ARTICLE

Applying Topic Modeling for Automated Creation of Descriptive Metadata for Digital Collections

Monika Glowacka-Musial

ABSTRACT

Creation of descriptive metadata for digital objects tends to be a laborious process. Specifically, subject analysis that seeks to classify the intellectual content of digitized documents typically requires considerable time and effort to determine subject headings that best represent the substance of these documents. This project examines the use of topic modeling to streamline the workflow for assigning subject headings to the digital collection of New Mexico State University news releases issued between 1958 and 2020. The optimization of the workflow enables timely scholarly access to unique primary source documentation.

INTRODUCTION

Digital scholarship relies on digital collections and data. In the influential book *Digital_Humanities*, Anna Burdick and her associates affirm that humanistic knowledge production depends on collection building and curation.[1] Access to historical documents and data resources is essential for the development of new research questions and methodologies.[2] This project utilizes topic modeling to support building a digital collection of institutional news releases. It is one of the initiatives to apply digital technologies in the library workflows.

NEW MEXICO STATE UNIVERSITY NEWS RELEASES

In response to a growing scholarly and public interest in original university press announcements, the digitization of past NMSU print news releases was approved in September 2013. Sixty years of news releases starting from the late 1950s to the present were to be included. One of the arguments presented in justification of the project was that these institutional news briefs have a truly unique historical value. Researchers view university press announcements as anchors in the history of NMSU and the region, particularly for dating events and initiatives. They also find official communications essential for studying the way the news was framed by participants and the university administration.

Historically, the relationships between the university and the local media had always been a major concern of college administrators: how to respect the freedom of the press, while ensuring responsible and factual journalism, and how to build an effective partnership that would benefit both sides?[3] To address these questions, the administration early on established the college's Information Services that have issued news releases about campus events, programs, and developments in the college's research, teaching, and service. These formal news reports representing the perspective of the university have been regularly distributed to local and worldwide media for many decades. This collection has become one of the most popular primary sources documenting a history of the Southwestern educational institution.

Monika Glowacka-Musial (monikagm@nmsu.edu) is Assistant Professor/Metadata Librarian, New Mexico State University Library.

https://doi.org/10.6017/ital.v41i2.13799

Since the beginning of the digitization project, thousands of press releases had been scanned, described, and added to the digital collection. Currently, the collection features press releases issued by the university between 1958 and 1974. There is still a lot to be done. The most time-consuming element in the process is adding metadata, including Library of Congress Subject Headings, to individual news releases. With decreasing personnel, dwindling library resources, and competing work priorities, the progress on the project has slowed substantially. Its revitalization requires a fresh, problem-solving approach that would allow for a significant reduction of time that catalogers spend on metadata creation. In search for a viable solution, topic modeling, a computational tool for classifying large collections of texts, was put to the test and generated promising results. The following sections describe the tools, data, and process created for this experiment in some detail.

TOPIC MODELING AND ITS APPLICATIONS

Topic modeling (TM) is one of the methodologies used in natural language processing (NLP). It was specifically designed for text mining and discovering hidden patterns in huge collections of documents, images, and networks.[4] According to practitioners, topic modeling is best viewed as a statistical tool for text exploration and open-ended discovery.[5] It has been used extensively in computer science, genetics, marketing, political science, journalism, and digital humanities for the last two decades. A growing literature on topic modeling applications provides clear evidence of its viability.[6] Examples of TM applications in digital social sciences and humanities include finding geographic themes from GPS-associated documents on social media platforms such as Flickr and Twitter,[7] selecting news articles on opposition to Euro currency from Financial Times data,[8] identifying paragraphs on epistemological concerns in English and German novels,[9] tracking research trends in different disciplines,[10] and revealing dominant themes in newspapers,[11] governance literature,[12] and Wikipedia entries.[13]

Topic modeling was applied in addition to text mining to enhance access to large digital collections by providing minimal description and enriching metadata, including subject headings.[14] Also, a possibility of using topic modeling to determine the subject headings for books on Project Gutenberg was explored.[15]

Topic modeling in a nutshell

Topic models help to identify the contents of document collections. Topic modeling is a process of discovering clusters of words that best represent a set of topics. Figure 1 shows the basic idea behind topic modeling.

A large collection of text documents (the scrolls on top) consists of thousands of words (shown symbolically at the bottom). The algorithm seeks for the most frequent words that tend to occur in proximity and clusters them together. Each cluster, referred to as a topic, has a set of words with their probabilities of belonging to a given topic. Each document in the collection displays a set of combined topics to different degrees. Here, documents are seen as mixtures of topics, and topics are seen as mixtures of words.[16] Topics also provide context to words. Documents that have similar combinations of topics tend to be related. As a result, a large collection of text documents can be represented by a limited set of topics (as presented by icons in the middle of the figure).

