N.A. Ramsey (Durham, NC)	1	0.24%	1892			
Observer Printing Co (Raleigh, NC)	1	0.24%	1888			
Page Trust Co.	1	0.24%	1916			
Raleigh Stationery Company	1	0.24%	1896			
Samuel L. Adams (Durham, NC)	1	0.24%	1902			
Seeman Printery (Durham, NC)	1	0.24%	1911			

Southern Directory Co.	1	0.24%	1887			
Stone and Kendall	1	0.24%	1892			
The Educator Company (Durham, NC)	1	0.24%	1897			
Thompson, Breed and Crofill (Newburgh, NY)	1	0.24%	1887			
Turner, M'Lean and Losee Directory Co. (Raleigh, NC)	1	0.24%	1886			
Walker, Evans and Cogswell (Charleston, SC)	1	0.24%	1890			
Wilmington Messenger / Messenger Steam Presses	1	0.24%	1889			

Given similarities observed in the various Miller directories, including formatting and editor (Ernest Miller and, later, Chas Miller), these various publishers were combined into one catch-all publisher category:

- Miller-Commercial Directory Co. (N=60)

- Miller (Chas)-Southern Directory Co. (N=43)

- Miller-Piedmont Directory Co. (N=40)

- Miller-Piedmont Directory Co.-Hackney and Moale (N=5)

- Miller-Commercial Directory Co. (Piedmont) (N=2)

- Miller-Southern Directory Co. (N=2)

- Miller Press (N=1)

- Miller-Commercial Directory Co.-Hackney and Moale (N=1)

Subsequent visualizations adhere to this grouping unless otherwise indicated. To facilitate attribute analysis, the Millers were then disaggregated into two categories, where Miller-Piedmont Directory Co. and Miller-Piedmont Directory Co.-Hackney and Moale (N=45 combined) were removed from the larger group to form the "Miller-Piedmont" cluster. A small amount of processing was similarly performed on some of the

other publishers for grouping purposes; all of the individual Baldwin entries were brought

together under Baldwin Directory Co.

Table 3. Top ten publishers, with all Millers combined.

Publisher	Count	Years
Hill Directory Co.	189	1897-1963
Miller Press/Piedmont/Commercial/Southern Directory Co.	154	1907-1947
Baldwin Directory Co.	14	1935-1940
Chas S. Gardiner	6	1915-1918
Edwards, Broughton and Co.	6	1880-1893
Maloney Directory Co.	6	1899-1901
Walsh Directory Co. (Charleston, SC)	5	1905-1910
Carolina Directory Co.	2	1932-1933
E. F. Turner and Co.	2	1889-1890
P. Heinsberger (Wilmington, NC)	2	1865-1871

All told, publishers with at least five publications account for nearly 92% of the entire

digitized collection (with Millers combined). In light of this, and because there were so

many publishers with just one or two directories, only those with at least five volumes

were used in the attribute analysis.

A spatial visualization of the distribution of publishers confirms that there is a

degree of regional clustering across nearly fifty North Carolina towns, spanning the first

half of the twentieth century, with a handful of mid-to-late nineteenth century directories

(Figure 5).[13] The map was created using the open-source, Google-based batchgeo.com.[14]

Hill Directory Co., for instance, was based in Richmond, VA and accounted for 45.54%

---

[13] The span of years is representative not of the entire collection of N.C. city directories per se, but of those digitized and available from the Internet Archive. Different institutions have different policies about scanning material. The University of North Carolina-Greensboro has digitized through 1963. UNC-Chapel Hill has only recently begun digitizing directories after 1923, having determined that post-1923 directories should still be considered in the public domain.

[14] An interactive map is available at http://batchgeo.com/map/143617761673e23037159ea75d1cc56e.

of the entire collection. Its geographic spread seems largely rooted in the Piedmont

region, with additional coverage of the coast in Wilmington (Figures 6 and 7).
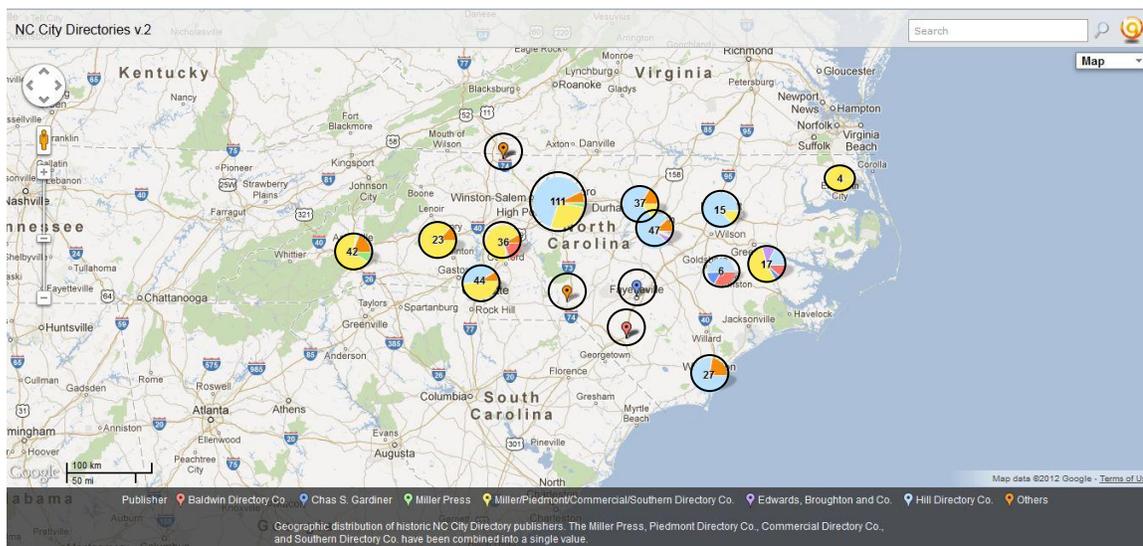
Figure 5. Geographic distribution of publishers across North Carolina.
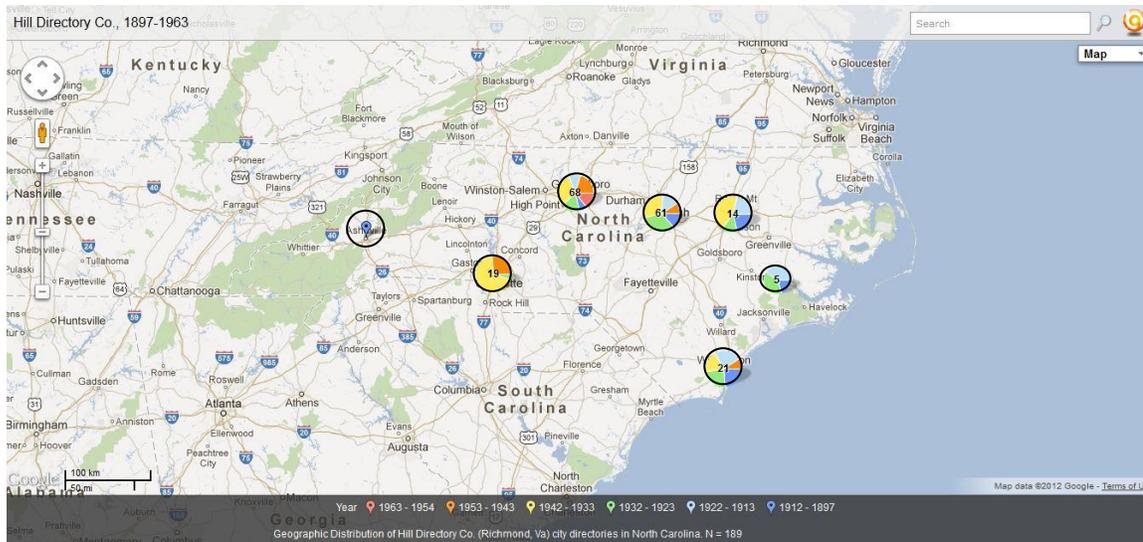
Figure 6. Geographic and temporal distribution of Hill Directory Co. N=189.[15]

---

[15] Interactive map can be accessed at http://batchgeo.com/map/c0c7840e0391e6c99a27915f45452633.
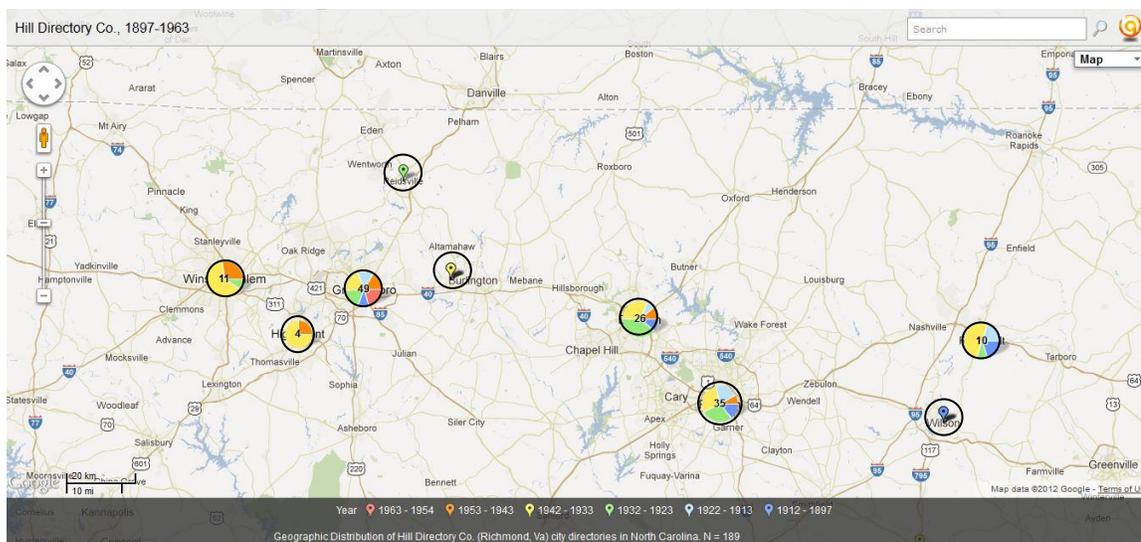
Figure 7. A closer look at Hill's distribution in the Piedmont region.

