Exploring Evaluative Methods for Large-Scale Local History

Katrina Fenlon
Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel St, MC-492
Champaign, IL
kfenlon2@illinois.edu

ABSTRACT

Local history is a topic of proven interest to library users. Digital cultural heritage collections and aggregations offer an ever-growing wealth of digitized, primary source materials to support local history research. Yet methods for strategic development and evaluation of local history collections and aggregations remain sparse. In this paper we present a multimodal pilot study to explore methods for evaluating local history coverage in digital collections and aggregations. We explore the possibility of employing automatic classification to identify items and collections relevant to local history, and conduct an exploratory metadata analysis to discern features of collections that contribute to their usefulness for local history research.

Keywords

Metadata, digital libraries, local history, aggregations, cultural heritage.

INTRODUCTION

Digital cultural heritage collections and aggregations offer an ever-growing wealth of digitized, primary source materials to support local history research. Intensive digitization efforts have transformed the local collections of libraries, historical societies, and archives. In turn, cultural heritage aggregations, such as the Digital Public Library of America (DPLA, http://dp.la/), seek to capitalize on these efforts by integrating local materials from different places. To support local history research, a critical mass of items centered on a location is necessary but not sufficient. Research collections, including those that researchers construct for their own use, do more than contain extensive raw material: they exhibit thematic coherence and purposeful selection, are constituted of heterogeneous data types, and are designed to support research (Unsworth, 2000; Palmer, 2004). In short, local history collections of high utility conform to the development principle of contextual mass (Palmer, 2004; Palmer, et al., 2010). Beyond geographic coverage, what aspects of a collection or aggregation contribute to its usefulness as a local history resource?

In this paper we present a multimodal pilot study to explore methods for evaluating local history coverage in digital collections and aggregations.

BACKGROUND

Previous work identified technical challenges to evaluating local history coverage in aggregations, including data heterogeneity, location disambiguation, and visualizing spatial information at different levels of granularity (Fenlon & Varvel, 2013). In this paper, we turn to a conceptual challenge: understanding what factors of a collection or aggregation contribute to its usefulness for local history research.

Local history is the most popular subject of historical research in public libraries (D'Elia et al., 2002; Pettigrew et al., 2002; Fenlon & Varvel, 2013). Research on local and special-collections development rose in the last decades of the 20th century, but was not adapted to the era of large-scale, digital aggregations (e.g., Dunaway & Baum, 1984; Hitchcock, 1990; Phillips, 1995; Graham, 1998). Recent work has recognized the potential of digital collections to serve the high demand for local history (Reid & Macafee, 2007), but focuses largely on the act of digitization. While there are some guideposts for practitioners (RUSA, 2006), the literature remains weak on specific methods for the purposeful development and evaluation of digital, local history collections.

METHODS

This pilot evaluation employed two separate experiments to (1) identify items and collections relevant to local history, and (2) explore what features contribute to their usefulness. The first experiment attempted to classify items as relevant or irrelevant to local history using a supervised

ASIST 2013, November 1-6, 2013, Montreal, Quebec, Canada.

classification algorithm. The second was an exploratory analysis of features of collections that contribute to usefulness for local history research.

The data for this study comes from the DPLA open data provider, which offers more than 2.5 million item records from libraries, museums, and archives across the U.S.

Local history Classification Experiment

This part of the pilot study aimed to determine whether a simple classification effort could successfully distinguish cultural heritage items that are directly relevant to local history from those that are indirectly or not relevant to local history. This is a fine line to draw, since most cultural heritage items are relevant to the history of somewhere. While all cultural heritage collections seem likely to have tangential relevance to the study of local history, for example by providing context for local history items, our focus is on immediate, explicit relevance. We aim to test whether descriptive metadata (beyond spatial information) holds clues to whether an item is likely to serve local history needs. The reasoning is that, if such items can be identified, we may be able to discover latent local history collections within massive, messy cultural heritage aggregations. Note that this phase of analysis targets the item-level record. Our goal is not to optimize classification at the outset, but to rapidly discern the viability of this method for our purpose.

