



Exploring Reader-Generated Language to Describe Multicultural Literature

Denice Adkins, University of Missouri, USA
Jenny S. Bossaller, University of Missouri, USA
Heather Moulaison Sandy, University of Missouri, USA

Abstract

How do readers describe multicultural fiction works? While in library and information science (LIS) we have the language of *appeal factors* and *genre trends* to describe works of fiction, these linguistic choices may not be used by readers to describe their own responses and reactions to works that provide cultural affirmation of one's own culture or exposure to learning different cultures. In this research, text mining processes are employed to harvest reader-generated book reviews and subsequently analyze the words readers use to describe award-winning multicultural fiction on the retailer site Amazon.com. Our goal with this study is to provide LIS professionals an insight into readers' perspectives related to multicultural fiction. We describe our methodology of engaging in topic modeling as described by Jockers and Mimno (2013) as applied to multicultural fiction reviews. First, we explore the construction and processing of a corpus of reader reviews of multicultural fiction titles, then we model topics using a topic modeling toolkit to generate topics from these reviews. Through this analysis, we determine consistent terms used to describe multicultural fiction that can be used to indicate common reader experience and identify topics. Closing discussion reflects on whether librarians can use text mining of reader reviews to enhance their reader advisory services for readers seeking books that represent multiple and/or diverse cultures.

Keywords: Amazon reviews; appeal factors; multicultural fiction; multicultural literature; topic modeling

Publication Type: research article

Introduction

We are living in an era of socio-cultural movements that affirm identity in various ways. Movements are on the streets and online with hashtag identifiers such as #BlackLivesMatter, #EverydaySexism, #LaGenteUnida, and #AmINext, to name a few. Readers and librarians have entered this social discourse with social media identifiers such as #BlackBooksMatter, #1000BlackGirlBooks, #WeNeedDiverseBooks, and #ReadWomen. Additionally, there have been many other calls to diversify and celebrate books that chronicle the lives of people who are often insufficiently represented in library stacks. Given the responsive demands that these various social movements infer, we ask the question: How can librarians increase their ability to locate and select culturally relevant and authentic fiction? This paper describes a method of analyzing the words written by readers of diverse literature in order to discern themes and characteristics that readers find appealing. Findings can be used by librarians



to enhance their collections and their fluidity in discussing books with readers as well as with publishers, who serve as gatekeepers.

This paper begins with a question: How do readers describe multicultural fiction works? For the purposes of this paper, “multicultural literature” is used within the context of the U.S. publishing industry to describe literary works produced by or intended for audiences who come from African American, Hispanic/Latinx, Native American, or Asian American backgrounds. We acknowledge that these are broad classes that incorporate multitudes of diverse experiences, and we also assert that this paper serves as more of a model for a method than as a paper that includes definitive results about readers.

In this research, we use text mining processes to explore the words and phrases that readers use to describe multicultural literature. Our goal is to provide library and information science (LIS) professionals an insight into readers’ perspectives related to multicultural fiction. Although in LIS we utilize professional terms such as *appeal factors* and *genre trends* to describe works of fiction, these linguistic choices may not be used by readers to describe their own responses and reactions to works that provide cultural affirmation or potentially expose readers to different cultures. By using a new method of exploring readers’ descriptions of reading experiences, we hope to determine (a) whether there are consistent terms used to describe multicultural fiction that can be used to indicate common reader experiences; and (b) whether librarians can use text mining of reader reviews to enhance their reader advisory (RA) services for readers of multicultural fiction.

Literature Review

For years, librarians have sought methods to connect readers to books. For example, directed reading programs during the first real RA push of the 1920s suggested books that might uplift readers’ spirits or elevate their taste (Saricks, 2005). Lawrence (2017) summarizes the trends in RA, from the directed RA interviews of the 1920s in which librarians produced customized bibliographies for the patron, to its mid-century decline and its renaissance in the 1980s. Dilevko and Magowan (2007) point out many reasons for the reemergence of RA in the 1960s: technical education was on the rise, and libraries offered a chance to bolster liberal education for the masses, who also needed bibliotherapy because of societal problems. The study of popular culture gave librarians license to collect and recommend high-demand books over serious reading. Another important event in modern RA was the 1982 debut of Rosenberg’s *Genreflecting*, a guide to popular reading that told readers to, “never apologize for your reading tastes” (p. 5). This publication gave librarians a new and acceptable means of connecting readers to the books that they really wanted to read.

The library’s collection is most often built around books from mainstream publishers in the languages that will best serve the most people in the service area. There are many gatekeepers, then, at different levels, beginning with authors, in terms of what is written; publishers, in terms of how and where books are published; and the librarian themselves, due to the fact that RA services are limited to people who want to talk to the librarian. Minority (e.g., immigrants, African Americans, Latino/Latina, LGBTQ) readers might feel as if librarians do not speak their language. Sometimes they do not (literally), but sometimes there is a cultural barrier. We think that librarians can overcome some cultural barriers, given the right tools.

Inclusion and Exclusion: What is Published?

An assumption exists in the literature that large publishers like the big five (i.e., Hachette, HarperCollins, Macmillan, Penguin Random House, and Simon & Schuster) are not embracing diverse adult fiction; indeed, the publishing industry has a reputation for being overly white in its literary and authorial representation (Doherty, 2016), including the smaller presses (cf. Lee & Low Books, 2016). Doherty (2016) explores this phenomenon in her study of four indie publishers' responses to the #WeNeedDiverseBooks campaign. As a result, she found that the lack of representation caused those publishers "to ask for books by writers of color more openly than before and make diversity top priority in their slush pile by tagging writers of color" (p. 5), implying that indie publishers might be more inclined to seek out diverse fiction.

The 2015 Diversity Baseline Survey (DBS), conducted by Lee & Low Books, examined gender, race/ethnicity, sexual orientation, and disability in the publishing industry. The researchers found that the publishing industry is hiring new people (although the average age of employees is down, along with their compensation), but that a steady 89% of the employees are white and 77% are women. However, the researchers did find that publishers were advancing more diverse and multicultural books (Milliot, 2015). The DBS collected data on employees in the publishing industry and also on book reviewers. The data on book reviewers is important because librarians and other information professionals rely on book reviews to support their collection development efforts. The DBS's outcomes suggest that if publishers, editors, and book reviewers are predominantly white, cis-gendered, heterosexual non-disabled women (as the DBS showed), then the perspectives that determine what enters the distribution channels may be limited as well (Lee & Low Books, 2016).