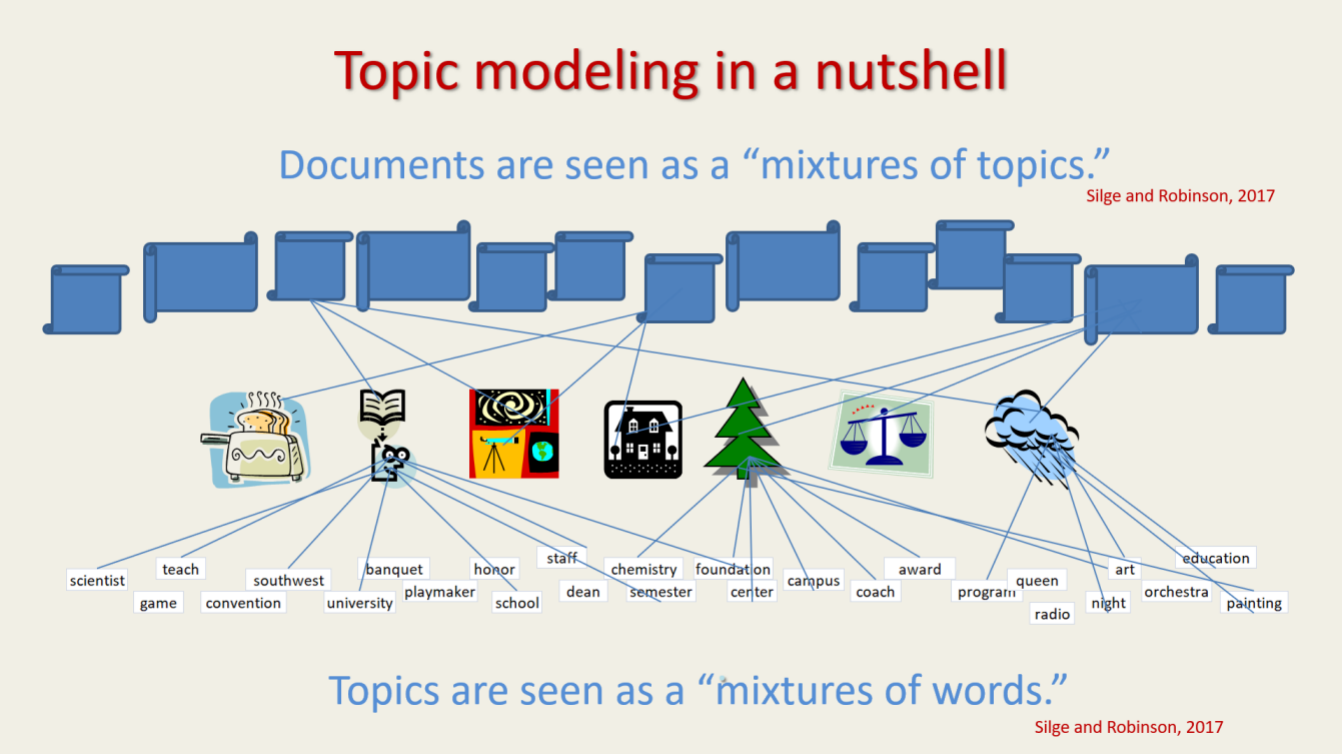

Figure 1. Basic idea behind topic modeling.

Topics and subject headings combined

The original purpose of topic modeling, as formulated by David Blei and his associates in 2003, was to make large collections of texts more approachable for scholars by organizing texts automatically based on latent topics.[17] These hidden topics can be discovered, measured, and consequently used by scholars to navigate the collection.

The purpose of assigning subject headings is to identify "aboutness," or simply subject concepts, covered by the intellectual content of a given work, and then again collocate related works.[18] Since both topic models and subject headings have a similar purpose, although very different methodology and scale, we decided to combine them and make topic models a prerequisite for assigning subject headings. In such a scenario, the computer deals with the scale of text collections that are beyond human reading capacity and catalogers then fine-tune the results generated by the algorithm. The following Methods section shows subsequent stages involved in the process of semiautomated assignment of subject headings to documents.

METHODS

Overview

For topic modeling, we used the algorithm of Latent Dirichlet Allocation (LDA).[19] LDA takes a document-term matrix, with rows corresponding to documents, and columns corresponding to terms (words) and, based on semirandom exploration, finds optimal probabilities of topics in documents (called *gammas*), and probabilities of terms in topics (called *betas*).

After LDA generates a set of topics that best represent the collection of news releases, each topic is associated with several subject headings that were previously assigned to news releases by catalogers. For a new news release, LDA finds a set of most representative topics. Subject headings

associated with the dominant topics are combined into a list of subject candidates presented to a cataloger.

The last step involves a cataloger using a short list of subject candidates for selecting subject headings for news releases.

Training data

Training data used in this project consists of over 6,000 news releases (from 1958 to 1967) annotated with metadata. Only two metadata properties—titles and subject headings—were considered. Created by catalogers, both properties reflect the content of news releases accurately, although mistakes may sometimes happen. The values from the titles field were converted into a document-term matrix that, in turn, became an input for the algorithm. Texts produced by OCR on original news releases were not included in the analysis due to their poor quality.

Detailed steps of the proposed method:

1. Topic modeling on training data:
    a. Run standard preprocessing of training text data, including tokenization, stop words removal, and stemming.
    b. Run topic modeling (LDA) where each document from the training data set is assigned a set of topics (subsets of words), each one with a measurable contribution to the document.
2. Assignment of subject headings to topics.[20] For each topic:
    a. Select a number of documents with the highest probability (*gamma*) for the topic. We used 400.
    b. Gather a set of subject headings assigned to documents produced in 2.a. and arrange them with decreasing frequency (*freq*) of occurrence in the set.
3. Assignment of subject headings to a new document.
    a. Assign to the new document *gammas* (probabilities) of topics using the LDA model trained in 1.b.
    b. In subsequent topics, for each subject heading calculate its weight in the document as a product of its frequency in the topic (*freq*) and probability of the topic (*gamma*) in the document; for subject headings duplicated across topics, sum up their weights across topics.
    c. Create a list of candidate subject headings processed in 3.b. in descending order with respect to their weights in the document.