By comparison, the various Miller presses, while based in Asheville, represent a more

expansive reach, stemming from Asheville in the western mountains all the way to

Elizabeth City in the northeast part of the state (Figures 8 and 9). This suggests that city

directory publishing was often a regional enterprise.
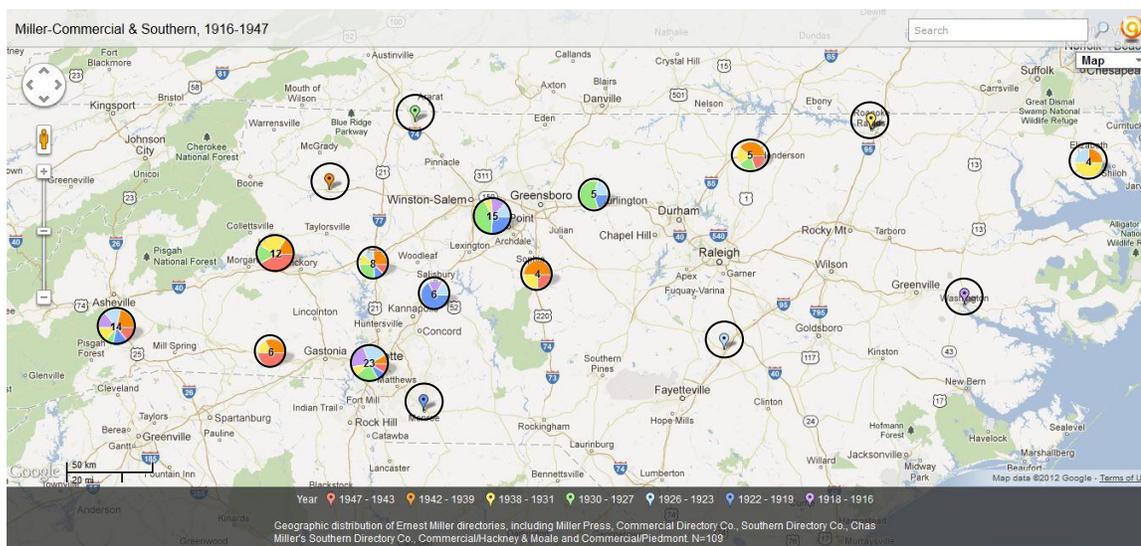
Figure 8. Geographic and temporal distribution of the Millers. N=109.[16]

---

[16] Interactive map available at http://batchgeo.com/map/eb891bf29310f3c35d81352bfc5e1e8a.
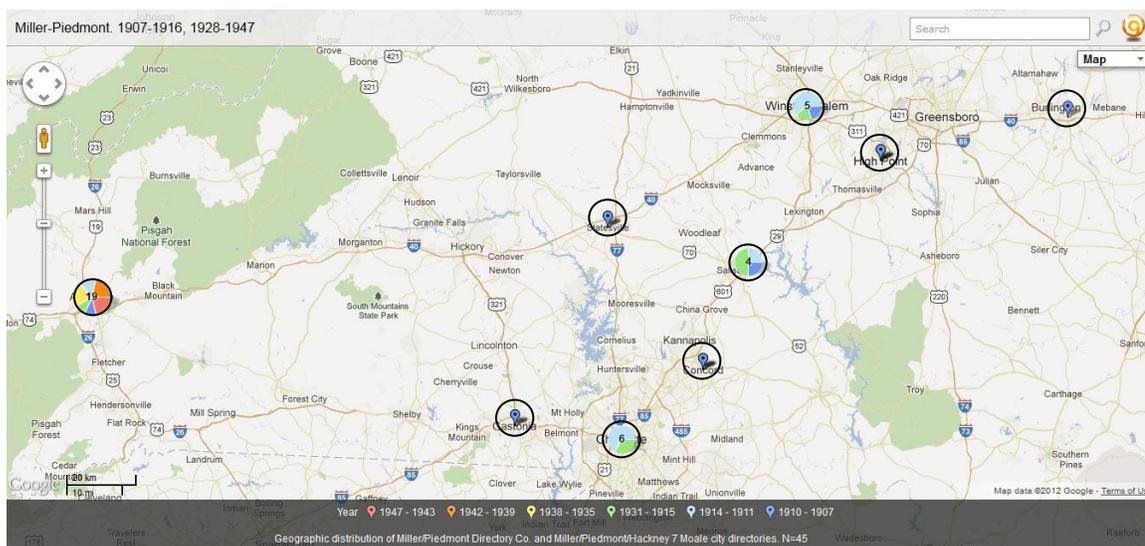
Figure 9. Geographic and temporal distribution of the Miller-Piedmonts. N=45.[17]

Taking a closer look, it appears there is quite a range of both locations and time periods

for most publishers, making it difficult to set parser parameters based solely on geo-

spatial factors (Table 4).

Table 4. Geographic distribution by publisher.

Publisher	City	City Count	Year
Baldwin Directory Co. 1935-1940	Albermarle	1	1937
	Concord	2	1938, 1940
	Kinston	1	1936
	Goldsboro	1	1938
	Lexington	1	1937
	Lumberton	1	1938
	New Bern	1	1937
	Salisbury	3	1935, 1938, 1940
	Reidsville	1	1935
	Thomasville	1	1935

---

[17] Interactive map available at http://batchgeo.com/map/9d03149c7a5964af1df4b55cd8bf0176.

	Washington	1	1937
Chas S. Gardiner 1915-1918	Dunn	1	1918
	Fayetteville	1	1915
	Goldsboro	1	1916
	Lexington	1	1916
	Lumberton	1	1916
	New Bern	1	1916
Edwards, Broughton and Co. 1879-1893	New Bern	2	1880, 1893
	Raleigh	3	1880, 1883, 1893
	Winston-Salem, Greensboro	1	1879
Hill Directory Co. 1897-1963	Asheville	1	1906
	Burlington (inc Graham, Haw River and Elon College)	3	1935, 1943, 1946
	Charlotte	15	1932-1947
	Durham	26	1903-1949
	Gastonia	4	1934, 1936, 1942, 1947
	Goldsboro	1	1934
	Greensboro	49	1909-1963
	High Point	4	1933, 1940, 1942, 1948
	Kinston	2	1923, 1928
	New Bern	3	1907, 1918, 1920
	Raleigh	35	1903-1948
	Reidsville	1	1932
	Rocky Mount	10	1908-1942
	Wilmington	21	1897-1947
	Wilson	3	1912, 1916, 1930
	Winston-Salem	11	1932-1947
Maloney Directory Co. 1899-1901	Greensboro	2	1899, 1901
	Asheville	2	1899, 1900
	Raleigh	2	1899-1901
Miller (Chas)-Southern Directory Co. 1931-1947	Asheboro	4	1937, 1939, 1941, 1947
	Canton	2	1937, 1942
	Elizabeth City	3	1938, 1938, 1942
	Greenville	6	1936-1947
	Henderson	4	1938, 1940, 1942, 1947
	Hendersonville	5	1937-1945
	Hickory	3	1937, 1943, 1947
	Mooresville	1	1939
	Morganton	3	1931, 1939, 1943
	N. Wilkesboro/Wilkesboro	1	1939

	Roanoke Rapids	2	1938, 1942
	Shelby	6	1937-1947
	Statesville	3	1938, 1940, 1944
Miller-Commercial Directory Co. 1916-1937	Asheville	4	1917, 1918, 1922, 1930
	Burlington, Graham and Haw River	4	1920, 1924, 1927, 1929
	Charlotte	11	1916-1930
	Concord	2	1920, 1922
	Dunn	1	1926
	Elizabeth City	1	1923
	Gastonia	4	1918, 1921, 1923, 1930
	Greenville	1	1926
	Hendersonville	3	1921, 1924, 1926
	Hickory	1	1930
	High Point	8	1919-1930
	Lenoir	2	1930, 1937
	Monroe	1	1922
	Mount Airy	1	1928
	Oxford	1	1929
	Reidsville	1	1929
	Salisbury-Spencer	4	1917, 1919, 1922, 1924
	Statesville	4	1922, 1925, 1928, 1930
	Thomasville	1	1930
	Winston-Salem	5	1916-1931
Miller-Piedmont Directory Co. 1908-1947, gap between 1916-1928	Asheville	19	1909-1916; 1931-1947
	Burlington, Graham and Haw River	1	1909
	Charlotte	6	1911-1915, 1931
	Concord	3	1908, 1913, 1916
	Gastonia	2	1910, 1913
	High Point	3	1908, 1910, 1913
	Salisbury-Spencer	2	1913, 1928
	Statesville	1	1916
	Winston-Salem	3	1910, 1912, 1913
Miller-Piedmont Directory Co.-Hackney and Moale 1907-1915	Salisbury-Spencer	2	1907, 1915
	Statesville and Iredell County	1	1909
	Winston-Salem	2	1911, 1915
Walsh Directory Co. (Charleston, SC) 1905-1910	Charlotte	4	1905, 1907, 1909, 1910
	Winston-Salem	1	1908
Other Millers (excludes 19th Century Miller Press)			
Miller-Commercial	Greenville	1	1916

	Washington	1	1916
Directory Co. (Piedmont)			
Miller-Southern Directory Co.	Lenoir	2	1943, 1945
Miller-Commercial Directory Co.-Hackney and Moale	Winston-Salem	1	1918

Looking at it from a slightly different perspective, there is a similarly diverse spread of

publishers for most towns (Table 5).

Table 5. Distribution of publishers by town.

City	Publisher	Count	Years
Albermarle	Baldwin Director Co.	1	1937
Asheboro	Miller (Chas)-Southern Directory Co.	4	1937, 1939, 1941, 1947
Asheville	Franklin Printing and Publishing Co., J.S. McIlwaine, publisher	1	1896
	Hill Directory Co.	1	1906
	Maloney Directory Co.	2	1899-1900
	Miller-Commercial Directory Co.	4	1917, 1918, 1922, 1930
	Miller-Piedmont Directory Co.	19	1909, 1910, 1912, 1913, 1914, 1916, 1931, 1935, 1936, 1937, 1938, 1940, 1941, 1942, 1943, 1944, 1945, 1947
	Southern Directory Co.	1	1887
	Walker, Evans and Cogswell (Charleston, SC)	1	1890
Asheville and Buncombe County	Baughman Brothers	1	1883
Burlington (inc Graham, Haw River and Elon College)	Hill Directory Co.	3	1935, 1943, 1946
Burlington, Graham and Haw River	Miller-Commercial Directory Co.	4	1920, 1924, 1927, 1929
	Miller-Piedmont Dir.	1	1909