This concern is not new; in 1965, Nancy Larrick wrote about "The All-White World of Children's Books" (Larrick, 1965). Noted LIS youth services scholar Sandra Hughes-Hassell confirms that, "the need for multicultural literature is even greater today than it was in 1965" (2013, p. 212). As 21st century demographics in the U.S. continue to skew towards a majority of non-white youth who will grow into the mainstream adult population, the need to see books that appreciate citizens' specific heritages and cultures will become requisite. The 21st century will continue to see the growth of nonwhite populations, and children's literature should reflect such demographic trends (p. 215).

Appeal Factors and Multicultural Literature

In 1989, Joyce G. Saricks and Nancy Brown introduced the concept of appeal factors that could be used to help readers find books that match their interests and tastes. "Through trial and error, these two practitioners developed a method for thinking about books in terms that mattered to readers" (Smith, 2015, p. 13). These appeal factors included pace, characterization, story line, and frame. Appeal factors were meant to be a "universal language," so to speak, that would allow librarians and readers to classify works of fiction based not on characters and plot so much as factors that draw links between literature and genres. Saricks (2005) affirms that "most readers are not usually looking for a book on a certain subject. They want a book with a particular 'feel'" (p. 40).

Using appeal factors has thus become the de facto method that librarians use to connect readers with books, but some authors have attempted to expand the relevance of appeal factors. For instance, Dali (2013) cites a "generally growing dissatisfaction with the original definition and

traditionally outlined scope of appeal” (p. 475). She mentions additions to the concept of appeal factors, such as the inclusion of linguistic style, learning potential, format, genre, and subject. Dali’s (2014) expansion of appeal factors includes both book appeal factors, as indicated by Saricks and others, reader-driven appeal factors such as a reader’s curiosity, the process of reading, and the role of reading in a reader’s life (pp. 32-33). In Dali’s (2013) summary of discussions of appeal factors, she notes that authenticity has been mentioned as a book appeal factor, as has connecting with characters. Topics such as “cultural empathy and understanding,” “recognition [of self] in books,” “reading about people who are similar to us and different from us,” and “confirmation of readers’ own . . . experiences” are classified as appeals beyond the book (Dali, 2013, pp. 484-485). Dali suggests that readers may want to explore multicultural fiction to expand their intercultural connections with others (2014, p. 28).

One of the most popular RA tools for librarians is the classic *Genreflecting* series, originally written by Betty Rosenberg in 1982, now by Diana Tixier Herald and other authors and in its seventh edition. How well does *Genreflecting* cover multicultural fiction? The sixth edition (2006) covers very little in the way of multicultural fiction. A review of the work finds sections on “African Americans in the West” (pp. 116-117) under historical fiction, “Diversity in Detection” (pp. 173-175) under crime, a brief review of ethnic romance, and another brief review of Asian fantasy. As noted by Dali (2014), appeal factors are intended to be a universal language to describe fiction; what unites, rather than what separates. Saricks (2005) explains that this was a conscious choice, speaking to the universal nature of the elements of appeal:

In developing a list of popular genres, there is always the temptation to create genres from groups of authors who are really linked in other ways, perhaps by subject or even gender, rather than genre. African American authors, for example, do not constitute a separate genre, since books by these authors are not written to a particular, identifiable pattern as by definition genre fiction always is. Toni Morrison and Mystery writer Eleanor Taylor Bland are both African American writers, but each has her own following of fans, and their books differ dramatically - they belong on the Mystery genre list, not a separate list that merely groups authors by race or gender. (p. 32)

Despite the relatively small amount of coverage of multicultural fiction in *Genreflecting*, there are a number of other American Library Association (ALA) RA guides that cover many fiction genres that include chapters (or entire books) about multicultural fiction. For instance, there are guides to romance, mystery, historical fiction, and street lit. These books discuss who reads the genre and why it is appealing for specific types of readers, alluding to universal appeal factors.

RA Tools for Multicultural Materials and Audiences

Alma Dawson and Connie Van Fleet (2001) discussed the need for increased RA services for multicultural audiences. They noted increased mainstream interest in multicultural literature, defined as “the literature about persons or groups that differ in some way (ethnically, racially, culturally, linguistically, by sexual orientation, or disabilities) from the sociopolitical Euro-American mainstream of the United States” (p. 250). They posit a pattern of development among literature where early writings focus on self-definition and identity, later writings on the transition between cultures.

In 2004, Dawson and Van Fleet published the first RA guide focused on African American literature. That publication covers the same genres as other RA guides (e.g., crime/detective fiction, romance, inspirational fiction, etc.). *Latino Literature: A Guide to Reading Interests* was published in 2009 as part of the *Genreflecting* series (Martinez, 2009). Like the RA title for African American Literature, *Latino Literature* also focuses on popular and traditional genres such as mystery, historical fiction, and romance, specifying works written by Latinx authors or relevant for Latinx readers. Editor Sara E. Martinez includes some key concepts specific to Latinx books, including “themes of immigration, political upheaval, the refugee experience, and the search for cultural identity” (Martinez, 2009, p. 84).

A Readers’ Advisory Guide to Street Literature (2012) written by Vanessa Irvin Morris focuses specifically on the literary genre known as urban fiction or street lit. In the forward of the book, street lit author Teri Woods discusses how her first novel, *True to the Game*, was repeatedly rejected by publishers, so she decided to publish it herself. After selling her novel firstly to local then national bookstores, Woods established her own independent publishing company to promote her own and others’ urban fiction works. Woods’ resounding success alerted established publishers to the lucrative possibilities of street lit. Woods’ experience as an author coupled with her experience working with inner-city teen readers provides the knowledge to describe more literary appeal factors—such as relationships, identity, accurate representation of street life, and pace—particular to the street lit genre.