IMPLEMENTATION

There is a growing number of tools used for topic modeling.[21] For this project, we used the R programming language, which has many packages for data preprocessing and topic modeling (TM).[22]

Below are listed R packages used for this project:

- *topicmodels* with functions: LDA() producing topic models, posterior() for assigning topics to test documents by pretrained models and perplexity() for perplexity calculation[23]
- *tidytext* with tidying functions that allow for re-arrangements and exploring data as well as for interpreting the models
- *textstem* for preprocessing data, including stemming and lemmatization

- *tidyr*, *dplyr*, and *stringr* for data and strings manipulation and arrangements
- *ggplot2* for data visualizations

The code related to topic modeling was mostly reused from the DataCamp class on topic modeling.[24] Occasionally, data.table data structure was applied instead of data.frame.

In addition to standard stop-words, custom stop-words including initials, names of weekdays, and dates were removed from the corpus using function anti_join().

For finding topics in test documents by a pretrained model, function posterior() from the R package topicmodels was used.[25] The extra step needed before using function posterior() was to align the new document with the document-term matrix used for training the LDA model.[26]

RESULTS

For assessing the method's performance, we adopted the idea of recall. In this specific context, recall is defined as the fraction of original subject headings (i.e., those assigned to a document manually by a cataloger) that are present on the list of candidate subject headings produced by the method.

The average recall is estimated using a leave-one-out setting.[27] Once a single test document is set aside, the LDA model is trained on the remaining documents and recall is calculated for the tested document using the list of candidate subject headings produced by the method. Then, recall is averaged over a set of testing documents. This approach produces an estimate of the method's performance if tested on a new document.

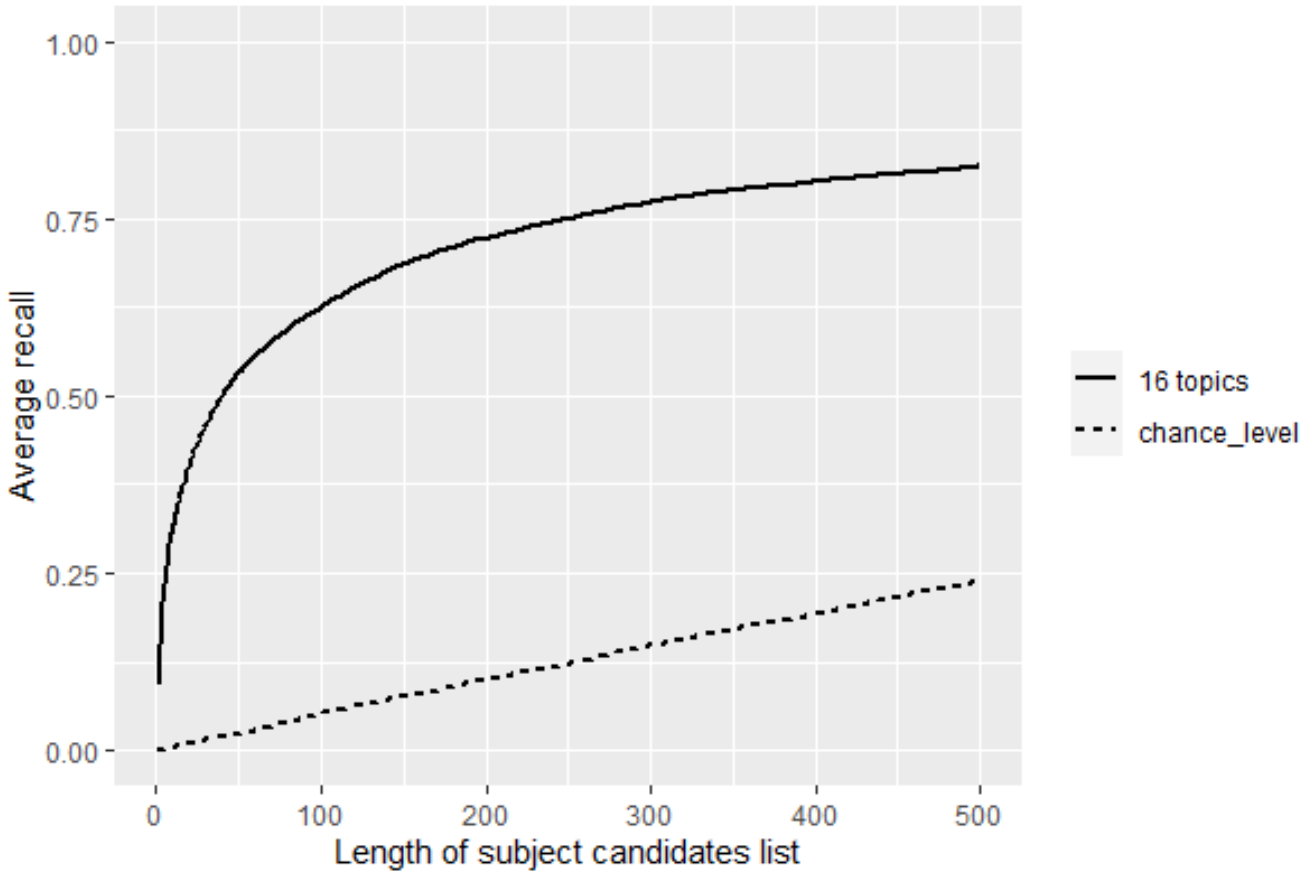

Figure 2. Average recall as a function of size of list with subject headings candidates.