Canton	Miller (Chas)-Southern Directory Co.	2	1937, 1942
Charlotte	Beasley and Emerson	1	1875
	Hill Directory Co.	15	1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1947
	Miller-Commercial Directory Co.	11	1916, 1917, 1918, 1920, 1923, 1925, 1926, 1927, 1928, 1929, 1930
	Miller-Piedmont Directory Co.	6	1911, 1912, 1913, 1914, 1915, 1931
	Walsh Directory Co. (Charleston, SC)	4	1905, 1907, 1909, 1910
Concord	Baldwin Dir. Co.	2	1938, 1940
	Interstate Directory Co. (Charlotte, NC)	1	1902
	Miller-Commercial Directory Co.	2	1920, 1922
	Miller-Piedmont Directory Co.	3	1908, 1913, 1916
Dunn	Chas S. Gardiner	1	1918
	Miller-Commercial Directory Co.	1	1926
Durham	E. F. Turner and Co.	1	1889
	Hill Directory Co.	26	1903, 1905, 1907, 1915, 1919, 1923-1936, 1938, 1939, 1941, 1942, 1943, 1947, 1949
	Levi Branson (Raleigh, NC)	1	1887
	N.A. Ramsey (Durham, NC)	1	1892
	Samuel L. Adams (Durham, NC)	1	1902
	Seeman Printery (Durham, NC)	1	1911
	The Educator Company (Durham, NC)	1	1897
Elizabeth City	Miller (Chas)-Southern Directory Co.	3	1936, 1938, 1942
	Miller-Commercial Directory Co.	1	1923
Fayetteville	Chas S. Gardiner	1	1915
Gastonia	Hill Directory Co.	4	1934, 1936, 1942, 1947
	Miller-Commercial Directory Co.	4	1918, 1921, 1923, 1930
	Miller-Piedmont Dir.	2	1910, 1913

Goldsboro	Baldwin Dir. Co.	1	1938
	Chas S. Gardiner	1	1916
	Hill Directory Co.	1	1934
Greensboro	Chase Brenizer	1	1896
	E.F. Turner and Co.	1	1890
	Hill Directory Co.	49	1904, 1905, 1907, 1912, 1913, 1915, 1917, 1918, 1920-1931, 1933-1943, 1945-1947, 1949-1951, 1953-1955-1963
	Maloney Dir. Co.	2	1899, 1901
Greensboro, Reidsville	Thompson, Breed and Crofill (Newburgh, NY)	1	1887
Greensboro, Salem and Winston	Interstate Directory Co. (Atlanta, GA)	1	1884
Greenville	Miller (Chas)-Southern Directory Co.	6	1936, 1938, 1940, 1942, 1944, 1947
	Miller-Commercial Directory Co.	1	1926
	Miller-Commercial Directory Co. (Piedmont)	1	1916
Henderson	Miller (Chas)-Southern Directory Co.	4	1938, 1940, 1942, 1947
Hendersonville	I. E. Maxwell (Hendersonville, NC)	1	1915
	Miller (Chas)-Southern Directory Co.	5	1937, 1939, 1941, 1943, 1945
	Miller-Commercial Directory Co.	3	1921, 1924, 1926
Hickory	Brady Printing Co.	1	1920
	Home Directory Co. (Hickory, NC)	1	1935
	Miller (Chas)-Southern Directory Co.	3	1937, 1943, 1947
	Miller-Commercial Directory Co.	1	1930
High Point	Hill Directory Co.	4	1933, 1940, 1942, 1948
	Miller-Commercial Directory Co.	8	1919, 1921, 1923, 1925, 1927, 1928, 1929, 1930
	Miller-Piedmont Directory Co.	3	1908, 1910, 1913
Kinston	Baldwin Dir. Co.	1	1936
	Hill Directory Co.	2	1923, 1928

Lenoir	Miller Press	1	1939
	Miller-Commercial Directory Co.	2	1930, 1937
	Miller-Southern Directory Co.	2	1943, 1945
Lexington	Baldwin Dir. Co.	1	1937
	Chas S. Gardiner	1	1916
Lumberton	Baldwin Dir. Co.	1	1938
	Chas S. Gardiner	1	1916
Monroe	Miller-Commercial Directory Co.	1	1922
Mooresville	Miller (Chas)-Southern Directory Co.	1	1939
Morganton	Miller (Chas)-Southern Directory Co.	3	1931, 1939, 1943
Mount Airy	J. Edwin Carter and A. Kyle Sydnor	1	1913
	Miller-Commercial Directory Co.	1	1928
New Bern	Baldwin Dir. Co.	1	1937
	Chas S. Gardiner	1	1916
	Edwards, Broughton and Co.	2	1880, 1893
	Hill Directory Co.	3	1907, 1918, 1920
North Wilkesboro, Wilkesboro	Miller (Chas)-Southern Directory Co.	1	1939
Oxford	Miller-Commercial Directory Co.	1	1929
Raleigh	Edwards, Broughton and Co.	3	1880, 1883, 1887
	Hill Directory Co.	35	1903, 1905, 1907, 1909, 1911, 1913, 1915, 1917-1919, 1921-1943, 1945, 1948
	J. H. Chataigne (Raleigh, NC)	1	1875
	Maloney Dir. Co.	1	1899
	Maloney Directory Co.	1	1901
	Observer Printing Co (Raleigh, NC)	1	1888
	Raleigh Stationery Company	1	1896
	Turner, M'Lean and Losee Directory Co. (Raleigh, NC)	1	1886

Reidsville	Baldwin Directory Co.	1	1935
	Hill Directory Co.	1	1932
	Miller-Commercial Directory Co.	1	1929
Richmond County	Page Trust Co.	1	1916
Roanoke Rapids	Miller (Chas)-Southern Directory Co.	2	1938, 1942
Rocky Mount	Hill Directory Co.	10	1908, 1912, 1914, 1920, 1930, 1934, 1936, 1938, 1940, 1942
Salisbury	Baldwin Directory Co.	3	1935, 1938, 1940
Salisbury-Spencer	Miller-Commercial Directory Co.	4	1917, 1919, 1922, 1924
	Miller-Piedmont Directory Co.	2	1913, 1928
	Miller-Piedmont Directory Co.-Hackney and Moale	2	1907, 1915
Shelby	Miller (Chas)-Southern Directory Co.	6	1937, 1939, 1941, 1943, 1945, 1947
Statesville	Carolina Directory Co.	1	1932
	Miller (Chas)-Southern Directory Co.	3	1938, 1940, 1944
	Miller-Commercial Directory Co.	4	1922, 1925, 1928, 1930
	Miller-Piedmont Directory Co.	1	1916
Statesville and Iredell County	Miller-Piedmont Directory Co.-Hackney and Moale	1	1909
Thomasville	Baldwin Dir. Co.	1	1935
	Carolina Dir. Co.	1	1933
	Miller-Commercial Directory Co.	1	1930
Washington	Baldwin Dir. Co.	1	1937
	Miller-Commercial Directory Co. (Piedmont)	1	1916
Waynesville and Haywood County	Hackney and Moale	1	1906
Wilmington	Benj. R. Sherriff	1	1877
	Frank D. Smaw	1	1866
	Geo. H. Kelley	1	1860

	Hill Directory Co.	21	1897, 1905, 1907, 1909, 1911, 1913, 1917-1919, 1922, 1924, 1926, 1928, 1930, 1932, 1934, 1940, 1943, 1947
	P. Heinsberger (Wilmington, NC)	2	1865, 1871
	Wilmington Messenger/Messenger Steam Presses	1	1889
Wilson	Hill Directory Co.	3	1912, 1916, 1930
Winston-Salem	Hill Directory Co.	11	1932-1939, 1943, 1946, 1947
	Miller-Commercial Directory Co.	5	1916, 1921, 1922, 1930, 1932
	Miller-Commercial Directory Co.-Hackney and Moale	1	1918
	Miller-Piedmont Directory Co.	3	1910, 1912, 1913
	Miller-Piedmont Directory Co.-Hackney and Moale	2	1911, 1915
	Walsh Directory Co. (Charleston, SC)	1	1908
Winston-Salem, Greensboro	Edwards, Broughton and Co.	1	1879

There does not appear to be a strong correlation between the size or location of the town

and its range of publishers. Charlotte, for instance, surpassed Wilmington as the largest

N.C. city in the early part of the twentieth century (Hanchett, 1998), and yet Charlotte has

only slightly fewer total publishers than smaller towns, such as Raleigh, Winston-Salem,

and Durham. And Greensboro, which has far more directories than any other location,

has only four major publishers. This is a reminder that the digitized collection available at

IA is not necessarily representative of the entire extant historical record, and the nature of

the digital collection may change significantly as more directories come online.

Discussion

It is feasible to use patterns of attributes (Appendix 2) as the basis for revising and refining the original Python parser to capture a far wider swath of the collection than just the test bed directory. Creating data visualizations based on this set of attributes was stymied by the disproportionate amount of nominal data and the overwhelming lack of ratio data. It is nonetheless possible to draw some general conclusions based on an analysis of the attributes for the top publishers. These conclusions, grounded in historical variation, can inform the creation of a more nuanced parser that can be adapted to a wider range of directories across the state.

Perhaps the most significant conclusion drawn from collecting the publishing attributes for these 415 directories is that there are some major groupings that could be used to fine tune the parser. Hill Directory Co, responsible for 45.54% (189 of 415) of the collection, can be divided into three major chronological categories for the purposes of parsing data (Table 6).

Table 6. Breakdown of Hill directories. N=189

Year Range	Count	Percentage of Hill Directories
1897-1926	64	33.86%
1927-1952	113	59.79%
1953-1963	12	6.35%

The 1953-1963 range only covers Greensboro directories, for which there is a high degree of consistency (Table 9). It is uncertain how much this portion of the Hill collection will grow in the future. Such growth will depend on the willingness of

digitizing sponsors to extend their collections beyond the 1940s, which is the temporal endpoint for most of the directories in the overall collection. The University of North Carolina at Greensboro, more than other institutions, has been scanning later volumes; hence Greensboro is likely disproportionally represented in the digital collection.

The earlier Hill directories, like other earlier publications, are rather inconsistent, with many exceptions to the common attributes (Table 7). The bulk of the collection (1927-1952) also contains many exceptions, though there appears a measure of consistency among those exceptions (Table 8). It is conceivable that several versions of the parser could be built for Hill based on a combination of location and year. For instance, there is remarkable consistency across the post-1939 Charlotte and post 1940-Greensboro directory exceptions.

