

University of Kentucky

UKnowledge

Information Science Faculty Publications

Information Science

7-2020

Identifying Facets of Reader-Generated Online Reviews of Children's Books Based on a Textual Analysis Approach

Yunseon Choi

Valdosta State University

Soohyung Joo

University of Kentucky, soohyung.joo@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/slis_facpub



Part of the [Computational Linguistics Commons](#), and the [Library and Information Science Commons](#)

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Repository Citation

Choi, Yunseon and Joo, Soohyung, "Identifying Facets of Reader-Generated Online Reviews of Children's Books Based on a Textual Analysis Approach" (2020). *Information Science Faculty Publications*. 70.

https://uknowledge.uky.edu/slis_facpub/70

This Article is brought to you for free and open access by the Information Science at UKnowledge. It has been accepted for inclusion in Information Science Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Identifying Facets of Reader-Generated Online Reviews of Children's Books Based on a Textual Analysis Approach

Digital Object Identifier (DOI)

<https://doi.org/10.1086/708962>

Notes/Citation Information

Published in *The Library Quarterly*, v. 90, no. 3.

© 2020 by The University of Chicago. All rights reserved.

The copyright holder has granted the permission to post this article here.

Identifying Facets of Reader-Generated Online Reviews of Children's Books Based on a Textual Analysis Approach

Yunseon Choi and Soohyung Joo

ABSTRACT

With the increasing popularity of social media, online reviews have become one of the primary information sources for book selection. Prior studies have analyzed online reviews, mostly in the domain of business. However, little research has examined the content of online book reviews of children's books. Book reviews generated by book readers contain different aspects of information, such as opinions, feedback, or emotional responses, from the perspectives of readers. This study explores what aspects of the books are addressed in readers' reviews, and then it intends to identify categorical features or facets of online book reviews of children's books. We employed a textual analysis approach including the latent Dirichlet allocation topic modeling to analyze the content of book reviews. The results indicate that online book reviews exhibit different facets of the books, which can be used as access points by potential readers to help them select relevant books.

As online reviews are getting more popular on social media platforms, they have become one of the primary information sources that people can utilize for decision making before making the purchase (Liu 2006; Park, Lee, and Han 2007; Sparks, Perkins, and Buckley 2013; Chen, Kou, and Peng 2018). Prior studies have examined online reviews in different contexts; in particular, business and computer science researchers explored the use of online reviews focusing on marketing and consumer behavior, such as product and service reviews or movie reviews (Hu and Liu 2004; Ganu, Kakodkar, and Marian 2013; Zha et al. 2014; Mate 2015). As social networking sites or other commercial sites enabled book readers to share their opinions and reviews about books and communicate with other readers, a few studies paid attention to online book reviews to examine the helpfulness of online reviews on sales of books (Chevalier and Mayzlin 2006; Mudambi and Schuff 2010).

Despite the popularity of research on online reviews in business and marketing areas, relatively little research has been done on online reviews in library and information science. Book reviews have been long used as a selection tool for collection building and book purchase

(Jenkins 1996). Moreover, the importance of online book reviews has been recognized as a new type of book culture that enables easy, massive communications among readers (Driscoll and Rehberg Sedo 2019).

The purpose of this study is to develop categories of online book reviews. We used latent Dirichlet allocation (LDA) topic modeling and facet analysis to develop categories of online book reviews. Our conclusion of this analysis presents the successful mapping of the identified categories with the traditional schemes (Dublin Core [DC] and MACHine-Readable Cataloging [MARC]). This study is part of a larger research project that aims to investigate whether online reviews of children's books would represent significant factors that are useful for better categorization of children's books, leading to easier reader selection.

This article intends to explore what aspects of books are addressed in reader-generated book reviews at a social networking site. Based on the understanding of book review content, we aim to identify a range of categories, focusing on which facets readers are interested in. In particular, we are interested in the unique context of children's books. Children's books stimulate children's emotional experience and provide intellectual challenges for children (Kiefer 2004). Reading books allows children to acquire various literacy devices including story structures, schemes, and literary conventions (Bus, de Jong, and van Ijzendoorn 2007). Good books are valuable for children's language acquisition and literacy learning and contribute to their cognitive and emotional development (Saracho and Spodek 2010). Given the vast body of children's literature, it has become imperative to select materials for children effectively. Thus, we strive to examine whether online book reviews would help characterize the aspects of books by identifying different facets from the perspectives of book reviews. This article extends our previous work (Choi and Joo 2019) to provide a comprehensive analysis of online book reviews by identifying the facets of reader-generated online book reviews. Identifying underlying aspects within online book reviews is an important matter that provides clues for understanding readers' behaviors in selecting and evaluating books.