Figure 2 shows the dependence of average recall on length of list of candidate subjects produced by the method. Recall is averaged over 1,500 randomly selected test documents. The dashed line represents the chance level performance, i.e., when the method would produce a random subset of all subject headings available in the data. On a list of 100 suggested subject headings, the recall is on average above 0.6 and for a list of 500 candidate subject headings, above 0.8. Even though the average recall stays noticeably below 1 (recall value 1 would mean perfect performance), at the same time it is still considerably above the chance level. The results presented in figure 2 were produced by the LDA model trained with 16 topics.

One of the parameters affecting the method performance is the number of topics used by the LDA model. For finding the number of topics corresponding to the highest recall, an overall measure of recall across different lengths of the subject candidate list was defined as the cumulative recall for first 100 subject candidates. We assumed that 100 is a likely size of candidate lists that catalogers would be willing to go through. Figure 3 shows the cumulative recall for different numbers of topics, based on which 16 were chosen as the optimum. Interestingly, this corresponds well with the perplexity dependence on number of topics (fig. 4). The perplexity, a measure of model's surprise at the data, shows how the model fits the data—a smaller number means a better fit, i.e., a better topic model.[28]

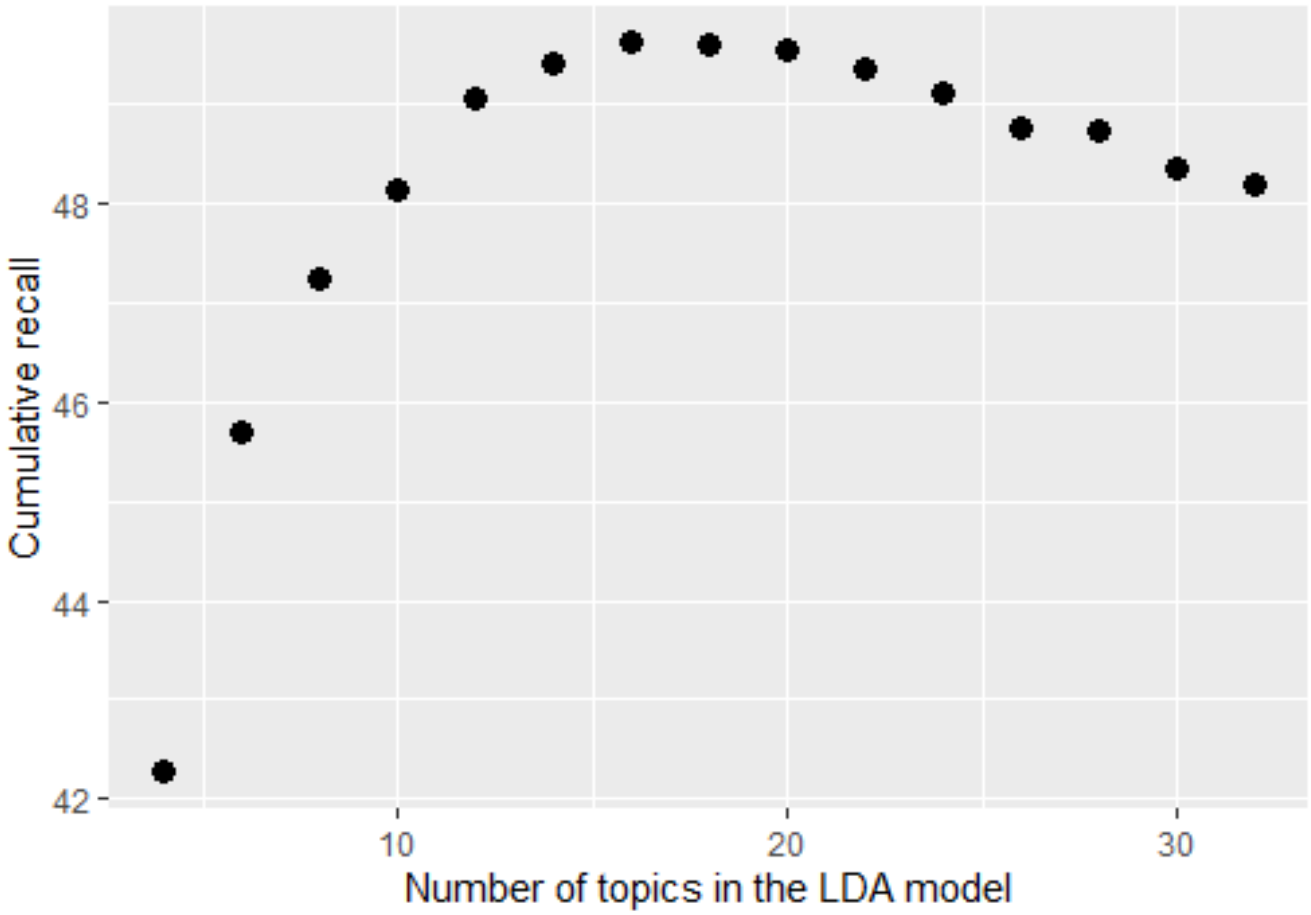

Figure 3. Cumulative recall as a function of number of topics in the LDA model.