Simone Gibson (2010) explains that many African American girls want to read street lit and that their literacy rates are higher than school records might indicate, but that their reading choices are not valued in school. Sandra Hughes-Hassell and Pradnya Rodge (2007), likewise, described a “literacy ceiling” (p. 22) that many young people reach in middle or high school, which might be broken if students are given the opportunity to engage in more (and more personally appealing) leisure reading. Street lit is a valid example of how the reading experience of self-selected titles helps youth readers to make sense of their lived worlds; readers “learn from their experiences by the conclusion of the story, passing along advice that results in the formation of a cautionary tale” (Gibson, 2010, p. 567). The books can provide a guide for life and substantial escapism, but many educators will not use the books because of “vulgar themes, nonstandard language use, stereotypical portrayals, . . . [and] poor writing construction of many of the texts” (p. 569). Fortunately, librarians can provide street lit novels and connect with the genre’s readers because they do not have the same pedagogical constraints as classroom teachers. However, Irvin Morris found that within days of leaving her position at the Free Library of Philadelphia, her carefully curated (and popular) street lit section had been dismantled, reflecting a lack of knowledge on the part of the librarian who took over the collection of its appeal to readers (or possibly a prejudice against its content). Irvin Morris is doubtlessly not alone in her quest to curate such collections.

Of course, African American girls are not alone in a quest for a good book. What do we know about fiction’s appeal to immigrants? Dali (2010) points out that public librarians are very good at services for immigrants in two areas: “coping skills (e.g., ESL; basic literacy . . . citizenship and exam preparation . . .)” and “arts and culture (e.g., library programs aimed at celebrating ethnic heritage . . . etc.)” (p. 215). However, she found that librarians are not as good at connecting immigrant users to leisure or pleasure reading. While Dali (2010) admits that almost no public library is going to have a multicultural collection that rivals ethnic bookstores, librarians can best collect and manage relevant collections by asking immigrants what they like to read in their own language and then connecting the readers to books in English that have

similar themes. Dali's main point is that librarians can connect to readers by listening and acknowledging their expressed reading interest, tastes and needs, and by making liberal use of electronic resources like the RA database, *Novelist*, for high caliber RA services. *Novelist*'s conceptual framework largely draws on appeal factors, but the database also has user guides to direct readers on how to find books by culture and diverse characters. Dali notes that her interviews with immigrants suggests that appeal factors may be "applicable irrespective of the ethnic and linguistic origin" of the reader (2014, p. 24).

Overcoming Publishing Problems

In the *Guide to Street Lit* (Irvin Morris, 2012), readers reported that one of the things that draws them into a story is authenticity of the characters, of the portrayal of difficult lives and situations, and of the cultural language used by the characters (i.e., not Anglo-American English). The reader can relate to the characters; they have seen (if not experienced) similar situations. This relatability harkens back to Louise Rosenblatt's theory of reader response (1938) and to Janice Radway's (1984) study of romance readers: people want stories that they can relate to.

However, it is a lack of authenticity that Sanchez (2014) writes about regarding protagonists and characters of color in crime fiction (mysteries), explaining that "English has become the lingua franca of world literature" (p. 1). Good intentions in promoting world literature are not enough; librarians must also be aware of English and Western bias in publishing. Sanchez explains that much "ethnic detective fiction" is actually written by white Anglos, that the voices of the characters are inauthentic, and that the characters are presented as "the other" or exoticized. The etic voice is present because it is more palatable or more relatable to the publishers. Similarly, Poddar (2016) discusses gatekeeping by publishers, who promote "false exoticism" in postcolonial fiction to perpetuate cultural tropes. Poddar is specifically talking about diasporic writers who write to appeal to mainstream publishers, rather than cultural appropriation by Anglo writers.

Bryoni Birdi and Mostafa Syed (2011) found that the items being published were a barrier to minority readers in England. Citing a BBC interview regarding Muslims in fiction, one Muslim convert pointed out that "publishers...are reluctant to commission...novels which portray Muslim cultures positively, since they felt, as one publisher put it, that readers would be 'confused,' and the book would not sell" (p. 3). Birdi and Syed (2011) also found that very few people even came to the library looking for Black British or Asian fiction in English, or Gay/Lesbian fiction. They suggested using themed displays to "encourage people to find the elusive 'good read', ... remove fears and prejudices in an entirely unobtrusive way, to present wider reading choices to all library users" (p. 18). In her later work, Birdi (2014) further explored minority ethnic English language fiction to identify reader types. Using a Likert scale to measure attitudes and opinions, she found that readers of minority fiction were more likely to be members of a minority group. While readers of minority fiction are more likely to be minorities, as Birdi pointed out, themed displays of minority fiction can draw all readers in, encouraging them to find the "elusive 'good read'" (Birdi & Syed, 2011, p. 18) while removing prejudices about minorities.

Again, appeal is the most common method that librarians use to talk about fiction reading, despite limited coverage or terminology to help readers find multicultural works. Are there ways to expand this limited terminology? We suggest going directly to the readers' voices via online reader-generated book reviews. This source is not entirely new; for instance, Wanda Brooks and Lorraine Savage (2009) used Amazon book reviews to qualitatively assess appeal and reader

response to street lit. They found that readers enjoyed the relatability of the characters, and that the books really drew them in; readers were unable to put the books down. Our research, instead, takes a bird's-eye view of the readers' descriptions, using a quantitative text analysis approach to find common words and themes in reviews of multicultural fiction. We used a corpus of reader reviews drawn from award-winning fiction for this experiment, but plan to go deeper into independently published works from other genres (e.g., LGBTQ, manga, etc.) for future studies; this demonstrates how the model might be employed while providing an analysis of award-winning fiction.

Topic Modeling as Method of Analysis

Topic modeling is an approach that allows researchers to analyze large bodies of text in a way that is relatively straightforward. Saxton (2018) points out that topic modeling can be performed on any kind of documents, as long as the documents are electronic. Xiao, Ji, Li, Zhuang, and Shi (2018) use topic modeling to analyze online reviews in conjunction with other indicators to predict users' ratings of consumer products on two retail websites. According to Jockers and Mimno (2013), who analyzed a large corpus of nineteenth century English literature, there are a number of benefits to this method, including the ability to study a corpus much larger than possible for a single person to read. A second benefit is that topic modeling allows researchers to establish themes using the words of the source—as a result, these themes have the potential to be complex, instead of the somewhat reductionist themes that might otherwise be produced as a result of overgeneralizations and personal bias after reading. Our project uses topic modeling as a method of analysis—by selecting a large corpus of reader reviews of multicultural works of fiction in the words of the readers, and then uses this corpus to build topic models as a way of testing the questions of the terminology used by readers of multicultural fiction to describe their experiences.