Table 7. Hill Directory Co. Part I. 1897-1926. N=64

Attribute	Common Attribute	Exception
Year range	1897-1963	
City range	Map[18]	
Directories appearing before the general directory	A range of one or several directories including: Business Directory or Buyer's Guide Street Directory or Street Guide Miscellaneous	New Bern 1907 has nothing before general directory
Beginning page of general directory	Ranges from as low as page 45 to as high as page 213, with the bulk falling in the 100s	
Separate colored directory	No	
Street directory included	Y, though those for Durham, Rocky Mount and Wilmington frequently do not provide household listings. Many have 2 columns. When the header differs from general directory in later	None for New Bern (1907-1908, 1918-1909, 1920-1921) Kinston (1923-1924) Street Directories with 1 col:

---

[18] Interactive map available at http://batchgeo.com/map/c0c7840e0391e6c99a27915f45452633.

	years, it is typically: first appearing street name page number last name on the page	Rocky Mount (1908-1909, 1912-1915, 1919-1920) Wilmington (1897, 1905, 1907, 1909-1914) Wilson (1912-1913, 1916-1917)
Listing of abbreviations	All include general and job abbreviations but nothing else	
Title of general directory	All include "City Directory" in title	All directories between 1921 and 1926 also have "Hill Directory Co."
Racial designation	*	
Married status indicated	No, except when listing a widow's husband's name in parenthesis	Greensboro (1917, 1920) includes marital status
Number of columns	1	The following have 2 columns: Wilmington 1897 Durham 1925 Greensboro 1926
Header or footer	Header	
Even Header	1897-1910: First alpha name page number last alpha name on page 1911-1921: page number city name year directory 1922-1926: includes "Hill Directory Co"	
Odd Header	1897-1910: First alpha name page number last alpha name on page 1911-1921: city name year Directory page number 1922-1926: city name City Directory (year) page number	
Repeating last names	Names repeat in full	
Other symbols	When employed, the following symbols mean: h – head of household or resides at rms – rooms bds – boarder dash used when building number is missing	

Table 8. Hill Directory Co. Part II. 1927-1952. N=113

Attribute	Common Attribute	Exception
Year range	1897-1963	
City range	Map	
Directories appearing before the general directory	Either none or a Buyer's Guide of advertisements (not a real directory)	

Beginning page of general directory	Ranges from 17 to 93, with the bulk falling somewhere between pages 21 and 33	Greensboro 1927 starts on page 105
Separate colored directory	No	High Point 1933
Street directory included	Yes – all with 2 columns	The following Street Directories have 3 columns: Charlotte (1939 onward) Greensboro (1940-1950) Raleigh (1948) Wilmington (1947) Winston-Salem (1946-1948)
Listing of abbreviations	In addition to general/job abbrev., all directories after 1931 include name abbrev. About 27% include special abbrev. A handful include abbrev for firms, streets and suburbs, and occupants.	
Title of general directory	All include "Hill's and "City Directory"	
Racial designation	(c)	High Point 1940 and 1948 do not include parenthesis.
Married status indicated	Y	Number of children indicated for Greensboro (1935-1943)
Number of columns	1	The following have 2 columns: Charlotte (1932-1947) Durham (1941-1949) Greensboro (1927-1952) Raleigh (1948) Wilmington (1947) Winston-Salem (1932-1939, 1943, 1946-1948)
Header or footer	Header through 1930, none from 1931-1951. Where no header appears, the page number is centered at top of page.	The following have header: Burlington (1934) Goldsboro (1935) High Point (1933) Reidsville (1932)
Even Header	1927-1930: includes "Hill Directory Co." Where headers exist (1932-1935): includes "Hill Directory Co"	
Odd Header	1927-1930: city name City Directory (year) page number 1932-1935 where applicable: city name City Directory (year) page number	High Point (1933): city name [race] population (year)
Repeating last names	1927-1940 names repeat in full	*1927-1940 exceptions (use "):* Charlotte (1934-1939) Winston Salem (1932-1939)

Even Header	N/A	
Odd Header	N/A	
Repeating last names	Name cont. w. --	Name cont. w. " for Greensboro 1961, 1962
Other symbols	When employed, the following symbols mean: h – head of household o in circle – homeowner	

Given the total number of publishers and the variety within publisher groupings, it is not feasible for the parser to rely solely on publisher clusters. The number of exceptions, even among the major publishers, suggests that constructing the parser around publisher groupings will probably not capture enough of the collection. This approach, however, does suggest a more fluid solution. Rather than build a series of parsers around publishers, it might be more efficient to build a skeletal parser accompanied by an interactive system for setting parameters depending on the publisher's attributes (or exceptions). With the proper interface built over the parser, a user could feed in the specific parameters based on the attributes described here. These parameters could be input into the parser behind the scenes, making for an agile and extensible parser in the end.

Another approach for building the parser would be to combine attributes from different publishers, and use the complete list of directories (Appendix 1) as a lookup mechanism. In this implementation, a user need only select a particular directory, and the system would apply the closest parser based on the directory's corresponding attributes. In order for this to work even moderately well, the attributes would have to be prioritized, so that a higher ranked attribute would be weighed more heavily when applying a parser. Number of columns and header formatting should be treated with the most importance, since these are the two aspects of the directory most likely to trip the parser. Secondly,

race should be weighed heavily, as there are only eight variations for its depiction (Table 1), and since it can prove critical for analytical purposes.

The column structure is perhaps the easier of the two top-tier attributes to handle, as the default for most directories is one column for the general alphabetical directory; the major exceptions are larger cities and later years (e.g. nineteenth-century Wilmington, Charlotte since the 1920s, and post-1953 Greensboro). Likewise, the wording in headers is absolutely critical for guiding the parser and eliminating noise. In nearly all cases where a header exists, there are only a few variations of items that even- and odd-paged headers contain in some configuration or another: publisher's name (typically even only); city name, publication year, the word "Directory" or "City Directory," or the first/last name on the page (or the first three letters of the names). Page numbers are almost always included in the header. An iterator in the parser could be used to detect the directory page number in the header (or at the top of the page where no header exists and where the page number is visible in the scan). Likewise, year could possibly be plugged into the parser for detection purposes. This is particularly the case for directories that fall on the lower and upper bound of the temporal range (an assessment of most of the nineteenth-century directories is not covered here).

Ultimately, a combination of these two approaches might prove most successful. Returning to the columns attribute, those directories with two columns tend to be clustered into a few publisher groupings: the Miller group (5 instances, all for Charlotte between 1926 and 1930), the Miller-Piedmont series (1 occurrence for Charlotte in 1931), and Hill Directory Co.: three instances for the 1897-1926 grouping (Wilmington, Durham and Greensboro); multiple occurrences for 1927-1952 (Charlotte 1932-1947,

Durham 1941-1949, and Greensboro 1927-1952). Finally all of the Greensboro directories after 1953 have two columns. A convergence between location, year, and publisher appears appropriate for this particular attribute. This suggests, then, that the most agile parser is one that is responsive to particular attributes; for some, sticking with the publisher is the sound approach, while for other attributes, it might make more sense to take a spatial or spatial-temporal approach.

Limitations

It is uncertain how well any of these approaches will work until the parser can be adapted and tested, at least on a subset of the entire N.C. city directory collection. To that end, it is unclear to what degree, if any, this parser will scale *beyond* North Carolina. It is highly likely that a similar process of attribute collection will be necessary for another state/region (perhaps a subset of such directories can be assessed as the project grows). This will first require a comprehensive inventory of digitized directories, since the Internet Archive's collection is incomplete. As of 27 June 2012, there were only approximately 839 directories spanning cities across the county. Nineteen states had no directories whatsoever hosted on IA.[19] The attribute collection process is time consuming and, as yet, still one that must be completed manually. While it is acceptable to employ a large amount of manual hours on the front end if it facilitates faster processing and automation for others, figuring out a way to at least partially automate or crowdsource attribute coding will be critical for scaling up to a nationwide city directory collection. One bright spot for scaling up, though, is that many of these publishers, including Hill and Miller/Piedmont, reached well beyond North Carolina. Piedmont, for instance,

---

[19] The total number of directories is questionable, as searches for directories yielded many irrelevant results, including state-wide gazetteers, government directories, and telephone directories.

covered cities in Alabama, Georgia, Illinois, Kentucky, Louisiana, Mississippi, Missouri, South Carolina, Tennessee, Texas and Virginia in 1903; Hill covered many more cities in many of the same states, particularly towards the mid-twentieth century.[20]

Moreover, it is uncertain how well the N.C. parser will do given that only the major publishers' attributes were taken into consideration here. There is a strong logic for ignoring smaller publishers with only a few directories. Yet as the number of digitized directories continues to increase, the composition of publishers may change in unexpected ways. For instance, the Internet Archive might see an influx of directories from the Carolina Directory Co., which in this collection only represent two out of 415 items. There is no way to predict how many towns and for how many years this publisher operated short of tracking the company's own paper trail.

Yet another limitation of this approach is its lack of attention to advertisements. While ads can be processed with OCR, the DIL parser is not being trained to look at ads, since they are difficult to parse. Yet advertisements comprise an important part of the historical record, particularly since these directories were marketed aggressively to business owners as a way to boost their client base. Not only are ads very informative and, in some cases, quite amusing, they are also important windows into everyday life. They can help fill in gaps that other content in the city directories cannot provide.[21]

One final limitation of this approach, though perhaps less pressing, is that the parser does not yet account for street directory listings. The street directory proved essential to the manual mapping of Charlotte 1911; the 4,000 markers were initially

---

[20] Based on the "Directory of Publishers" included in many Piedmont and Hill publications. See, for instance, Miller-Piedmont's Concord 1913 directory and Hill's Greensboro directory for 1958.
[21] Many cultural historians and media critics have produced important analyses and readings (particularly gendered readings) of advertisements. See, for instance, Peiss (1998).

culled from the street index because it was easier to delimit the geographic boundaries of the dataset. Had there not been a street index, the team would have been forced to scan through the entire alphabetical listing and pull out those entries falling within the project's geographic borders. The street directory entries were then cross-referenced with the complete listings in the general alphabetical directory to capture spouse and occupation. It is probable that the final parser will not need to iterate through the street directory because its output will enable searching and sorting by street from the data generated with the alphabetical listing. However, there may be other uses of the street directory that have not yet been predicted. What is the cost of ignoring this section of the directory?

Future Steps and Implications

This project proposes several theoretical approaches to improving the city directory parser. The next step is to implement some of these ideas, test iteratively on increasingly larger subsets of the collection, and measure the success rate. Once satisfied that the parser performs relatively well on a subset, it will be tested on the entire N.C. collection. And once success rates reach about 90%, work can begin to generalize beyond the N.C. directories.

The DIL envisions developing a simple but powerful interface tool for parsing such data, which could be created in tandem with parser refinement. This web interface would sit over the Python parser, with a corresponding list of attributes and/or some sort of authority file/s on the backend. Users would input information about a directory and, based on the parameters, the program would run the appropriate script. This would give

other scholars, students, and the public greater access to the content already available on Internet Archive.