Five DPLA data providers were sampled for this experiment. We limited the sample to state library collections, which, by definition, tend to hold localized content from diverse institutions. We retrieved sample data from the DPLA data provider, via bulk download.[1] The records were pared down to title, description, and subject fields, which contain free-text descriptive information. A training set of 250 records was randomly selected from the aggregate, and manually evaluated for direct relevance to local history. A further random sample of 500 records was taken from each data provider, for a total unlabeled sample of 2500 records. We employed the open-source MALLET[2] (MAchine Learning for LanguagE Toolkit) software for classification. We used the default Naïve Bayes classifier, with the option to remove stopwords. The classifier was trained on the manual training set and applied to the unlabeled data. The classifier assigned probabilities of relevance (to local history) to each of the 2500 metadata records. Afterward, we selected a subset of 250 newly classified records for manual evaluation of the classifier's success. Results are discussed below.

Contextual Mass – Exploratory Analysis

Having piloted a method for discerning basic relevance, we undertook to identify aspects of collections (and, ultimately, aggregations) that contribute to their usefulness for local history research. Our forward-looking goal is to isolate evaluative metrics for aggregations and collections. We aim to begin to formalize measurable features of the aforementioned development principle of contextual mass.

Drawing on the literature on thematic research collections, we hypothesized the desirable semantic characteristics of local history collections. For now we leave aside syntactic or technical data characteristics. Within any given collection, we hypothesize that:

- **Spatial** fields should predominantly (but perhaps not exclusively) indicate focused coverage of a region, state, or lower geographic division.
- **Type** and **format** fields should indicate a diverse variety of item types.
- **Date** fields should indicate a wide span of temporal coverage, with deeper coverage of events of particular importance.
- **Subject, title, and description** fields should indicate a diverse variety of topics, connecting to or intersecting with the umbrella topic of local history.

These collection-specific criteria, which must be considered preliminary and unproven despite their grounding in the literature, will be modified and expended for aggregations in future work.

This phase of the study tests our criteria against the model of extant local history collections, which libraries and historical societies have curated as thematic research collections. We conducted an analysis of item- and collection-level metadata of extant collections for hallmarks of contextual mass, such as those given by the criteria above.

The DPLA Metadata Application Profile[3] accommodates minimal collection description, but collections are not accessible through the portal via search or browse. Therefore, we relied on the API, which enables some collection-specific queries. We found 10 collections in the DPLA that self-identified as pertaining to local history. From those, we selected a few for manual, qualitative analysis of both the free-text collection descriptions and records of constituent items. For the sake of concision, we present in this paper a single case: the Tracy W. O'Neal Collection.[4]

In the case of the O'Neal collection, summary information on type, temporal, and spatial coverage were given by the

[1] http://dp.la/info/developers/download/

[2] http://mallet.cs.umass.edu/

[3] http://dp.la/info/map/

[4] Available at host institution: http://digitalcollections.library.gsu.edu/cdm/landingpage/collection/oneal

collection description. Our exploration did not require further analysis of these fields. We focus here on subject, description, and title: topical and largely free-text elements that are notoriously difficult to analyze. We retrieved topical metadata for all items in this collection from the DPLA API. We normalized the records and converted them into word tokens for aggregate analysis. In addition to stopwords, we removed field-specific, omnipresent or administrative phrases (such as "local identification number"). Because subjects obeyed a controlled vocabulary, there was no need to tokenize them. Finally, we conducted a crude analysis of term frequencies.