Method

The premise of this project was to determine whether reviews of multicultural books would reveal user-generated language to indicate the books' appeal. This research took place in three phases: developing the corpus, analyzing said corpus by using topic modeling techniques, and assessing the results. We present the first two phases as part of our methodology, and the results and their assessment as part of our data analysis.

Developing and Processing the Corpus

A list of 50 award-winning multicultural adult fiction titles was generated by our collaborative reviewing of major literary award lists such as the National Book Award, the PEN/Faulkner Book Award, and other award lists that covered a period of 10 years (2008-2017). The choice of books was based largely on our need for testing the limits of the method of analysis in this exploratory work: limitations included books that were relatively recent and that would have a large number of reader reviews and commentary available for data mining. We chose to look at 50 books to ensure that multiple works were available for each of the cultural groups involved in the study (African American, Hispanic/Latinx, Immigrants, Asian American). Reviews for those titles were obtained from Amazon.com customer reviews, a known source of reader information that could serve as a limited public forum where reviewers had some freedom to express their authentic reader responses.

Review data was scraped using the Webscraper.io extension to the Chrome web browser and downloaded as a comma-separated values file. Each file consisted of a collection of reviews for a particular book title, and each file was processed individually. The text of the narrative reviews was exported into a plain-text file.

Punctuation and general stopwords were removed from each file, and the most frequently-used words were reviewed to determine if any specific stopwords might need to be removed. After deliberation, it was decided that authors' and characters' names should be removed from the corpus. After removal of more specific stopwords, reviews were tokenized (split up into individual words) so that each file could be processed individually, and all files could be processed collectively.

Analyzing the Corpus

The first stage of analysis involved using Voyant Tools, a web-based text analysis tool, to review our existing corpus. Voyant Tools is available at <http://voyant-tools.org/>, which enables users to do basic text analysis using a “bag of words” format, analyzing words individually rather than contextually. We used Voyant as an entry point into our data analysis, looking at some basic topics such as word frequency and word collocation. We analyzed the entire corpus file for the most frequently used words in the reviews.

The second stage of processing involved analyzing the corpus using the MALLET Topic Modeling toolkit. MALLET is a downloadable Java application available at <http://mallet.cs.umass.edu/index.php>. MALLET uses the graphical distribution topic model called Latent Dirichlet Allocation (LDA) to look at word use in context and create topic clusters based on the text. That is, LDA looks at the words used in each combination of documents (e.g., all the words used in the reviews of Book 1, all the words used in the reviews of Book 2, etc.). Based on those words, MALLET uses statistical algorithms to infer the topics of the documents based on the frequency with which various words are used together, producing “meaningfully ambiguous” results that guide researchers toward making their own judgments (Underwood, 2012, para. 16). When studied over time, these word relationships may signal changing discourse (e.g., Underwood, 2012) or societal change (e.g., Nelson, n.d.). The value of topic modeling is that “[statistical topic] models do not require annotated training data and do not attempt to analyze linguistic structures, they are simple to run and robust to variation in language and data quality [and] scale to large data sets” (Jockers & Mimno, 2012, p. 2).

We ran MALLET against the entire corpus file, looking at the top five, 10, and 20 topics indicated. Looking at these smaller numbers allowed us to analyze the corpus reductively, as we found that looking at 50 and 100 topics resulted in analyses of individual titles' reviews rather than the whole corpus.

Reviewing the topic models generated, we opted to run Stanford CoreNLP Named Entity Recognition (software that analyzes text looking for named entities and then extracts those entities into a list) against the file to determine which named entities were mentioned in the reader reviews, such as author names, country names, and character names. Available for free download at <https://stanfordnlp.github.io/CoreNLP/>, Stanford CoreNLP offers several features for text analysis, including tokenization (splitting the corpus into individual tokens, in this case, words), lemmatization (creating lemmas, or base forms, for words to allow searching with truncation), identification of parts of speech used in a text, and identification of named entities

such as person, location, and organization, among others (Manning, Surdeanu, Bauer, Finkel, Bethard, & McClosky, 2014). Author names and character names are typically “specific only to the text in which they occur” (Jockers & Mimno, 2012, p. 5), and so may increase noise in the analysis. This is a concern that Underwood (2012) also notes, with discussion about the nature of authorial “signatures” in their word choice and topic selections.

Author and character names recognized by Stanford CoreNLP’s Named Entity Recognition (NER) program were added to the stopwords list in order to get a better sense of the words used by readers to describe and review these books. While the NER function did a good job recognizing traditional American names, it was less reliable in recognizing ethnically-derived names, such as “Isabel” and “Nguyen.” Therefore, it is possible that some character names are included in the reviews analyzed. After removing stopwords, we ran Voyant-Tools and the MALLET Topic Modeling Toolkit again, using these results for our final analysis.

Results

We downloaded 33,178 reader reviews for the 50 books in our data pool. The number of reviews per book ranged from 11, for *I Hotel* by Karen Tei Yamashita, to 4,178, for *The Underground Railroad* by Colson Whitehead. Prior to stopword removal, the file contained 2,077,474 words; after stopword removal, the file contained 1,092,786 words, and after removal of authors’ and characters’ names, 856,998 words.

Table 1. Twenty most frequently used words in review corpus

Word	Frequency of Use
book	28,024
read	16,488
story	14,834
characters	8,449
novel	7,070
great	6,310
stars	6,200
like	6,040
good	5,894
life	5,635
written	5,175
love	4,879

Word	Frequency of Use
time	4,579
reading	4,511
just	4,388
family	4,266
writing	4,009
author	3,846
interesting	3,651
really	3,634