Finally, and perhaps most ambitiously, the DIL would like to automate the spatialization of city directory data by batch geo-locating addresses. The process of automatically assigning reasonably accurate latitude and longitude, as with the rest of this project, comes up against inconsistencies in the historical record. Many towns experienced street renumbering and/or redevelopment through programs such as Urban Renewal. Many of the addresses in these directories simply no longer exist. Furthermore, some of the addresses listed in these directories may be completely fabricated. Rose-Redwood reports that "sometimes city directory publishers … took it upon themselves to number the buildings without getting an official stamp of approval from the city government," as in the case of Mobile, Alabama in 1837 (p. 294). Though it is possible to use historical maps, particularly georectified Sanborn maps, to correct for this, the Sanborns only cover central business districts and their environs. Such an undertaking requires historical problem-solving in combination with computational and GIS methodologies.

As this project demonstrates, historical data sets cannot be treated as pure data. They require an understanding of the nuances and peculiarities always found in the historical record. By combining historical thinking with computational approaches, digital humanists can build smarter tools with wider reaches, thereby making our cultural heritage more accessible to more people.

Bibliography

AddressingHistory Project. Retrieved from http://addressinghistory.edina.ac.uk/. Accompanying blog retrieved from http://addressinghistory.blogs.edina.ac.uk/blog/. Project updates retrieved from http://addressinghistory.blogs.edina.ac.uk/category/project-updates/.

Bontcheva, K., Maynard, D., Cunningham, H., & Saggion, H. (2002). Using human language technology for automatic annotation and indexing of digital library content. In M. Agosti and C. Thanos (Eds.), *Research and Advanced Technology for Digital Libraries, Proceedings of the 6th European Conference, ECDL 2002, Rome, Italy,* 141-154. doi: 10.1007/3-540-45747-X_46

Cohen, D. (2006). From Babel to knowledge: data mining large digital collections. *D-Lib Magazine* Volume 12, Number 3 (March, 2006): 6-19 and is reprinted here with permission. Retrieved from http://chnm.gmu.edu/essays-on-history-new-media/essays/?essayid=40.

Digging into Data Challenge Website, http://www.diggingintodata.com/.

Goldstein, S. (1954). City directories as sources of migration data. *American Journal of Sociology* 60, no. 2, 169-176.

Hanchett, T. (1998). *Sorting out the new south city: race, class, and urban development in Charlotte, 1875-1975*. Chapel Hill: UNC Press.

Hershberg, T. (1978). The new urban history: 'toward an interdisciplinary history of the city.' *Journal of Urban History* 5, no. 1, 3-40.

Hundt, M., Denison, D., & Schneider, G. (2012). Retrieving relatives from historical data. *Literary and Linguistic Computing* 27, no. 1, 3-16. doi: 10.1093/llc/fqr049

Liu, A. (2012). The state of the digital humanities: a report and a critique. *Arts and Humanities in Higher Education* 11, 1-2, 8-41. doi:10.1177/1474022211427364

Lutz, M. (2009). *Learning Python*, 4th edition. Sebastopol, California: O'Reilly.

Peiss, K. (1998). *Hope in a jar: the making of America's beauty culture*. New York: Metropolitan Books.

Pekar, V. & Evans, R. (2007). Discovery of language resources on the web: information extraction from heterogeneous documents. *Literary and Linguistic Computing* 22, no. 3, 329-343. doi: 10.1093/llc/fqm010

Rose-Redwood, R.S. (2008). Indexing the great ledger of the community: urban house numbering, city directories, and the production of spatial legibility. *Journal of Historical Geography* 34, 286-310. doi: 10.1016/j.jhg.2007.06.003

Sculley, D. & Pasanek, B. (2008). Meaning and mining: the impact of implicit assumptions in data mining for the humanities. *Literary and Linguistic Computing* 23, no. 4, 409-424. doi: 10.1093/llc/fqn019

Stuart Macdonald. (2011, August 22). AddressingHistory Phase 2 [Web log post]. Retrieved from http://addressinghistory.blogs.edina.ac.uk/2011/08/22/addressinghistory-phase-2/.

Thernstrom, S. (1971). Reflections on the new urban history. *Dædalus* 100, no. 2, 359-375.

Williford, C., Henry, C., & Friedlander, A. (June 2012). One culture: computationally intensive research in the humanities and social sciences. A report on the experiences of the first respondents to the Digging into Data Challenge. Retrieved from the Council on Library and Information Resources website on 22 June 2012: http://www.clir.org/pubs/reports/pub151/pub151.pdf.

Appendix 1. North Carolina City Directories
The complete list of N.C. directories hosted on Internet Archive as of 2 May 2012.
N=415

Albermarle (1937), Baldwin Directory Co.

Asheboro (1937-1938), Miller (Chas)-Southern Directory Co.
Asheboro (1939-1940), Miller (Chas)-Southern Directory Co.
Asheboro (1941-1942), Miller (Chas)-Southern Directory Co.
Asheboro (1947-1948), Miller (Chas)-Southern Directory Co.

Asheville (1887), Southern Directory Co.
Asheville (1890), Walker, Evans and Cogswell (Charleston, SC)
Asheville (1896-1897), Franklin Printing and Publishing Co., J.S. McIlwaine, publisher
Asheville (1899-1900), Maloney Directory Co.
Asheville (1900-1901), Maloney Directory Co.
Asheville (1906-1907), Hill Directory Co.
Asheville (1909), Miller-Piedmont Directory Co.
Asheville (1910), Miller-Piedmont Directory Co.
Asheville (1912), Miller-Piedmont Directory Co.
Asheville (1913), Miller-Piedmont Directory Co.
Asheville (1914), Miller-Piedmont Directory Co.
Asheville (1916), Miller-Piedmont Directory Co.
Asheville (1917), Miller-Commercial Directory Co.
Asheville (1918), Miller-Commercial Directory Co.
Asheville (1922), Miller-Commercial Directory Co.
Asheville (1930), Miller-Commercial Directory Co.
Asheville (1931), Miller-Piedmont Directory Co.
Asheville (1935), Miller-Piedmont Directory Co.
Asheville (1936), Miller-Piedmont Directory Co.
Asheville (1937), Miller-Piedmont Directory Co.
Asheville (1938), Miller-Piedmont Directory Co.
Asheville (1939), Miller-Piedmont Directory Co.
Asheville (1940), Miller-Piedmont Directory Co.
Asheville (1941), Miller-Piedmont Directory Co.
Asheville (1942), Miller-Piedmont Directory Co.
Asheville (1943), Miller-Piedmont Directory Co.
Asheville (1944), Miller-Piedmont Directory Co.
Asheville (1945-1946), Miller-Piedmont Directory Co.
Asheville (1947), Miller-Piedmont Directory Co.

Asheville and Buncombe County (1883-1884), Baughman Brothers

Burlington (inc Graham, Haw River and Elon College) (1935), Hill Directory Co.
Burlington (inc Graham, Haw River and Elon College) (1943), Hill Directory Co.

Burlington (inc Graham, Haw River and Elon College) (1946), Hill Directory Co.

Burlington, Graham and Haw River (1909-1910), Miller-Piedmont Directory Co.
Burlington, Graham and Haw River (1920-1921), Miller-Commercial Directory Co.
Burlington, Graham and Haw River (1924-1925), Miller-Commercial Directory Co.
Burlington, Graham and Haw River (1927-1928), Miller-Commercial Directory Co.
Burlington, Graham and Haw River (1929-1930), Miller-Commercial Directory Co.

Canton (1937-1938), Miller (Chas)-Southern Directory Co.
Canton (1942-1943), Miller (Chas)-Southern Directory Co.

Charlotte (1875-1876), Beasley and Emerson
Charlotte (1905-06), Walsh Directory Co. (Charleston, SC)
Charlotte (1907), Walsh Directory Co. (Charleston, SC)
Charlotte (1909), Walsh Directory Co. (Charleston, SC)
Charlotte (1910), Walsh Directory Co. (Charleston, SC)
Charlotte (1911), Miller-Piedmont Directory Co.
Charlotte (1912), Miller-Piedmont Directory Co.
Charlotte (1913), Miller-Piedmont Directory Co.
Charlotte (1914), Miller-Piedmont Directory Co.
Charlotte (1915), Miller-Piedmont Directory Co.
Charlotte (1916), Miller-Commercial Directory Co.
Charlotte (1917), Miller-Commercial Directory Co.
Charlotte (1918), Miller-Commercial Directory Co.
Charlotte (1920), Miller-Commercial Directory Co.
Charlotte (1923-1924), Miller-Commercial Directory Co.
Charlotte (1925), Miller-Commercial Directory Co.
Charlotte (1926), Miller-Commercial Directory Co.
Charlotte (1927), Miller-Commercial Directory Co.
Charlotte (1928), Miller-Commercial Directory Co.
Charlotte (1929), Miller-Commercial Directory Co.
Charlotte (1930), Miller-Commercial Directory Co.
Charlotte (1931), Miller-Piedmont Directory Co.
Charlotte (1932), Hill Directory Co.
Charlotte (1933), Hill Directory Co.
Charlotte (1934), Hill Directory Co.
Charlotte (1935), Hill Directory Co.
Charlotte (1936), Hill Directory Co.
Charlotte (1937), Hill Directory Co.
Charlotte (1938), Hill Directory Co.
Charlotte (1939), Hill Directory Co.
Charlotte (1940), Hill Directory Co.
Charlotte (1941), Hill Directory Co.
Charlotte (1942), Hill Directory Co.
Charlotte (1943), Hill Directory Co.
Charlotte (1944), Hill Directory Co.

Charlotte (1945-1946), Hill Directory Co.
Charlotte (1947), Hill Directory Co.

Concord (1902), Interstate Directory Co. (Charlotte, NC)
Concord (1908), Miller-Piedmont Directory Co.
Concord (1913-1914), Miller-Piedmont Directory Co.
Concord (1916-1917), Miller-Piedmont Directory Co.
Concord (1920-1921), Miller-Commercial Directory Co.
Concord (1922-1923), Miller-Commercial Directory Co.
Concord (1938), Baldwin Directory Co.
Concord (1940), Baldwin Directory Co.

Dunn (1918-1919), Chas S. Gardiner
Dunn (1926-1927), Miller-Commercial Directory Co.

Durham (1887), Levi Branson (Raleigh, NC)
Durham (1889-1890), E. F. Turner and Co.
Durham (1892), N.A. Ramsey (Durham, NC)
Durham (1897), The Educator Company (Durham, NC)
Durham (1902), Samuel L. Adams (Durham, NC)
Durham (1903-1904), Hill Directory Co.
Durham (1905-1906), Hill Directory Co.
Durham (1907-1908), Hill Directory Co.
Durham (1911-1912), Seeman Printery (Durham, NC)
Durham (1915-1916), Hill Directory Co.
Durham (1919-1920), Hill Directory Co.
Durham (1923), Hill Directory Co.
Durham (1924), Hill Directory Co.
Durham (1925), Hill Directory Co.
Durham (1926), Hill Directory Co.
Durham (1927), Hill Directory Co.
Durham (1928), Hill Directory Co.
Durham (1929), Hill Directory Co.
Durham (1930), Hill Directory Co.
Durham (1931), Hill Directory Co.
Durham (1932), Hill Directory Co.
Durham (1933), Hill Directory Co.
Durham (1934), Hill Directory Co.
Durham (1935), Hill Directory Co.
Durham (1936-1937), Hill Directory Co.
Durham (1938), Hill Directory Co.
Durham (1939), Hill Directory Co.
Durham (1941), Hill Directory Co.
Durham (1942), Hill Directory Co.
Durham (1943), Hill Directory Co.
Durham (1947), Hill Directory Co.