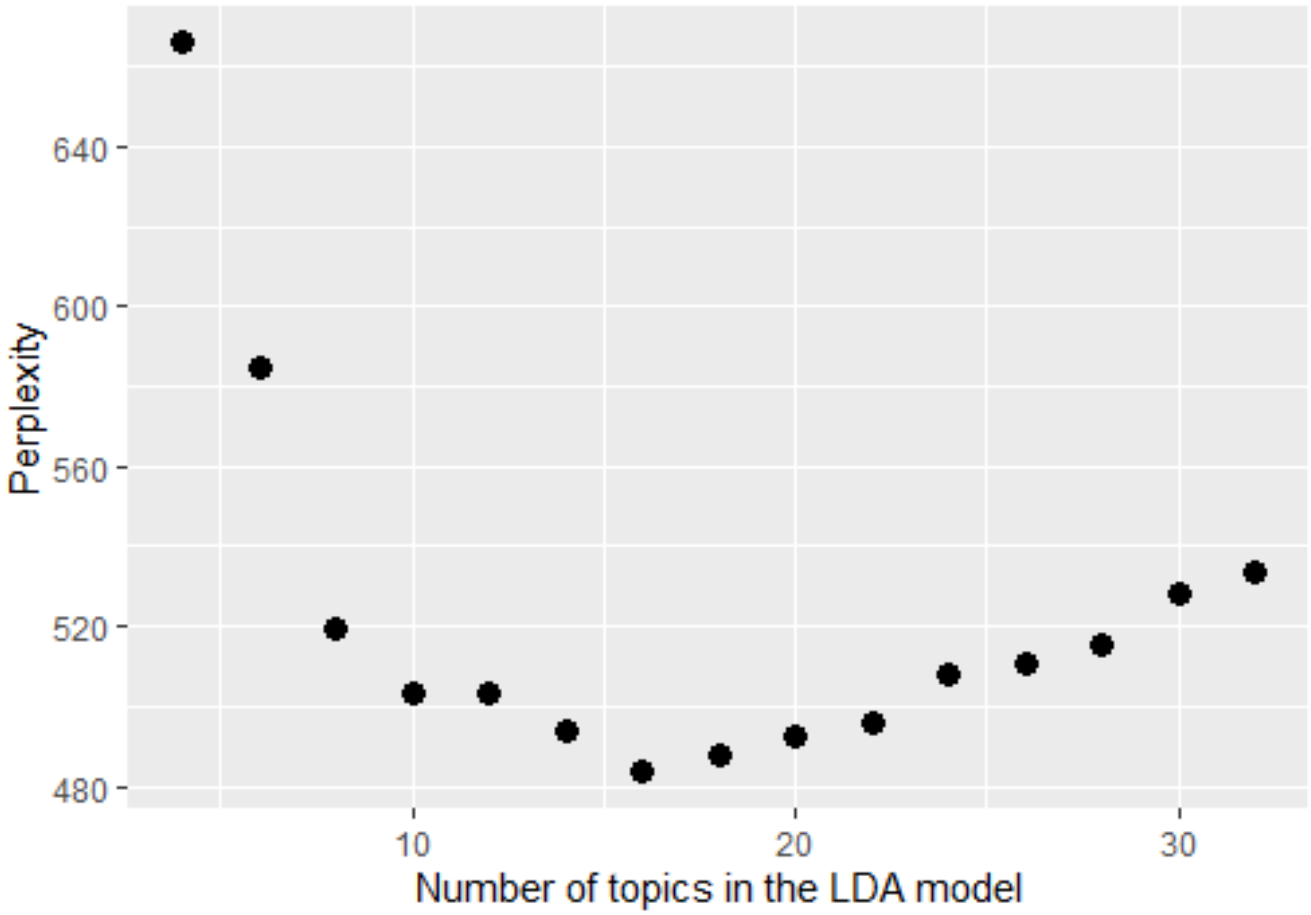

Figure 4. Perplexity of the LDA model as a function of number of topics in the LDA model.

To give a better idea about the method performance, figure 5 shows the distribution of recall for individual test documents, for a list of 100 subject headings. Since most documents in the training data have just a few subject headings, there is only a small set of discrete values possible for recall for individual documents. The distribution is wide, with a fraction of documents with no subject heading present on the proposed list (recall = 0) but also with a bigger fraction of documents fully covered by the list (recall = 1).