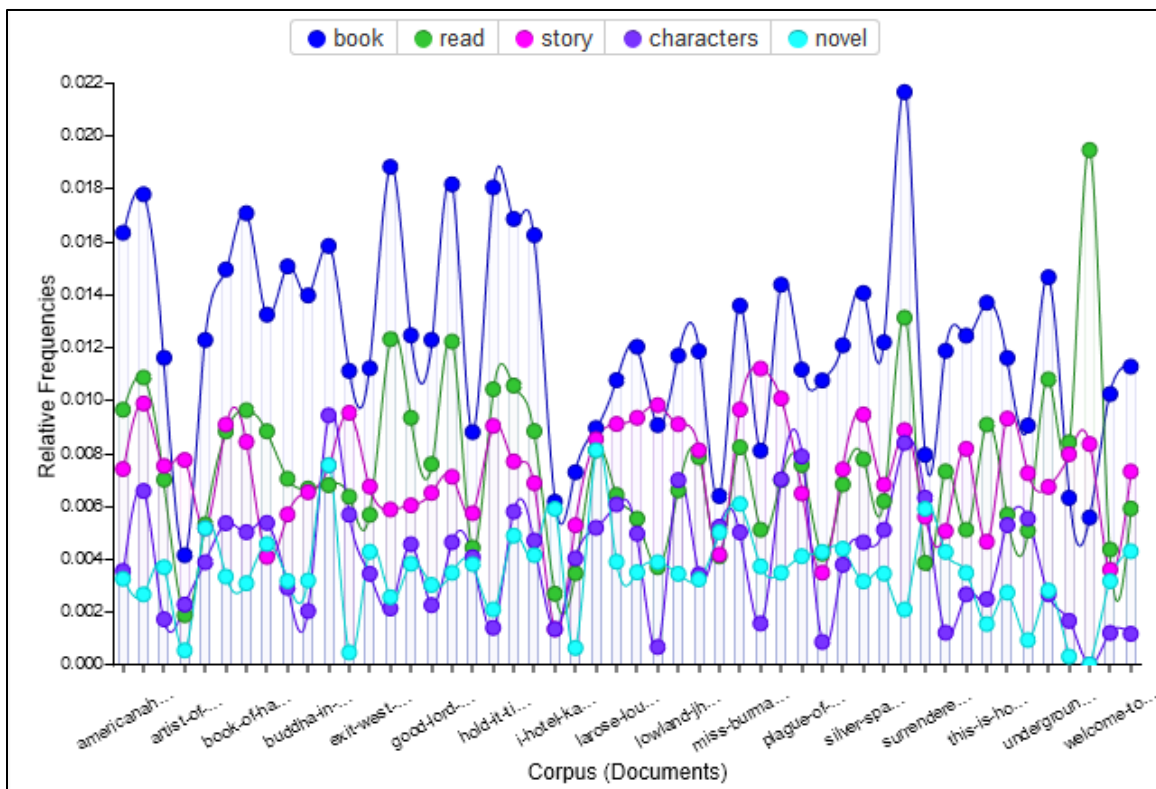


Figure 1. Voyant distribution of the five most frequently used words in the corpus, based on all 50 titles

Characteristics of User-Generated Review Texts

This corpus is comprised of reviews written for the most part by layperson readers. Some reviewers for Amazon are semi-professional, in the sense that they receive books for free in exchange for their reviews. As of 2016, Amazon has banned reviewers receiving financial compensation for their reviews (Perez, 2016), but based on web articles with titles like “Confessions of a paid Amazon reviewer” (Chen, 2017), it is possible that some of these reviewers have received money or goods in exchange for their reviews. Nonetheless, these reviewers hold a variety of opinions about these books.

Table 1 shows the most frequently-used words in the corpus after the removal of common English-language stopwords such as “the” and “and.” Several words were related to the reading experience or process, such as book, novel, and story. These results were generated from Voyant-Tools, using their standard list of stopwords.

Figure 1 shows the distribution of the five most frequently used words throughout the corpus: book, read, story, characters, and novel. The word “book,” in dark blue, is used most frequently across all 50 titles, with “read” (green), and “novel” (light blue) also describing the reading experience and supporting Dali’s assertion that the process of reading factors into reading motivations. The term “characters” (purple) supports characterization as being an important appeal factor for readers. The term “story” (pink) may support story line as an appeal factor.

Processing the corpus through Stanford CoreNLP indicates the parts of speech used in reviews and the types of concepts addressed. CoreNLP finds 2,077,970 words total in the data set. Of those, 493,800 (24%) are nouns, 351,521 (17%) are verbs, 223,078 (11%) are adjectives, and 127,058 (6%) are adverbs. These words are used to describe the story, elements within the story such as characters and plot, readers’ reactions, and readers’ thoughts related to the book.

The NER function identifies people’s names, but it also identifies words describing general concepts. For instance, NER identified 21,066 (1%) words as being related to place - names of countries, cities, and states, as well as identifiers of nationality such as American or Vietnamese. There were 3,549 words in the category “Cause of Death,” such as “war,” “disease,” and “violence,” while 1,273 words fell into the “Criminal Charges” category, such as “murder,” “terrorism,” and “genocide.” Another 1,224 words related to ideology, including “middle class,” “family values,” and “feminism,” while only 395 words were deemed to relate to religion, including “Judaism,” “Islam,” “Christianity,” “Hinduism,” and “atheism.”

Topic Models

Before running topic models with MALLET, we removed four of the five most frequently-occurring words (“book,” “read,” “story,” and “novel”) because we felt that, while they are important, they are not important to distinguish anything useful for the purposes of this paper; they were a way for users to say what they were reading, and were therefore more like stopwords. We then ran a model to generate 20 topics, 10 topics, and five topics, based on words that have a tendency to co-occur in the same texts. Jockers and Mimno found that analyzing a full novel resulted in topics that were too broad and ended up using novels segmented into 1000-word chunks (2013, p. 754). For our files, the average number of words in the reviews for each title is 41,551, ranging from a low of 360 words for *The Water Museum* by Luis Alberto Urrea, to a high of 195,758 for *Americanah* by Chimamanda Ngozi Adichie.

Table 2 shows the 20-topic model. The column labeled “Interpretation” contains our general interpretation of the words identified by MALLET as representing the words most frequently used together. Several words are repeated in multiple categories, suggesting that topics are wide-ranging and might be refined. This 20-topic model demonstrates some of the meaningful ambiguity noted by Underwood (2012, para. 16), including words that do not seem immediately relevant to a specific topic. This 20-topic model demonstrates the need for further stop word refinement, as evidence of book titles (such as “Forty Acres,” “The Sympathizer,” and “House of Spirits”) are all apparent in these word choices. Further, though we attempted lemmatization (word stemming) with StanfordNLP, this function did not work effectively, as demonstrated with the grouping of words such as “vietnam” with “vietnamese” and “american” with “americans.” In interpreting the nature of topics identified, we tended to group general concepts together and ignore the words that seemed particularly out of place, such as “sparrow” and “sag” in the Family and Childhood row. However, this indicated a need for refinement.

Table 3 shows a reduced topic model, returning 10 instead of 20 topics. To some degree, these topics reflect the topics above, but start becoming more general.