Durham (1949), Hill Directory Co.

Elizabeth City (1923-1924), Miller-Commercial Directory Co.
Elizabeth City (1936-1937), Miller (Chas)-Southern Directory Co.
Elizabeth City (1938-1939), Miller (Chas)-Southern Directory Co.
Elizabeth City (1942-1943), Miller (Chas)-Southern Directory Co.

Fayetteville (1915-1916), Chas S. Gardiner

Gastonia (1910-1911), Miller-Piedmont Directory Co.
Gastonia (1913-1914), Miller-Piedmont Directory Co.
Gastonia (1918-1919), Miller-Commercial Directory Co.
Gastonia (1921-1922), Miller-Commercial Directory Co.
Gastonia (1923-1924), Miller-Commercial Directory Co.
Gastonia (1930-1931), Miller-Commercial Directory Co.
Gastonia (1934), Hill Directory Co.
Gastonia (1936), Hill Directory Co.
Gastonia (1942), Hill Directory Co.
Gastonia (1947), Hill Directory Co.

Goldsboro (1916-1917), Chas S. Gardiner
Goldsboro (1934), Hill Directory Co.
Goldsboro (1938), Baldwin Directory Co.

Greensboro (1890-1891), E. F. Turner and Co.
Greensboro (1892-1893), Stone and Kendall
Greensboro (1896-1897), Chase Brenizer
Greensboro (1899-1900), Maloney Directory Co.
Greensboro (1901), Maloney Directory Co.
Greensboro (1903-1904), Hill Directory Co.
Greensboro (1905-1906), Hill Directory Co.
Greensboro (1907-1908), Hill Directory Co.
Greensboro (1909-1910), Hill Directory Co.
Greensboro (1912-1913), Hill Directory Co.
Greensboro (1913-1914), Hill Directory Co.
Greensboro (1915-1916), Hill Directory Co.
Greensboro (1917), Hill Directory Co.
Greensboro (1918-1919), Hill Directory Co.
Greensboro (1920), Hill Directory Co.
Greensboro (1921), Hill Directory Co.
Greensboro (1922), Hill Directory Co.
Greensboro (1923), Hill Directory Co.
Greensboro (1924), Hill Directory Co.
Greensboro (1925), Hill Directory Co.
Greensboro (1926), Hill Directory Co.
Greensboro (1927), Hill Directory Co.

Greensboro (1928), Hill Directory Co.
Greensboro (1929), Hill Directory Co.
Greensboro (1930), Hill Directory Co.
Greensboro (1931), Hill Directory Co.
Greensboro (1933), Hill Directory Co.
Greensboro (1934), Hill Directory Co.
Greensboro (1935), Hill Directory Co.
Greensboro (1936), Hill Directory Co.
Greensboro (1937), Hill Directory Co.
Greensboro (1938), Hill Directory Co.
Greensboro (1939), Hill Directory Co.
Greensboro (1940), Hill Directory Co.
Greensboro (1941), Hill Directory Co.
Greensboro (1942), Hill Directory Co.
Greensboro (1943), Hill Directory Co.
Greensboro (1945), Hill Directory Co.
Greensboro (1946), Hill Directory Co.
Greensboro (1947-1948), Hill Directory Co.
Greensboro (1949), Hill Directory Co.
Greensboro (1950), Hill Directory Co.
Greensboro (1951-1952), Hill Directory Co.
Greensboro (1953), Hill Directory Co.
Greensboro (1954), Hill Directory Co.
Greensboro (1955), Hill Directory Co.
Greensboro (1956), Hill Directory Co.
Greensboro (1957), Hill Directory Co.
Greensboro (1958), Hill Directory Co.
Greensboro (1959), Hill Directory Co.
Greensboro (1960), Hill Directory Co.
Greensboro (1961), Hill Directory Co.
Greensboro (1962), Hill Directory Co.
Greensboro (1963), Hill Directory Co.

Greensboro, Reidsville (1887), Thompson, Breed and Crofill (Newburgh, NY)

Greensboro, Salem and Winston (1884), Interstate Directory Co. (Atlanta, GA)

Greenville (1916-1917), Miller-Commercial Directory Co. (Piedmont)
Greenville (1926-1927), Miller-Commercial Directory Co.
Greenville (1936-1937), Miller (Chas)-Southern Directory Co.
Greenville (1938-1939), Miller (Chas)-Southern Directory Co.
Greenville (1940-1941), Miller (Chas)-Southern Directory Co.
Greenville (1942-1943), Miller (Chas)-Southern Directory Co.
Greenville (1944-1945), Miller (Chas)-Southern Directory Co.
Greenville (1947-1948), Miller (Chas)-Southern Directory Co.

Henderson (1938-1939), Miller (Chas)-Southern Directory Co.
Henderson (1940-1941), Miller (Chas)-Southern Directory Co.
Henderson (1942-1943), Miller (Chas)-Southern Directory Co.
Henderson (1947-1948), Miller (Chas)-Southern Directory Co.

Hendersonville (1915), I. E. Maxwell (Hendersonville, NC)
Hendersonville (1921-1922), Miller-Commercial Directory Co.
Hendersonville (1924-1925), Miller-Commercial Directory Co.
Hendersonville (1926-1927), Miller-Commercial Directory Co.
Hendersonville (1937-1938), Miller (Chas)-Southern Directory Co.
Hendersonville (1939-1940), Miller (Chas)-Southern Directory Co.
Hendersonville (1941-1942), Miller (Chas)-Southern Directory Co.
Hendersonville (1943-1944), Miller (Chas)-Southern Directory Co.
Hendersonville (1945-1946), Miller (Chas)-Southern Directory Co.

Hickory (1920-1921), Brady Printing Co.
Hickory (1930-1931), Miller-Commercial Directory Co.
Hickory (1935), Home Directory Co. (Hickory, NC)
Hickory (1937-1938), Miller (Chas)-Southern Directory Co.
Hickory (1943-1944), Miller (Chas)-Southern Directory Co.
Hickory (1947-1948), Miller (Chas)-Southern Directory Co.

High Point (1908), Miller-Piedmont Directory Co.
High Point (1910-1911), Miller-Piedmont Directory Co.
High Point (1913), Miller-Piedmont Directory Co.
High Point (1919), Miller-Commercial Directory Co.
High Point (1921-1922), Miller-Commercial Directory Co.
High Point (1923-1924), Miller-Commercial Directory Co.
High Point (1925-1926), Miller-Commercial Directory Co.
High Point (1927), Miller-Commercial Directory Co.
High Point (1928-1929), Miller-Commercial Directory Co.
High Point (1929-1930), Miller-Commercial Directory Co.
High Point (1930-1931), Miller-Commercial Directory Co.
High Point (1933), Hill Directory Co.
High Point (1940), Hill Directory Co.
High Point (1942-1943), Hill Directory Co.
High Point (1948), Hill Directory Co.

Kinston (1923-1924), Hill Directory Co.
Kinston (1928), Hill Directory Co.
Kinston (1936), Baldwin Directory Co.

Lenoir (1930-1931), Miller-Commercial Directory Co.
Lenoir (1937-1938), Miller-Commercial Directory Co.
Lenoir (1939-1940), Miller Press
Lenoir (1943-1944), Miller-Southern Directory Co.

Lenoir (1945-1946), Miller-Southern Directory Co.

Lexington (1916-1917), Chas S. Gardiner
Lexington (1937), Baldwin Directory Co.

Lumberton (1916-1917), Chas S. Gardiner
Lumberton (1938), Baldwin Directory Co.

Monroe (1922-1923), Miller-Commercial Directory Co.

Mooresville (1939-1940), Miller (Chas)-Southern Directory Co.

Morganton (1941-1942), Miller (Chas)-Southern Directory Co.
Morganton (1939-1940), Miller (Chas)-Southern Directory Co.
Morganton (1943-1944), Miller (Chas)-Southern Directory Co.

Mount Airy (1913-1914), J. Edwin Carter and A. Kyle Sydnor
Mount Airy (1928-1929), Miller-Commercial Directory Co.

New Bern (1880-1881), Edwards, Broughton and Co.
New Bern (1893), Edwards, Broughton and Co.
New Bern (1907-1908), Hill Directory Co.
New Bern (1916-1917), Chas S. Gardiner
New Bern (1918-1919), Hill Directory Co.
New Bern (1920-1921), Hill Directory Co.
New Bern (1937), Baldwin Directory Co.

North Wilkesboro, Wilkesboro (1939-40), Miller (Chas)-Southern Directory Co.

Oxford (1929-1930), Miller-Commercial Directory Co.

Raleigh (1875-1876), J. H. Chataigne (Raleigh, NC)
Raleigh (1880-1881), Edwards, Broughton and Co.
Raleigh (1883), Edwards, Broughton and Co.
Raleigh (1886), Turner, M'Lean and Losee Directory Co. (Raleigh, NC)
Raleigh (1887), Edwards, Broughton and Co.
Raleigh (1888), Observer Printing Co (Raleigh, NC)
Raleigh (1896-1897), Raleigh Stationery Company
Raleigh (1899-1900), Maloney Directory Co.
Raleigh (1901), Maloney Directory Co.
Raleigh (1903), Hill Directory Co.
Raleigh (1905-1906), Hill Directory Co.
Raleigh (1907-1908), Hill Directory Co.
Raleigh (1909-1910), Hill Directory Co.
Raleigh (1911-1912), Hill Directory Co.
Raleigh (1913-1914), Hill Directory Co.