Helpfulness of Online Reviews

Prior studies have confirmed the helpfulness of online reviews for business decisions (Liu 2006; Mudambi and Schuff 2010; Zhu and Zhang 2010; Kim and Gupta 2012; Yu et al. 2012; Huang et al. 2018). Online reviews help consumers in their decision-making process to choose a product or service (Mudambi and Schuff 2010), and consumers' social interactions via social media provide different values for both business and consumers (Hajli 2014). Consumers are more interested in other consumers' opinions rather than the product information provided by the vendor (Ridings and Gefen 2004). Therefore, analyzing and understanding these consumers' online reviews has increasingly become an essential part of the business decision.

The content of movie reviews has been an interest to researchers. For example, researchers investigated movie reviews by analyzing review text and discussed its implications for online

consumers (Liu 2006; Yu et al. 2012; Gaikar, Marakarkandy, and Dasgupta 2015). Dipak Damodar Gaikar, Bijith Marakarkandy, and Chandan Dasgupta (2015) analyzed movie reviews posted on Twitter, focusing on text sentiment, as a way to predict box office movie performance. Similarly, Xiaohui Yu and colleagues (2012) also analyzed sentiments in movie reviews to examine the relationships between review content and the future sales of movies. On the other hand, Yong Liu (2006) found that both negative and positive messages significantly influence box office revenue.

A few studies have examined online book reviews (Chevalier and Mayzlin 2006; Dimitrov et al. 2015; Driscoll and Rehberg Sedo, 2019). Judith A. Chevalier and Dina Mayzlin (2006) analyzed online book reviews on commercial sites, including Amazon.com and BarnesandNoble.com, to investigate the relationships between the content of online reviews and sales of books. They found that the impact of a negative review on the decrease in sales of books was much higher than the impact of a positive review on the increase in sales of books. Stefan Dimitrov and colleagues (2015) compared the Amazon.com review and Goodreads.com with regard to reviewers' behaviors and engagement. They found that Goodreads reviewers are more engaged toward a discussion of book characteristics whereas Amazon users tend to author more purchase-oriented reviews. They also investigated sentiment of review text on one genre, biography, and no significant difference was observed in sentiment scores on Goodreads.com and Amazon.com. Their study also revealed that Amazon ratings tend to be extreme and more focused on decision making. Beth Driscoll and DeNel Rehberg Sedo (2019) analyzed online book reviews of seven best-selling fiction and nonfiction books in Goodreads. They investigated the language and sentiments expressed in the reviews and discovered how online book reviews represent intimate reading experiences, emotional reactions to the book, and even social connections by addressing related book elements such as authors and other readers.

Methodology

Data Collection

We chose sample books for the analysis from the list of Newbery Medal Winners books. Using the random number generation tool, we randomly selected 20 books from the list (see table A1). Then, we manually scraped online reviews posted by users from Goodreads.com. A total of 3,062 reviews, consisting of 55,856 tokens, were collected. A token indicates a meaningful unit of text, most often a certain word or term, that is used for further analysis in text mining (Silge and Robinson 2017). The data was collected during the period from November 2017 to June 2018. All non-English reviews were excluded from the data set. The collected reviews cover the period from 2006 to 2018. This study did not identify users' profiles, and the collected online book reviews did not include any identification information. However, during the content analysis of the reviews, we found that reviews were written not only by adults (such as parents or teachers) but also by

children readers themselves. We believe this is because children sometimes lead to the purchase or borrowing of books, although they are not the target market for book purchasing.

Data Analysis

The collected text underwent preprocessing, including tokenization, stopword removal, and stemming. We employed LDA topic modeling to analyze the content of reviews. The LDA topic model is an unsupervised machine learning technique that can be used to discover hidden themes or topics based on the analysis of the semantic structure of the text (Blei, Ng, and Jordan 2003; Blei 2012). LDA topic modeling has been adopted to analyze the content of online reviews in various contexts, such as travel, products, or medical services (Ma et al. 2013; McAuley and Leskovec 2013; Xianghua et al. 2013; Guo, Barnes, and Jia 2017). LDA topic modeling can be a compelling method to identify different facets or aspects from the collection of unstructured text. As to the number of topics, we decided to set it as 15 topics for each model after examining multiple sets of topics. We initially examined 10, 12, 15, and 20 topics for each book respectively, and it turned out 15 topics would exhibit most coherent topics by covering different aspects of the books. Then, open coding was employed to interpret the topic model results qualitatively. We analyzed the top 10 high-probability terms for each topic qualitatively and then determined to which facet a topic would be relevant to describe the chosen book. In this way, we were able to define different categories of review content, such as audience, genre, subject, evaluation, and others. More importantly, we further conducted a mapping of the facets with the DC and MARC to identify which facets of the books are covered by the review content. We compared the facets identified from the topic modeling analysis with knowledge organization systems (KOS) such as DC metadata standards (National Information Standards Organization 2007) and MARC record standards (Library of Congress 2020).