Table 2. Twenty topic models with interpretations and the top 20 co-occurring words for each topic

Interpretation	Topic Model Words
African American, music and history	jazz thriller black acres forty war music half paris history blues band blood musicians braggsville louis german african group plot
Characterization	written people world interesting life time part tale plot place told events fiction important journey telling descriptions takes narrative experience
Readers’ feelings	time character reader work hard pages readers beautifully difficult give left times words day woman young novels kind chapter wanted
Family and childhood	family mother father girls age coming brother parents growing summer boys silver brothers girl childhood friends sparrow harbor secret sag
Immigration	immigrant family immigrants immigration doors country dream interesting good people city west refugees home america exit dreamers cameroon timely war
Family	family hurricane sing esch bones salvage father mother dog mississippi love poverty china kayla pop unburied ghosts children brother characters
Vietnam	war vietnam vietnamese sympathizer perspective written narrator american south spy view interesting general communist saigon north man excellent end june
Format (short fiction)	short collection brooklyn beautiful august animals pakistan family young narrator novella novellas cards poetry father mexican debut final loteria

Interpretation	Topic Model Words
	death
Violence (tone)	characters history violence jamaica jamaican pages hard political language difficult violent understand killings york challenging patois graphic dialect interesting english
Readers' feelings	great good author found long bit made recommend style dont worth back lot find time make authors makes times full
Family	family house characters spirits detroit chile love supremes history great good native loved women friends enjoyed wonderful notebook magical generations
Asian American experience	india japanese lowland interesting indian brothers history short women style lives written america family calcutta brides namesake experience buddha brother
Characters' perspectives	american women point voice unique narrative reader person thought man subject early view small perspective told order simply experiences men
African American experience, immigration	race america love american nigeria african black nigerian great interesting culture experience written loved perspective immigrant blog racism enjoyed americans
East Asian families	family history japan korean characters generations japanese koreans interesting historical korea african pachinko great saga written slavery homegoing lives generation
African American historical fiction	railroad underground slavery history slave slaves historical fiction freedom actual white plantation black people south real characters escape good america
Historical fiction	history good historical onion child bride girl lord slave god bird browns man slavery boy war interesting civil funny fiction
Readers' feelings	great good loved written put enjoyed author american club ending thought excellent recommend real page felt love end interesting black
Latin American experience, language	spanish dominican life history love republic family language culture footnotes references dont slang words understand english lose voice narrator sister
Families and characters	characters life love feel end lives beautiful felt family loved enjoyed wonderful human real didnt home relationships disappointed good character

Table 3. Ten topic models with interpretations and the top 20 co-occurring words for each topic

Interpretation	Topic Model Words
Family and history, heritage	family characters history generations house african american spirits half chapter time homegoing generation trade written jazz families interesting native historical
Eastern Asian experience	japanese japan korean family history women koreans interesting korea written war historical characters pachinko generations brides people saga style culture
African immigrant perspectives	america race american love african immigrant nigeria written life black culture great experience interesting immigrants nigerian perspective people country loved
African American historical experience	slavery railroad history underground slave historical slaves fiction written people freedom white black american actual america interesting important plantation real
Vietnam War	war vietnam vietnamese written american country perspective sympathizer people doors refugees man narrator interesting west author south thought general spy
Readers' perceptions	time reader work hard world style pages worth understand found narrative difficult readers find point times character dont novels bit
Characterization and family	characters love life written beautiful mother young felt family lives people child beautifully american powerful things feel past marriage black
Family	family brothers india short brother indian lowland father collection written characters lives hurricane life parents house esch bones detroit children
Readers' perceptions	good great characters author loved enjoyed end time recommend interesting put written wonderful excellent years life didnt thought page made
Latinx experience	spanish dominican history love life language family republic fiction culture boy onion voice funny sing dont footnotes words lord references

Finally, Table 4 shows the topic models with the number of topics retrieved reduced to five.

As the number of topics is reduced, the models naturally become more general, perhaps pointing to universal appeal factors, or perhaps to topics that are viewed as safe or appealing by publishers. As noted by Dawson and Van Fleet (2001), initial ventures into multicultural publishing tend to be stories about origins and cultural transitions. The award-winning books

selected for this analysis have had to pass through numerous gatekeepers within the publishing industry. Stories of people immigrating to the U.S. from their home countries might be viewed as stories that reaffirm a superior or desirable status for the U.S. in the minds of readers. Historical novels describing cultural differences might be viewed as past history, without an acknowledgement of the effects of the past on the present. The data of this project reveals the need for more insight into the mainstream publishing process to shed light on why and how books are chosen for publication and marketing, and whether these stories were chosen for their universal or unique characteristics. In other words, readers read what is available, accessible, and therefore, known to them. The data also suggests the question: in what ways do librarians contribute to the readership of multicultural titles and subsequent reader reviews?

Table 4. Five topic models with interpretations and the top 20 co-occurring words for each topic

Interpretation	Topic Model Words
Historical touchpoints, international	history war slavery railroad good written underground historical slave great characters japanese time fiction interesting slaves japan vietnam vietnamese korean
Family	characters family good life love mother great written time father lives child children parents character loved brothers author feel back
Readers' feelings	reader people work world time women style author short find pages written lives told place life beautiful interesting young narrative
Latin American characters	spanish dominican history life great love characters good time dont language character funny understand didnt family hard lot years republic
Characterization and culture	great characters love american good life loved written america enjoyed interesting race author black recommend end character african excellent culture

Publishers' Status

To complement the data on reviews, this project also sought to understand the extent to which *independent* publishers (i.e., publishers not acquired by a larger company and operating as an "imprint" of a larger unit), also known as *indie presses*, were responsible for the publication of the award books selected for study since they represented multicultural points of view and since indie presses were perceived as being more open to supporting diversity (Doherty, 2016). To do this, the publishers of the books were recorded. The publishing houses were then evaluated to assess their status.

As shown in Figure 2, nine of the 50 books were published by five indie presses (one indie press, Algonquin Books, was responsible for four of the titles; the other four indie presses were Akashic Books, Bloomsbury USA, Cinco Puntos Press, and Coffee House Press and each was responsible for the publication of one book). The remaining 41 books in the list were published either directly by the big five publishers (i.e., Hachette, HarperCollins, Macmillan, Penguin Random House (PRH), and Simon & Schuster) or through one of their imprints; or through a standard, traditional

press (e.g., Anchor). No books on the list of selected resources was self-published.