Raleigh (1915-1916), Hill Directory Co.
Raleigh (1917), Hill Directory Co.
Raleigh (1918-1919), Hill Directory Co.
Raleigh (1919-1920), Hill Directory Co.
Raleigh (1921-1922), Hill Directory Co.
Raleigh (1922-1923), Hill Directory Co.
Raleigh (1923-1924), Hill Directory Co.
Raleigh (1924), Hill Directory Co.
Raleigh (1925), Hill Directory Co.
Raleigh (1926), Hill Directory Co.
Raleigh (1927), Hill Directory Co.
Raleigh (1928), Hill Directory Co.
Raleigh (1929), Hill Directory Co.
Raleigh (1930), Hill Directory Co.
Raleigh (1931), Hill Directory Co.
Raleigh (1932), Hill Directory Co.
Raleigh (1933), Hill Directory Co.
Raleigh (1934), Hill Directory Co.
Raleigh (1935), Hill Directory Co.
Raleigh (1936), Hill Directory Co.
Raleigh (1937), Hill Directory Co.
Raleigh (1938), Hill Directory Co.
Raleigh (1939), Hill Directory Co.
Raleigh (1940), Hill Directory Co.
Raleigh (1941), Hill Directory Co.
Raleigh (1942), Hill Directory Co.
Raleigh (1943), Hill Directory Co.
Raleigh (1945-1946), Hill Directory Co.
Raleigh (1948), Hill Directory Co.

Reidsville (1929), Miller-Commercial Directory Co.
Reidsville (1932), Hill Directory Co.
Reidsville (1935), Baldwin Directory Co.

Richmond County (1916-1917), Page Trust Co.

Roanoke Rapids (1938-1939), Miller (Chas)-Southern Directory Co.
Roanoke Rapids (1942-1943), Miller (Chas)-Southern Directory Co.

Rocky Mount (1908-1909), Hill Directory Co.
Rocky Mount (1912-1913), Hill Directory Co.
Rocky Mount (1914-1915), Hill Directory Co.
Rocky Mount (1920), Hill Directory Co.
Rocky Mount (1930), Hill Directory Co.
Rocky Mount (1934), Hill Directory Co.
Rocky Mount (1936), Hill Directory Co.

Rocky Mount (1938), Hill Directory Co.
Rocky Mount (1940), Hill Directory Co.
Rocky Mount (1942), Hill Directory Co.

Salisbury (1935), Baldwin Directory Co.
Salisbury (1938), Baldwin Directory Co.
Salisbury (1940), Baldwin Directory Co.

Salisbury-Spencer (1907-1908), Miller-Piedmont Directory Co.-Hackney and Moale
Salisbury-Spencer (1913-1914), Miller-Piedmont Directory Co.
Salisbury-Spencer (1915-1916), Miller-Piedmont Directory Co.-Hackney and Moale
Salisbury-Spencer (1917), Miller-Commercial Directory Co.
Salisbury-Spencer (1919-1920), Miller-Commercial Directory Co.
Salisbury-Spencer (1922-1923), Miller-Commercial Directory Co.
Salisbury-Spencer (1924-1925), Miller-Commercial Directory Co.
Salisbury-Spencer (1928-1929), Miller-Piedmont Directory Co.

Shelby (1937-1938), Miller (Chas)-Southern Directory Co.
Shelby (1939-1940), Miller (Chas)-Southern Directory Co.
Shelby (1941-1942), Miller (Chas)-Southern Directory Co.
Shelby (1943-1944), Miller (Chas)-Southern Directory Co.
Shelby (1945-1946), Miller (Chas)-Southern Directory Co.
Shelby (1947-1948), Miller (Chas)-Southern Directory Co.

Statesville (1916-1917), Miller-Piedmont Directory Co.
Statesville (1922-1923), Miller-Commercial Directory Co.
Statesville (1925-1926), Miller-Commercial Directory Co.
Statesville (1928-1929), Miller-Commercial Directory Co.
Statesville (1930-1931), Miller-Commercial Directory Co.
Statesville (1932-1933), Carolina Directory Co.
Statesville (1938-1939), Miller (Chas)-Southern Directory Co.
Statesville (1940-1941), Miller (Chas)-Southern Directory Co.
Statesville (1944-1945), Miller (Chas)-Southern Directory Co.
Statesville and Iredell County (1909-1910), Miller-Piedmont Directory Co.-Hackney and
      Moale

Thomasville (1930-1931), Miller-Commercial Directory Co.
Thomasville (1933-1934), Carolina Directory Co.
Thomasville (1935), Baldwin Directory Co.

Washington (1916-1917), Miller-Commercial Directory Co. (Piedmont)
Washington (1937), Baldwin Directory Co.

Waynesville and Haywood County (1906-1907), Hackney and Moale

Wilmington (1860-1861), Geo. H. Kelley

Wilmington (1865-1866), P. Heinsberger (Wilmington, NC)
Wilmington (1866-1867), Frank D. Smaw
Wilmington (1871), P. Heinsberger (Wilmington, NC)
Wilmington (1877-1878), Benj. R. Sherriff
Wilmington (1889), Wilmington Messenger/Messenger Steam Presses
Wilmington (1897), Hill Directory Co.
Wilmington (1905), Hill Directory Co.
Wilmington (1907), Hill Directory Co.
Wilmington (1909-1910), Hill Directory Co.
Wilmington (1911-1912), Hill Directory Co.
Wilmington (1913-1914), Hill Directory Co.
Wilmington (1917), Hill Directory Co.
Wilmington (1918), Hill Directory Co.
Wilmington (1919-1920), Hill Directory Co.
Wilmington (1922), Hill Directory Co.
Wilmington (1924), Hill Directory Co.
Wilmington (1926), Hill Directory Co.
Wilmington (1928), Hill Directory Co.
Wilmington (1930), Hill Directory Co.
Wilmington (1932), Hill Directory Co.
Wilmington (1934), Hill Directory Co.
Wilmington (1938), Hill Directory Co.
Wilmington (1940), Hill Directory Co.
Wilmington (1942), Hill Directory Co.
Wilmington (1943), Hill Directory Co.
Wilmington (1947), Hill Directory Co.

Wilson (1912-1913), Hill Directory Co.
Wilson (1916-1917), Hill Directory Co.
Wilson (1930), Hill Directory Co.

Winston-Salem (1908), Walsh Directory Co. (Charleston, SC)
Winston-Salem (1910), Miller-Piedmont Directory Co.
Winston-Salem (1911), Miller-Piedmont Directory Co.-Hackney and Moale
Winston-Salem (1912), Miller-Piedmont Directory Co.
Winston-Salem (1913), Miller-Piedmont Directory Co.
Winston-Salem (1915), Miller-Piedmont Directory Co.-Hackney and Moale
Winston-Salem (1916), Miller-Commercial Directory Co.
Winston-Salem (1918), Miller-Commercial Directory Co.-Hackney and Moale
Winston-Salem (1921), Miller-Commercial Directory Co.
Winston-Salem (1922), Miller-Commercial Directory Co.
Winston-Salem (1930), Miller-Commercial Directory Co.
Winston-Salem (1931), Miller-Commercial Directory Co.
Winston-Salem (1932), Hill Directory Co.
Winston-Salem (1933), Hill Directory Co.
Winston-Salem (1934), Hill Directory Co.

Winston-Salem (1935), Hill Directory Co.
Winston-Salem (1936), Hill Directory Co.
Winston-Salem (1937), Hill Directory Co.
Winston-Salem (1938), Hill Directory Co.
Winston-Salem (1939), Hill Directory Co.
Winston-Salem (1943), Hill Directory Co.
Winston-Salem (1946), Hill Directory Co.
Winston-Salem (1947-1948), Hill Directory Co.

Winston-Salem, Greensboro (1879-1880), Edwards, Broughton and Co.

Appendix 2. Publisher Attributes

Attributes for the top publishers (publishers with 5 or more directories in the collection).

Hill Directory Co. N=189

Hill Directory Co. Part I. 1897-1926. N=64

Attribute	Common Attribute	Exception
Year range	1897-1963	
City range	Map[22]	
Directories appearing before the general directory	A range of one or several directories including: Business Directory or Buyer's Guide Street Directory or Street Guide Miscellaneous	New Bern 1907 has nothing before general directory
Beginning page of general directory	Ranges from as low as page 45 to as high as page 213, with the bulk falling in the 100s	
Separate colored directory	No	
Street directory included	Y, though those for Durham, Rocky Mount and Wilmington frequently do not provide household listings. Many have 2 columns. When the header differs from general directory in later years, it is typically: first appearing street name page number last name on the page	None for New Bern (1907-1908, 1918-1909, 1920-1921) Kinston (1923-1924) Street Directories with 1 col: Rocky Mount (1908-1909, 1912-1915, 1919-1920) Wilmington (1897, 1905, 1907, 1909-1914) Wilson (1912-1913, 1916-1917)
Listing of abbreviations	All include general and job abbreviations but nothing else	
Title of general directory	All include "City Directory" in title	All directories between 1921 and 1926 also have "Hill Directory Co."
Racial designation	*	
Married status indicated	No, except when listing a widow's husband's name in parenthesis	Greensboro (1917, 1920) includes marital status
Number of columns	1	The following have 2 columns: Wilmington 1897 Durham 1925 Greensboro 1926
Header or footer	Header	
Even Header	1897-1910: First alpha name page	

[22] Interactive map may be accessed at http://batchgeo.com/map/c0c7840e0391e6c99a27915f45452633.

	number last alpha name on page 1911-1921: page number city name year directory 1922-1926: includes "Hill Directory Co"	
Odd Header	1897-1910: First alpha name page number last alpha name on page 1911-1921: city name year Directory page number 1922-1926: city name City Directory (year) page number	
Repeating last names	Names repeat in full	
Other symbols	When employed, the following symbols mean: h – head of household or resides at rms – rooms bds – boarder dash used when building number is missing	

Hill Directory Co. Part II. 1927-1952. N=113

Attribute	Common Attribute	Exception
Year range	1897-1963	
City range	[Map](#)	
Directories appearing before the general directory	Either none or a Buyer's Guide of advertisements (not a real directory)	
Beginning page of general directory	Ranges from 17 to 93, with the bulk falling somewhere between pages 21 and 33	Greensboro 1927 starts on page 105
Separate colored directory	No	High Point 1933
Street directory included	Yes – all with 2 columns	The following Street Directories have 3 columns: Charlotte (1939 onward) Greensboro (1940-1950) Raleigh (1948) Wilmington (1947) Winston-Salem (1946-1948)
Listing of abbreviations	In addition to general/job abbrevs., all directories after 1931 include name abbrevs. About 27% include special abbreviations. A handful include abbreviations for firms, streets and suburbs, and occupants.	
Title of general directory	All include "Hill's and "City Directory"	
Racial designation	(c)	High Point 1940 and 1948 do not include parenthesis.