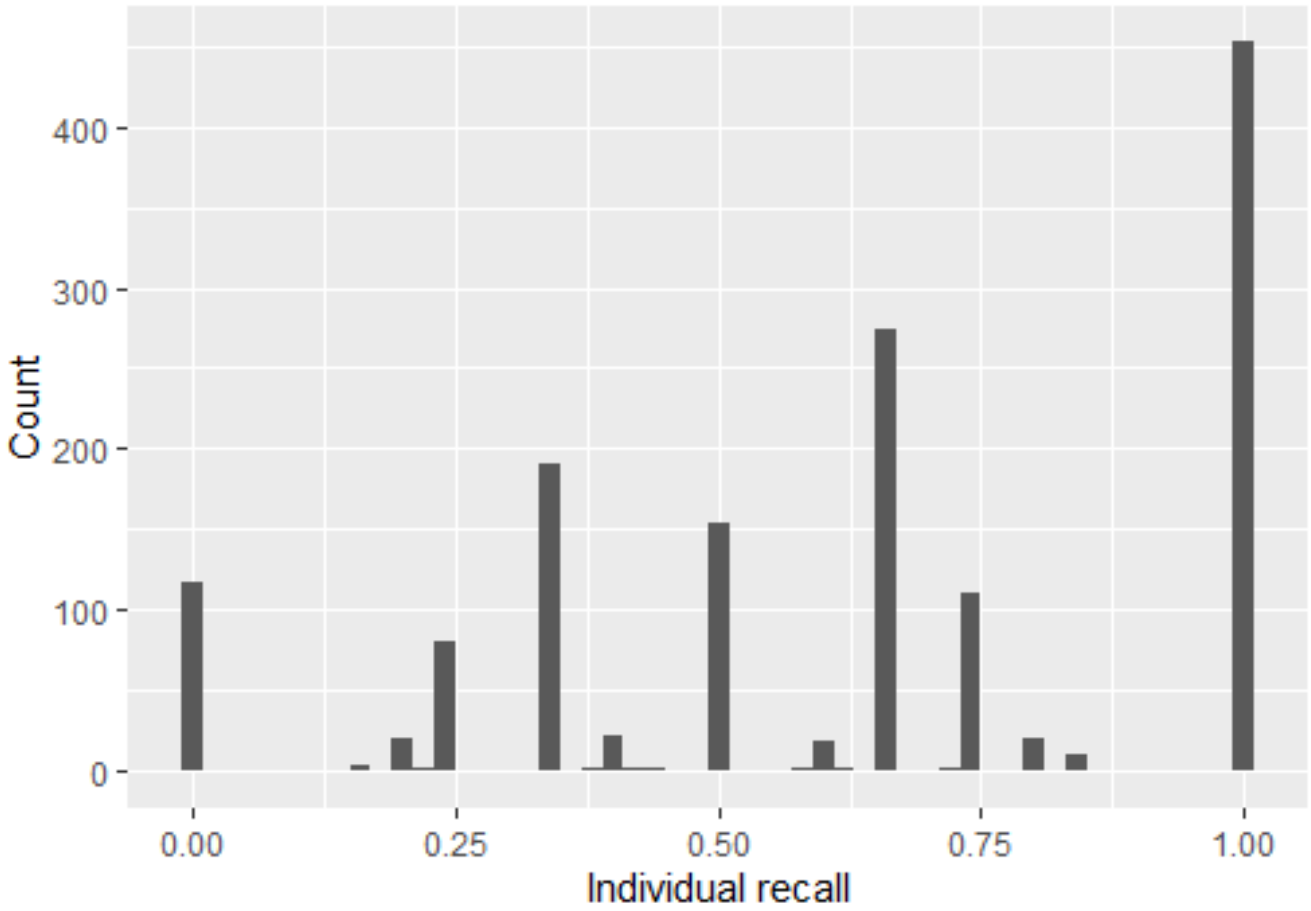

Figure 5. Distribution of recall across 1,500 test documents, for 100 subject candidates (for 16 topics).

The following examples show the sets of subject headings selected by the algorithm that include subject headings (in bold blue) chosen originally by catalogers.

Example 1
Title of news release: "'Romeo and Juliet' play - part of campus celebration for 400th anniversary of Shakespeare's birth"

Subjects	Weights
New Mexico State University. Playmakers	**0.280**
Theater	**0.143**
Students	0.080
Academic achievement	0.080
Theater--Production and direction	**0.075**
High school students	0.052
Competitions	0.048
New Mexico State University. College of Engineering	0.042
Plays	**0.041**
Debates and debating	0.038
New Mexico State University. Aggie Forensic Festival	0.036

Zohn, Hershel	0.034
Shakespeare, William, 1564-1616. A Midsummer Night's Dream	0.034
Forensics (Public speaking)	0.034
Frisch, Max, 1911-1991. Firebugs	0.027
Tickets	0.027
Theater rehearsals	0.027
New Mexico State University. College of Agriculture and Home Economics	0.022
Shakespeare, William, 1564-1616. Romeo and Juliet	**0.020**
Frisch, Max, 1911-1991	0.020
Performances	0.020
Garcia Lorca, Federico, 1898-1936. Casa de Bernarda Alba. English	0.020
Molière, 1622-1673. Bourgeois gentilhomme. English	0.020
Anniversaries	**0.014**
New Mexico State University. College of Teacher Education	0.012

Example 2

Title of caption to photo: "Locals Barbara Gerhard, Donna Herron, Lillian Jean Taylor rehearse for upcoming concert"

Subjects	Weights
Concerts	**0.123**
New Mexico State University. University-Civic Symphony Orchestra	**0.085**
INSTITUTION. Playmakers	0.077
United States. Air Force ROTC	0.073
United States. Army. Reserve Officers' Training Corps	0.062
Military cadets	0.058
Award presentations	0.054
Theater	0.039
Award winners	0.038
Scholarships	0.035
Music	**0.035**
Musicians	**0.031**
Awards	0.027
New Mexico State University. Department of Military Science	0.023
Theater--Production and direction	0.021
Kennecott Copper Corporation	0.019
Students	0.019
Glowacki, John	0.019
New Mexico State University Symphonic Band	0.015
New Mexico State University. University-Community Chorus	0.015
Lynch, Daniel	0.015
Drath, Jan	0.015
Performances	0.015

Military art and science	0.012
United States. Army--Inspection	0.012

DISCUSSION

The major advantage of the method described above is reducing a long list of Library of Congress Subject Headings that catalogers need to consult before assigning subject headings to news releases. It is important to note that this method produces subject headings that are already present in the training data. The list of available subject headings can be expanded by periodic updates of the training data to include all entries in the catalog, assuming catalogers will add, where needed, subjects not present so far in the data set.