FINDINGS AND DISCUSSION

Local history Classification Experiment

Table 1 gives the results of the classification process. 195 metadata records were classified as directly relevant to local history. Of those, our manual evaluation of a random sample of the classifier judgments determined that 10 records were incorrectly classified as relevant; in other words, 5% were false positives. For imaginable applications, this appears to be an acceptable error rate in the determination of relevance. For example, if this classifier were used to automatically augment subject metadata with a controlled term for local history, a meager 5% would be incorrectly labeled, and the classifier would err on the side of inclusivity. On the other hand, more than half of the documents were falsely rejected as irrelevant. This is an unacceptably high error rate. A closer look at the data suggests that the error rate is a result of our quick-and-dirty training process. It does not speak to the viability of the method for local history classification. The training set for future classification work will need to be significantly larger, and include a higher proportion of negative judgments. On the whole, classification appears to be promising for the identification of metadata records relevant to local history. This is a positive finding, given the subtle distinction we are aiming to make between cultural heritage records generally, and those pertaining specifically to local history. The classifier appears to have been equally successful across all five data providers, but the data providers were relatively homogeneous. Future work should explore the generalizability of a refined classifier to data from heterogeneous providers.

Contextual Mass – Exploratory Analysis

Our exploratory metadata analysis of the Tracy W. O'Neal Collection suggests that our hypothesized criteria for contextual mass in local history collections are on the right track. Spatial coverage for this collection is narrow in scope – indeed, focused on a single city. Temporally, the collection covers a wide swath from 1923-1975, with most items dating from 1950-1974. Some images in the collection, the description notes, are copy negatives of photographs dating as early as 1889. The collection is homogeneous in item-type, comprising 1,489 images (derived from a subset of more than 31,500 acetate negatives in the original collection). While we hypothesized that a local history research collection should exhibit a diversity of media types, this collection must be interpreted in its aggregation context, where other collections can be counted on to supply diversity.

	Classified Relevant	Classified Irrelevant
Correct Judgment	185	26
Incorrect Judgment	10 (5%)	29 (52%)

Table 1. Classifier and evaluation results.

How far can we get toward understanding the breadth and depth of topical coverage in a collection or aggregation, using a simple, bag-of-words approach? Our immediate concerns are three: (1) most prevalent topics and their proximity to local history; (2) the breadth or range of topical coverage; and (3) the depth of coverage of various topics (i.e., subject strengths and weakness of the collection).

Figures 1, 2, and 3 show the top title, description, and subject terms in the collection. All three fields show remarkable continuity with one another. This suggests that our combined approach to the analysis of these three fields was not misguided. The terms are notably localized, having to do with streets and avenues, local sports, transportation, and even particular buildings. A quick analysis of term frequencies across topical metadata fields appears to be a strong indicator of the local history focus of a collection.

Measuring topical breadth and depth are thornier problems. Future work will apply more advanced techniques for understanding topical coverage, such as topic modeling and language modeling visualizations, demonstrated in Fenlon, et al. (2012).

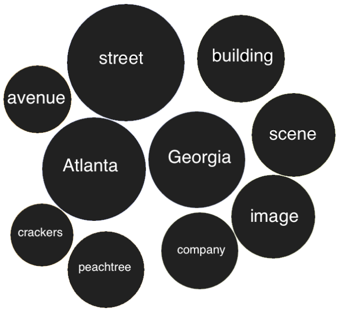

Figure 1. Top 10 Title Terms in Collection.

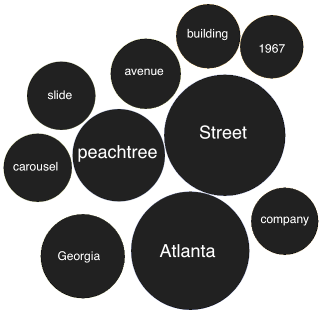

Figure 2. Top 10 Description Terms in Collection.

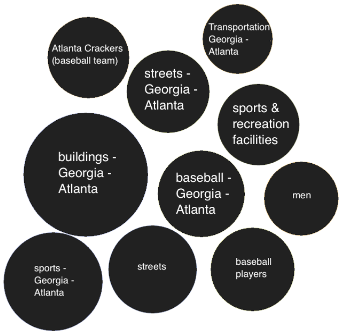

Figure 3. Top 10 Subject Terms in Collection.