Discussion

Based on our analysis, we find several terms used by readers to describe books that may not be reflected in reviews of more mainstream award-winning books that feature the dominant white culture: culture-specific terms (e.g. “African,” “Nigerian,” “Dominican”) focusing on identity, historical topics that have shaped culture and identity (e.g., Vietnam War, the institution of slavery and the legacy it has left). Readers’ advisors might look for similar terms or ideas in books to indicate their appeal or lack of appeal to a multicultural audience. We conclude that while there are great differences in the ways readers write about books, even the same book, there are some consistent terms and ideas used to describe multicultural fiction in this data set.

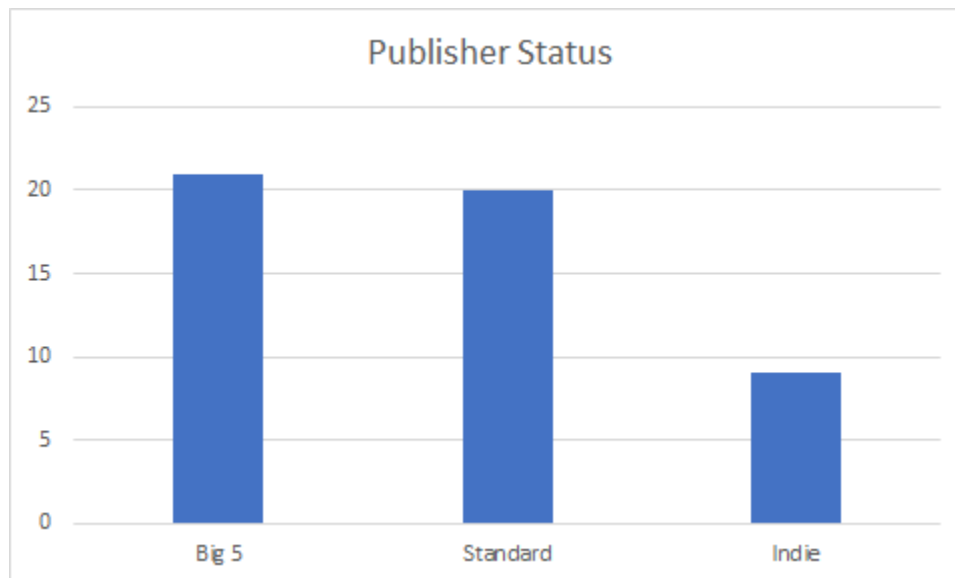


Figure 2. Publisher status for the 50 books selected for this project

All 50 of these books have garnered awards or recognition—as such, they have passed through the gatekeepers of publishing and awards committees. Some of the language used to describe these books was consistent with characteristics of literary fiction as described by Saricks (2009): depth of characterization and focus on relationships between characters. In the stories used in this analysis, family relations were very strongly represented. Many of the books focused on migration and immigration, which might be one explanation for the familial focus. To note, familial culture was both a source of pride and tension in the reader reviews. The portrayal of cultural beings experiencing cultural tensions may reinforce some people’s experience of culture and may be a new experience for other people. Readers’ advisors might focus on depth of cultural portrayal as being another measure by which to evaluate multicultural fiction for reader appeal.

Limitations and Future Research

This research acknowledges a number of limitations. First, the review of the literature, although sufficient to support the present analysis, is decidedly Anglo-American in scope. Future research

should investigate not only multicultural works in other languages and cultures, but also the publishing and access to those materials; a revised approach to presenting the literature in the field will be necessary to support that work. Next, Amazon was used as a source for the reviews because the platform represents a democratic forum for the exchange of ideas regarding the books themselves and has the potential to include a variety of voices and experiences. Amazon, as a platform, was chosen for reasons of convenience in light of the purpose of the project and the potential to collect data for analysis in this exploratory work. Future projects that do not have the same exploratory nature or that investigate other questions will want to use other sources for review data. This was exploratory work that tested a topics model method as a means of seeing how people described their reading experiences. Future projects will also utilize reviews and literature discussions that speaks to other identities (e.g., specifically youth of color, LGBTQ experiences, religious identity, etc.).

Conclusion

Text mining could be a successful way to determine appeals of multicultural literature, or indeed it could be used to analyze the appeal of other types of literature. We have pointed out terms used by readers to describe these works, and these terms may help librarians in recommending and describing multicultural fiction. These ideas should be incorporated into RA considerations to provide an additional avenue to connect readers with books that may reinforce their experience or expand their experience.

Text mining also presents a rich ground for future studies. Future research should be done to clarify several things, including whether our results for award-winning literature are consistent across popular multicultural literature such as street lit/urban fiction, a genre that has only recently received critical attention. Another potential topic for review is whether the racial background of a customer reviewer affects the content of the review—for example, if reviewers from a Mexican American culture point out aspects of cultural authenticity in a work that reviewers from other cultures might miss. Other research might analyze whether customer reviewers' language changes after public movements - for instance, whether reviewers use different words in reviews prior to and after the development of the #WeNeedDiverseBooks movement or accusations of inappropriate behavior against authors. This data could perhaps be found by mining text from discussions in other languages or in discussion forums that have a concentration of people with a particular identity (for instance, Latinx users).

Finally, this research suggests that librarians might want to seek books that are outside of traditional publication streams or by looking at book reviews outside of their normal venues. If indeed most book reviewers are white, cisgender, heterosexual non-disabled women, they could go to the readers themselves for additional advice on online forums such as Amazon, Goodreads, and other places where readers discuss books, to learn about what they are missing and to develop both physical and e-book collections that better meet the pleasure-reading needs of their community. Perhaps such efforts could eventually help diversify the profession, as well.