In this project we utilized metadata from just two fields: titles and subject headings. Although documents' titles are supposed to compactly represent the content of documents, we expect that the presented approach would give better results if the full text (OCR) was analyzed. In this project, the limiting factors were both quality of print copies and robustness of available OCR tools.

In some cases, subject annotations are imperfect, depending on skills and experience of catalogers. That also affects the performance of our method that relies on quality of subject assignments. On the other hand, there are cases when the method suggests subjects that are fitting the content of news releases but were not selected by catalogers. This indicates that the method can also be used to refine the existing annotations.

CONCLUSION

We propose a way to streamline the workflow of metadata creation for university news releases by applying topic modeling. First, we use this digital technology to identify topics in a large collection of text documents. Then, we associate the discovered topics with sets of subject headings. Finally, to a new document, we assign those subject headings that are associated with the document's most dominant topics. The proposed method facilitates the process of document annotation. It produces short lists of candidate subject headings that account for a significant part of original labeling performed by catalogers. This approach can be applied to support annotation of any large digital collection of text documents.

One of the advantages of applying topic modeling is that it produces numeric representations of text documents. These numeric representations can be used by advanced analytical methodologies, including machine learning, for numerous practical purposes in library workflows like text categorization, collocation of similar materials, enhancing metadata for digital collections, finding trends in government literature, etc.

In addition, mastering digital methodologies by librarians may open new ways of collaboration among them and digital scholars across university campuses. As Johnson and Dehmlow argue, "... digital humanities represent a clear opportunity for libraries to offer significant value to the academy, not only in the areas of tool and consultations, but also in collaborative expertise that supports workflows for librarians and scholars alike."[29] Digital technologies are best learned in hands-on practice. If librarians are to contribute to the development of digital scholarship, then

they need to learn how to apply new technologies to their own work. And since both librarians and humanists work with texts, they might have much to offer each other.

ENDNOTES

[1] Anne Burdick et al., *Digital_Humanities* (Cambridge, Massachusetts: The MIT Press, 2012), 32–33.

[2] Thomas G. Padilla, "Collections as Data Implications for Enclosure," *ACRL News* 79, no. 6 (2018), https://crln.acrl.org/index.php/crlnews/article/view/17003/18751; Rachel Wittmann, Anna Neatrour, Rebekah Cummings, and Jeremy Myntti, "From Digital Library to Open Datasets: Embracing a 'Collections as Data' Framework," *Information Technology and Libraries* 38, no. 4 (December 2019), https://doi.org/10.6017/ital.v38i4.11101.

[3] Gerald W. Thomas, *Academic Ecosystem: Issues Emerging in a University Environment* (Gerald W. Thomas, 1998), 159–64.

[4] David M. Blei, Andrew Ng, and Michael Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research* 3, no. 1 (2003); David M. Blei, "Topic Modeling and Digital Humanities," *Journal of Digital Humanities* 2, no. 1 (Winter 2012), http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/.

[5] Megan R. Brett, "Topic Modeling: A Basic Introduction," *Journal of Digital Humanities* 2, no. 1 (Winter 2012), http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/; Jordan Boyed-Graber, Yuening Hu, and David Mimno, "Applications of Topic Models," *Foundations and Trends® in Information Retrieval* 11, no. 2–3 (2017): 143–296.

[6] Boyed-Graber, Hu, and Mimno, "Applications of Topic Models," *Foundations and Trends® in Information Retrieval* 11, no. 2–3 (2017): 143–296; Rania Albalawi, Tet Hin Yeap, and Morad Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Frontiers in Artificial Intelligence* 3 (2020): 42, https://doi.org/10.3389/frai.2020.00042; Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey," (2017), https://www.ccs.neu.edu/home/vip/teach/DMcourse/5_topicmodel_summ/notes_slides/LDA_survey_1711.04305.pdf.

[7] Zhijun Yin et al., "Geographical Topic Discovery and Comparison," in *WWW: Proceedings of the 20th International Conference on the World Wide Web* (2011), https://doi.org/10.1145/1963405.1963443.

[8] David Andrzejewski and David Buttler, "Latent Topic Feedback for Information Retrieval," in *KDD '11: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2011), https://dl.acm.org/doi/10.1145/2020408.2020503.

[9] Matt Erlin, "Topic Modeling, Epistemology, and the English and German Novel," *Cultural Analytics* 1, no. 1 (May 1, 2017), https://doi.org/10.22148/16.014.