Married status indicated	Y	Number of children indicated for Greensboro (1935-1943)
Number of columns	1	The following have 2 columns: Charlotte (1932-1947) Durham (1941-1949) Greensboro (1927-1952) Raleigh (1948) Wilmington (1947) Winston-Salem (1932-1939, 1943, 1946-1948)
Header or footer	Header through 1930, none from 1931-1951. Where no header appears, the page number is centered at top of page.	The following have header: Burlington (1934) Goldsboro (1935) High Point (1933) Reidsville (1932)
Even Header	1927-1930: includes "Hill Directory Co." Where headers exist (1932-1935): includes "Hill Directory Co"	
Odd Header	1927-1930: city name City Directory (year) page number 1932-1935 where applicable: city name City Directory (year) page number	High Point (1933): city name [race] population (year)
Repeating last names	1927-1940 names repeat in full 1940-1941 name cont. w. " 1942-1945 name cont. w. --- 1946-1952 names cont. w. "	*1927-1940 exceptions (use "):* Charlotte (1934-1939) Winston Salem (1932-1939) *1940-1941 exceptions:* High Point (1940) and Rocky Mount (1940) repeat names; Raleigh (1940, 1941) uses --- *1942-1945 exceptions (use "):* Charlotte (1942, 1945, 1947) Durham (1942) Greensboro (1942, 1943, 1945, 1946) Winston-Salem (1946) *1946-1952 exceptions (---):* Burlington (1946) Gastonia (1947)
Other symbols	When employed, the following symbols mean: h – head of household rms – rooms bds – boarder o in circle – homeowner ( ) – may indicate suburb when after name Bell icon – has telephone	

	Dash used when building number is missing	

Hill Directory Co. Part III. 1953-1963. N=12

Attribute	Common Attribute	Exception
Year range	1897-1963	
City range	Greensboro. Map	
Directories appearing before the general directory	None	
Beginning page of general directory	1	Greensboro 1951-1952 begins on page 25
Separate colored directory	No	
Street directory included	Yes – all have 3 columns	
Listing of abbreviations	Includes general abbreviations, jobs, names and special abbreviations	
Title of general dir.	All include "Hill's and "City Directory"	
Racial designation	None listed after 1953	
Married status indicated	Yes	
Number of columns	2	
Header or footer	No header until 1960. Page numbers centered at top of page. After 1960, still just page number centered, but separated by lines on top and bottom.	
Even Header	N/A	
Odd Header	N/A	
Repeating last names	Name cont. w. --	Name cont. w. " for Greensboro 1961, 1962
Other symbols	When employed, the following symbols mean: h – head of household o in circle – homeowner	

Miller Directories. N=109

Attribute	Common Attribute	Exception
Year range	1916-1947	
City range	Map[23]	
Directories appearing before the general directory	A range of none to some, including Miscellaneous Directory Advertisers' Special Directory Numerical Telephone Directory	Street Directory for Asheville 1917, 1918 appears in the beginning.

---

[23] Interactive map may be accessed at http://batchgeo.com/map/eb891bf29310f3c35d81352bfc5e1e8a.

	Classified Buyers Guide	
Beginning page of general directory	101 – note that many of the scans skip front matter to start of General Directory. Salisbury-Spencer directories have separate listings for each town, but Salisbury begins at 101 (1919-1920, 1922-1923, 1924-1925)	
Separate colored directory	No	Winston-Salem (1916, 1918, 1921, 1922) has a separate directory for whites and colored
Street directory included	Yes. Most are 2 columns and frequently have a different header: first street name page number last name on page	
Listing of abbreviations	Yes	In addition, many of the 1929 and 1930 also include proper name abbreviations, as does Winston-Salem 1931.
Title of general directory	All have "City Directory" or "City and Suburban Directory" in title	Winston-Salem (1916, 1918, 1921, 1922) lists race in the header
Racial designation	*	Winston-Salem (1916, 1918, 1921, 1922) only uses * in street directory; no designation in Alpha Directory
Married status indicated	Yes	
Number of columns	1	2 columns in Charlotte (1926, 1927, 1928, 1929, 1930)
Header or footer	Header	
Even Header	All include page number and "Directory" in addition to city name and year	
Odd Header	All include page number and "Directory" in addition to city name and year	
Repeating last names	Names repeat in full	
Other symbols	Street index uses a range of symbols to denote houses (h), residences (r) and homeowners (o in circle). Dashes used when building number is missing in street index.	

Includes Miller Press, Miller/Commercial Directory Co., Miller/Southern Directory Co., Miller/Southern Directory Co. (Chas Miller), Miller/Commercial/Hackney and Moale, and Miller/Commercial (Piedmont).

Miller-Piedmont Directory Co. and Miller-Piedmont/Hackney and Moale. N = 45

Attribute	Common Attribute	Exception
Year range	1907-1916, 1928-1947	
City range	[Map](24)	
Directories appearing before the general directory	Some combination of: Special Advertisers' Directory Miscellaneous Directory Numerical Telephone Directory	No directories before Asheville (1913, 1931, 1935-1947) Concord (1908, 1913, 1916) Gastonia (1910, 1913) High Point (1908, 1911, 1913) Salisbury (1907) Statesville (1916) Winston-Salem (1910-1912)
Beginning page of general directory	Either page 65 or page 101	Asheville 1935-1947 (p 25-33) Salisbury/Spencer have separate general directories for each town, but beginning page adheres to common attribute
Separate colored directory	No	Winston-Salem 1911, 1912, 1913, 1915 have separate listings with no racial designations (except in Street Directories)
Street directory included	Yes, frequently a different header: first street name page number last name on page	
Listing of abbreviations	Yes	Name abbreviations appear for Asheville 1937-1947 and Statesville 1916. Further abbreviations included sporadically for Asheville 1937-1947.
Title of general dir.	All have "Directory" in the title	
Racial designation	*	(c) in Asheville 1935, Charlotte 1931 No designation in Winston-Salem 1912-1913, 1915 but * in street directory
Married status indicated	Yes	
Number of columns	1	2 columns in Charlotte 1931
Header or footer	header	None in Asheville 1936-1947 (just a page number centered at the top above ads and listings)
Even Header	All that have a header include page number and "Directory"	
Odd Header	All that have a header include "Directory" and page number	
Repeating last names	Names repeat in full	Name cont. w. " in resident

		directory for Asheville (1943-1947)
Other symbols	O in circle denotes homeowner in street index. Dashes used when building number is missing in street index.	Asheville (1937-1947) uses a bell icon in street index to connote when someone has a telephone.

Baldwin Directory Co. N=14

Attribute	Common Attribute	Exception
Year range	1935-1940	1935 (3 instances)
City range	map[25]	
Directories appearing before the general directory	Miscellaneous Directory	
Beginning page of general directory	41-77	
Separate colored directory	No	
Street directory included	Yes	
Listing of abbreviations	General/Jobs: Yes Names: Yes	1935 does not include names. 2 of the 3 for 1935 include street/suburb abbrevs.
Title of general dir.	Resident Directory	
Racial designation	©	1935 uses *
Married status indicated	Y	
Number of columns	1	Salisbury 1935 has 2
Header or footer	Header	
Even Header	All contain "Baldwin" or "Baldwin's"	
Odd Header	All have "City Directory"	
Repeating last names	Names repeat in full	Salisbury 1935 name cont. w. "
Other symbols	Home ownership denoted as (h)	1935 directories list as o in circle

Edwards, Broughton and Co. N=6

Attribute	Common Attribute	Exception
Year range	1879-1893	
City range	New Bern (1880-1881, 1893) Raleigh (1880-1881, 1883, 1887) Winston, Salem, Greensboro (1879-1880)	
Directories appearing before the general	Street Directory	Raleigh 1887 also has a Church Directory;

---

[25] Interactive map may be accessed at http://batchgeo.com/map/7b44e4824fa777bb6fb4f5ca156e2972.

directory		Raleigh 1883 includes a page of "Names received too late for alphabetical list"
Beginning page of general directory	Ranges from page 1 to page 39	
Separate colored directory	No	
Street directory included	Yes, but none are household listings	New Bern 1893
Listing of abbreviations	Yes	Raleigh 1883
Title of general directory	Includes "City Directory"	Winston, Salem, Greensboro just uses "Directory"
Racial designation	(c)	Raleigh 1883: col; Raleigh 1887: (col)
Married status indicated	No	
Number of columns	1	
Header or footer	Header	
Even Header	Includes page number and "Directory"	
Odd Header	Includes city name and page number	
Repeating last names	Names repeat in full	
Other symbols		

Chas. S. Gardiner. N=6

Attribute	Common Attribute	Exception
Year range	1915-1918	1918-1919 (1 instance)
City range	Map[26]	
Directories appearing before the general directory	None	
Beginning page of general directory	103 Scan skips front matter	1918 (begins page 7)
Separate colored directory	No	
Street dir. included	Yes	1918
Listing of abbreviations	Yes	
Title of general directory	Inconsistent across directories. All include the word "Directory" in the title.	
Racial designation	*	
Married status indicated	Yes	
Number of columns	1	
Header or footer	Header	
Even Header	page number city name N C (year) Directory	

---

[26] Interactive map may be accessed at http://batchgeo.com/map/4019d60fe7288a88b17ff6fd414091ac.

Odd Header	city name N C (year) Directory page number	
Repeating last names	Names repeat in full	
Other symbols	* also may also denote special business heading	

Maloney Directory Co. N=6

Attribute	Common Attribute	Exception
Year range	1899-1901	
City range	Asheville (1899-1900, 1900-1901) Greensboro (1899-1900, 1901) Raleigh (1899-1900, 1901)	
Directories appearing before the general directory	Street Index and Appendix	Asheville does not have Appendix
Beginning page of general directory	Ranges, but nothing before page 59. Most appear after page 97.	
Separate colored directory	Yes	Raleigh 1901
Street directory included	Yes	
Listing of abbreviations	Yes	Asheville 1900-1901
Title of general dir.	all include "City Directory" in title	
Racial designation	(c)	no racial designation for Asheville directories
Married status indicated	Yes	
Number of columns	1	
Header or footer	Header	
Even Header	Header includes include page number and city name	
Odd Header	All include "Directory" and page number	
Repeating last names	Name cont. w. " (not in street index)	1901 directories
Other symbols		Greensboro directories include additional designations: r - residence bds-boards rms-rooms

Walsh Directory Co. N=5

Attribute	Common Attribute	Exception
Year range	1905-1910	
City range	Charlotte (1905-1906, 1907, 1909, 1910) Winston-Salem (1908)	
Directories appearing before the general	Church Directory Miscellaneous Directory	

directory	Street Directory	
Beginning page of general directory	Ranges from page 157 to page 181	
Separate colored directory	Yes	
Street directory included	Yes – all have 2 columns	
Listing of abbreviations	Yes	
Title of general directory	Includes "White [or Colored] Department Walsh's Directory"	
Racial designation	None	
Married status indicated	Y	
Number of columns	1	
Header or footer	Header	
Even Header	First alpha name page number last alpha name on page	
Odd Header	Same as even	
Repeating last names	Names repeat in full	
Other symbols	r denotes residence bds denotes boards	