References

Birdi, B. (2014). *'We are here because you were there': An investigation of the reading of, and engagement with, minority ethnic fiction in UK public libraries* (Unpublished doctoral

- dissertation) University of Sheffield, Sheffield, United Kingdom.
- Birdi, B., & Syed, M. (2011). Exploring reader response to minority ethnic fiction. *Library Review*, 60(9), 816-831.
- Brooks, W., & Savage, L. (2009). Critiques and controversies of street literature: A formidable literary genre. *The ALAN Review*, 36(2), 48-55.
- Chen, Y. (2017, March 20). Confessions of a paid Amazon reviewer [blog post]. Retrieved from <https://digiday.com/marketing/vendors-ask-go-around-policy-confessions-top-ranked-amazon-review-writer/>
- Dali, K. (2010). Readers' advisory interactions with immigrant readers. *New Library World*, 111(5-6), 213-222.
- Dali, K. (2013). Hearing stories, not keywords: Teaching contextual readers' advisory. *Reference Services Review*, 41(3), 474-502.
- Dali, K. (2014). From book appeal to reading appeal: Redefining the concept of appeal in readers' advisory. *Library Quarterly*, 84(1), 22-48.
- Dawson, A., & Van Fleet, C. (2001). The future of readers' advisory in a multicultural society. In K. D. Shearer & R. Burgin (Eds.), *Readers advisor's companion* (pp. 249-269). Englewood, CO: Libraries Unlimited.
- Dawson, A., & Van Fleet, C. J. (Eds.). (2004). *African American literature: A guide to reading interests*. Westport, CT: Libraries Unlimited.
- Dilevko, J., & Magowan, C. F. (2007). *Readers' advisory service in North American public libraries, 1870-2005: A history and critical analysis*. Jefferson, NC: McFarland.
- Doherty, K. (2016). *We need diverse books and independent publishers: A Portland, Oregon, perspective* (Graduate Research Paper). Portland, OR: Portland State University. Retrieved from https://pdxscholar.library.pdx.edu/eng_bookpubpaper/17/
- Gibson, S. (2010). Critical readings: African American girls and urban fiction. *Journal of Adolescent & Adult Literacy*, 53(7), 565-574.
- Herald, D. T. (2006). *Genreflecting: A guide to popular reading interests* (6th ed.). Westport, CT: Libraries Unlimited.
- Hughes-Hassell, S. (2013). Multicultural young adult literature as a form of counter-storytelling. *The Library Quarterly*, 83(3), 212-228.
- Hughes-Hassell, S., & Rodge, P. (2007). The leisure reading habits of urban adolescents. *Journal of Adolescent & Adult Literacy*, 51(1), 22-33.
- Irvin Morris, V. (2012). *The readers' advisory guide to street literature*. Chicago, IL: ALA Editions.
- Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics*,

- 41(6), 750-769.
- Larrick, N. (1965, September 11). The all-white world of children's books. *The Saturday Review*, pp. 63-65, 84-85.
- Lawrence, E. (2017). Is contemporary readers' advisory populist?: Taste elevation and ideological tension in the *Genreflecting* series. *Library Trends*, 65(4), 491-507.
- Lee & Low Books. (2016, January 26). Where is the diversity in publishing? The 2015 diversity baseline survey results [blog post]. Retrieved from <http://blog.leeandlow.com/2016/01/26/where-is-the-diversity-in-publishing-the-2015-diversity-baseline-survey-results/>
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In K. Bontcheva & Z. Jingbo (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55-60). Stroudsburg, PA: ACL.
- Martinez, S. E. (2009). *Latino literature: A guide to reading interests*. Santa Barbara, CA: Libraries Unlimited.
- Milliot, J. (2015, October 16). The PW publishing industry salary survey 2015: A younger workforce, still predominantly white. *Publishers Weekly*. Retrieved from <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/68405-publishing-industry-salary-survey-2015-a-younger-workforce-still-predominantly-white.html>
- Nelson, R. K. (n.d.). Introduction. *Mining the Dispatch*. Retrieved from <http://dsl.richmond.edu/dispatch/pages/intro>
- Perez, S. (2016). Amazon bans incentivized reviews tied to free or discounted products. Retrieved from <https://techcrunch.com/2016/10/03/amazon-bans-incentivized-reviews-tied-to-free-or-discounted-products/>
- Poddar, N. (2016). 'Whiny assholes' or creative hustlers? On brownness, diaspora fiction, and Western publication. *Transition: An International Review*, 119, 92-106.
- Radway, J. (1984). *Reading the romance: Women, patriarchy, and popular culture*. Chapel Hill, NC: University of North Carolina Press.
- Rosenberg, B. (1982). *Genreflecting: A guide to reading interests in genre fiction*. Littleton, CO: Libraries Unlimited.
- Rosenblatt, L. M. (1938). *Literature as exploration*. New York, NY: D Appleton-Century.
- Sanchez, A. (2014). *On the poetics and politics of so-called 'ethnic' detective fiction: A chronotopic line-up of Peggy Blair's and Leonardo Padura's Cuban Crimes and Culprits*. Brussels, BE: Vrije Universiteit Brussel.
- Saricks, J. G. (2005). *Readers' advisory service in the public library* (3rd ed.). Chicago, IL:

- American Library Association.
- Saricks, J. G. (2009). *The readers' advisory guide to genre fiction* (2nd ed.). Chicago, IL: American Library Association.
- Saxton, M. D. (2018). A gentle introduction to topic modeling using Python. *Theological Librarianship*, 11(1), 18-27.
- Smith, D. (2015). Readers' advisory: The who, the how, and the why. *Reference & User Services Quarterly*, 54(4), 11-16.
- Underwood, T. (2012, April 7). Topic modeling made just simple enough [blog post]. Retrieved from <https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/>
- Xiao, D., Ji, Y., Li, Y., Zhuang, F., & Shi, C. (2018). Coupled matrix factorization and topic modeling for aspect mining. *Information Processing & Management*, 54(6), 861-873.

Denice Adkins (adkinsde@missouri.edu) is an associate professor at the iSchool at the University of Missouri. She was a Fulbright Scholar to Honduras in 2008, President of REFORMA (The National Association to Promote Library & Information Services to Latinos and the Spanish-Speaking) in 2012-2013, and Secretary-Treasurer for the Association of Library and Information Science Education (ALISE), 2014-2017. Her research interests include information needs of Midwestern immigrants, library services to diverse audiences, and public libraries.

Jenny S. Bossaller (bossallerj@missouri.edu) is an associate professor at the iSchool at the University of Missouri. Her teaching and research broadly encompasses public libraries and space, information policy, history, and related social and technological phenomena. Bossaller is currently Chair of the Library and Information Science Program at the University of Missouri and Chair of the Library History Round Table of the American Library Association.

Heather Moulaison Sandy (moulaisonhe@missouri.edu) is an associate professor at the iSchool at the University of Missouri. Moulaison Sandy's primary research focus is the organization of information in the online environment.