[10] Cassidy R. Sugimoto et al., "The Shifting Sands of Disciplinary Development: Analyzing North American Library and Information Science Dissertations Using Latent Dirichlet Allocation," *Journal of the American Society for Information Science and Technology* 62, no. 1 (January

2011), https://doi.org/10.1002/asi.21435; David Mimno, "Computational Historiography: Data Mining in a Century of Classics Journals," *Journal on Computing and Cultural Heritage* 5, no. 1 (April 2012): 3:1–3:19; Andrew J. Torget and Jon Christensen, "Mapping Texts: Visualizing American Historical Newspapers," *Journal of Digital Humanities* 1, no. 3 (Summer 2012), http://journalofdigitalhumanities.org/1-3/mapping-texts-project-by-andrew-torget-and-jon-christensen/; Andrew Goldstone and Ted Underwood, "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us," *New Literary History* 45, (2014): 359–84; Carlos G. Figuerola, Francisco Javier Garcia Marco, and Maria Pinto, "Mapping the Evolution of Library and Information Science (1978–2014) Using Topic Modeling on LISA," *Scientometrics* 112, (2017): 1507–35, https://doi.org/10.1007/s11192-017-2432-9; Jung Sun Oh and Ok Nam Park, "Topics and Trends in Metadata Research," *Journal of Information Science Theory and Practice* 6, no. 4 (2018): 39–53; Manika Lamba and Margam Madhusudhan, "Metadata Tagging of Library and Information Science Theses: Shodhganga (2013–2017)," paper presented at ETD 2018: Beyond the Boundaries of Rims and Oceans Globalizing Knowledge with ETDs, National Central Library, Taipei, Taiwan, https://doi.org/10.5281/zenodo.1475795; Manika Lamba and Margam Madhusudhan, "Author-Topic Modeling of DESIDOC Journal of Library and Information Technology (2008–2017), India," *Library Philosophy and Practice* (2019): 2593, https://digitalcommons.unl.edu/libphilprac/2593.

[11] David J. Newman and Sharon Block, "Probabilistic Topic Decomposition of an Eighteenth-Century American Newspaper," *Journal of the American Society for Information Science and Technology* 57, no. 6 (April 1, 2006): 753–67; Robert K. Nelson, "Mining the Dispatch," last modified November 2020, https://dsl.richmond.edu/dispatch/about; Tze-I Yang, Andrew Torget, and Rada Mihalcea, "Topic Modeling on Historical Newspapers," in *LaTeCH '11: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (2011), https://dl.acm.org/doi/10.5555/2107636.2107649; Carina Jacobi, Wouter van Atteveldt, and Kasper Welbers, "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling," *Digital Journalism* 4, no. 1 (2015), https://doi.org/10.1080/21670811.2015.1093271.

[12] Jonathan O. Cain, "Using Topic Modeling to Enhance Access to Library Digital Collections," *Journal of Web Librarianship* 10, no. 3 (2016): 210–25, https://doi.org/10.1080/19322909.2016.1193455; Alexandra Lesnikowski et al., "Frontiers in Data Analytics for Adaptation Research: Topic Modeling," *WIREs Climate Change* 10, no. 3 (2019): e576, https://doi.org/10.1002/wcc.576.

[13] Tiziano Piccardi and Robert West, "Crosslingual Topic Modeling with WikiPDA," in *Proceedings of The Web Conference 2021 (WWW '21)*, April 19–23, 2021, Ljubljana, Slovenia (ACM, New York), https://doi.org/10.1145/3442381.3449805.

[14] Cain, "Using Topic Modeling to Enhance Access to Library Digital Collections," 210–25; A. Krowne and M. Halbert, "An Initial Evaluation of Automated Organization for Digital Library Browsing," in *JCDL '05: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, (June 7–11, 2005): 246–255; David Newman, Kat Hagedorn, and Chaitanya Chemudugunta, "Subject Metadata Enrichment Using Statistical Topic Models," paper

presented at ACM IEEE Joint Conference on Digital Libraries JCDL'07, Vancouver, BC, June 17–22, 2007.

[15] Craig Boman, "An Exploration of Machine Learning in Libraries," *ALA Library Technology Report* 55, no. 1 (January 2019): 21–25.

[16] Julia Silge and David Robinson, *Text Mining with R: A Tidy Approach* (Sebastopol, California: O'Reilly Media, Inc., 2017), 90.

[17] Blei, Ng, and Jordan, "Latent Dirichlet Allocation."

[18] Arlene G. Taylor, *Introduction to Cataloging and Classification*, 10th ed. (Westport, Connecticut: Libraries Unlimited, 2006), 19–20, 301–14; Arlene G. Taylor and Daniel N. Joudrey, *The Organization of Information*, 3rd ed. (Westport, Connecticut: Libraries Unlimited, 2009), 303–28.

[19] Blei, Ng, and Jordan, "Latent Dirichlet Allocation."

[20] Silge and Robinson, *Text Mining with R*, 149.

[21] Albalawi, Yeap, and Benyoucef, "Using Topic Modeling Methods for Short-Text Data," 42.

[22] The R Project for Statistical Computing, https://www.r-project.org/.

[23] Bettina Grün and Kurt Hornik, "topicmodels: An R Package for Fitting Topic Models," *Journal of Statistical Software* 40, no. 13 (2011): 1–30, https://doi.org/10.18637/jss.v040.i13.

[24] Topic Modeling in R (DataCamp), https://learn.datacamp.com/courses/topic-modeling-in-r.

[25] Grün and Hornik, "topicmodels."

[26] Topic Modeling in R (DataCamp), chap. 3, https://learn.datacamp.com/courses/topic-modeling-in-r.

[27] Christopher M. Bishop, *Pattern Recognition and Machine Learning* (New York, NY: Springer Science + Business Media, 2006), 32–33.

[28] Blei, Ng, and Jordan, "Latent Dirichlet Allocation."

[29] Daniel Johnson and Mark Dehmlow, "Digital Exhibits to Digital Humanities: Expanding the Digital Libraries Portfolio," in *New Top Technologies Every Librarian Needs to Know*, ed. Kenneth J. Varnum, (Chicago: ALA Neal-Schuman, 2019), 124.