CONCLUSION AND NEXT STEPS

Our pilot evaluation of local history in state library collections drawn from the DPLA suggests that classification and even very simple text analysis of metadata records have great potential to help us identify items and collections relevant to local history, and evaluate their coverage. Future work will explore the refinement and generalization of the classification technique to identify local history items and collections. Future evaluations of contextual mass must improve on treatment of topical information in metadata records, probably through the use of advanced language modeling techniques such as topic modeling. The preliminary criteria of contextual mass, which we began to validate, must be rigorously tested, with the ultimate goal of producing formal metrics for evaluation of the local history coverage of collections. In addition, they must be adapted to the aggregation (rather than collection-specific) context. At least, the increase in scale will demand greater efficiency from our evaluative techniques. Other factors, not necessarily explicit in metadata, will be considered. Future work will explore the relationships between collections, the impact of institutional representation in aggregations, and the exciting potentiality of latent local history collections within aggregations.

REFERENCES

D'Elia, G., Jörgensen, C., Woelfel, J., & Rodger, E. J. (2002). The impact of the internet on public library use: An analysis of the current consumer market for library and internet services. *Journal of the American Society for Information Science & Technology*, *53*(10), 802–820. doi:10.1002/asi.10102

Dunaway, D. K., & Baum, W. K. (1984). *Oral history: an interdisciplinary anthology*. American Assoc. State Local History.

Efron, M., Organisciak, P., & Fenlon, K. (2011). Building topic models in a federated digital library through selective document exclusion. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*. Presented at ASIS&T 2011.

Fenlon, K., & Varvel, V. E. (2013). Local histories in global digital libraries: Identifying demand and evaluating coverage. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. Presented at the ACM/IEEE Joint Conference on Digital Libraries (JCDL 2013), Indianapolis, Indiana.

Fenlon, K., Organisciak, P., & Efron, M. (2012). Tooling the Aggregator's Workbench: Metadata Visualization Through Statistical Text Analysis. *Proceedings of the ASIS&T 2012 Annual Meeting*.

Graham, P. S. (1998). New roles for special collections on the network. *College & Research Libraries*, *59*(3), 232–239. Retrieved from http://crl.acrl.org/content/59/3/232.short

Hitchcock, E. R. (1990). Materials used in the research of state history: A citation analysis of the 1986 Tennessee Historical Quarterly. *Collection Building*, *10*(1/2), 52–54. doi:10.1108/eb023268

Palmer, C. (2004). Thematic Research Collections. In *A Companion to Digital Humanities*. Blackwell Publishing.

Palmer, C. L., Zavalina, O., & Fenlon, K. (2010). Beyond size and search: Building contextual mass in digital aggregations for scholarly use. In *Proceedings of the ASIS&T Annual Meeting*. Pittsburgh, PA.

Pettigrew, K. E., Durrance, J. C., & Unruh, K. T. (2002). Facilitating community information seeking using the Internet: Findings from three public library-community network systems. *Journal of the American Society for Information Science & Technology*, *53*(11), 894–903. doi:10.1002/asi.10120

Phillips, F. (1995). *Local History Collections in Libraries*. Libraries Unlimited.

Reference and User Services Association. (2006). *RUSA guidelines for establishing local history collections* (RUSA Guidelines). American Library Association. Retrieved from http://www.ala.org/rusa/resources/guidelines/guidelinesestablishing

Reid, P. H., & Macafee, C. (2007). The philosophy of local studies in the interactive age. *Journal of Librarianship & Information Science*, *39*(3).

Unsworth, J. (2000). Thematic Research Collections. Presented at the Modern Languages Association Annual Conference, Washington, D.C. Retrieved from http://www.iath.virginia.edu/~jmu2m/MLA.00/.

The columns on the last page should be of approximately equal length.