Results

Analysis of Latent Topic Terms

For all 20 books, we ran LDA topic modeling. For each book, we qualitatively analyzed the 10 highest-probability terms for each topic. Due to the limited space, we cannot report all 20 tables of topic modeling results herein. We selectively report three representative examples of the LDA topic modeling results: *Criss Cross*, *When You Reach Me*, and *The Giver*.

Figure 1 presents 15 topics extracted from the reviews of *Criss Cross*. The results reveal that book reviews cover different aspects of the book. For instance, topic 1 represents the audience of the book by showing high probabilities of terms like young, teenager, adult, age, and group. Topic 3 contains the terms that represent the format or physical description of the book. Topics 6 and 11 include emotional terms, such as “beautiful,” “favorite,” and “sweet.” Topics 7, 8, 10, and 14 represent mostly the subject of the book. Topic 15 indicates award notes, including “newbery,” “winner,” “win,” “award,” and “medal.”

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
young	noth	plot	charact	book
teenag	interest	love	feel	read
perkin	tri	write	just	year
live	live	real	book	love
adult	kid	teen	connect	written
small	bore	differ	moment	finish
enjoy	find	illustr	miss	ill
lynn	listen	experi	adolesc	pretti
group	world	find	grow	recommend
age	charact	town	happen	view
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
life	time	stori	novel	criss
peopl	that	confus	charact	cross
time	know	throughout	person	thought
beauti	part	write	chapter	girl
appreci	talk	differ	just	learn
favorit	stori	school	relat	play
wonder	work	interest	mani	reader
let	realli	author	love	read
author	thought	high	summer	boy
need	self	better	everi	word
Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
seem	book	realli	debbi	newberi
read	page	just	happen	read
kid	time	know	hector	winner
old	view	there	wish	feel
isnt	give	mayb	friend	newberri
amaz	read	care	interest	won
clear	understand	start	necklac	award
sweet	cant	mani	lenni	understand
without	funni	around	actual	almost
moment	subtl	quiet	line	medal

Figure 1. Latent Dirichlet allocation topic model results for *Criss Cross*

Figure 2 presents the topics extracted from the reviews of *When You Reach Me*. Topics 2, 12, and 15 involve terms related to the genre or format of the book, such as “mystery,” “fiction,” “time,” and “travel.” Topics 3, 4, and 6 are mostly about the subject of the book. Topics 5 and 8 indicate award notes, including “newbery,” “winner,” and “award.” Topic 7 represents the audience, such as “kid,” “school,” “young,” “adult,” and “children.”

Figure 3 shows the topics uncovered from the reviews of *The Giver*. Topics 1 and 3 include the author of the book, Lois Lowry. The term, “seri,” which is the stemmed term of “series,” in topic 2 provides a clue that this book is a series book. Topic 4 mostly represents “the audience” of the book. Topics 6, 7, 8, 10, 11, 12, 13, and 15 correspond to the subject of the book. Topic 9 represents the mixture of emotional terms (like “love”) and evaluative comments such as “great” and “amazing.”

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
show	reader	miranda	life	mom
year	mysteri	note	old	newberi
plot	fiction	friend	happen	part
back	understand	sal	grade	stori
game	lengl	mysteri	letter	grow
pyramid	puzzl	best	start	favorit
without	interest	know	york	detail
narrat	begin	talk	late	expect
fun	there	man	need	award
twist	believ	find	probabl	give
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
love	kid	read	stead	book
friendship	school	time	reach	read
thought	young	wrinkl	novel	know
great	adult	enjoy	rebecca	write
felt	children	madelein	fact	charact
still	rememb	hard	sort	love
recommend	won	pick	anyth	person
bit	pretti	winner	job	mani
time	absolut	award	question	differ
author	year	quick	today	finish
Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
just	stori	book	realli	time
veil	charact	age	just	travel
world	wonder	girl	review	peopl
someon	page	reach	alway	work
mean	mysteri	love	whole	star
sometim	reader	middl	better	sens
beauti	middl	give	ive	least
tri	seem	wrinkl	mind	along
element	feel	seem	cant	explain
togeth	complex	everyth	sad	exact

Figure 2. Latent Dirichlet allocation topic model results for *When You Reach Me*

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
lowri	charact	stori	children	enjoy
book	just	thought	peopl	realli
present	short	lowri	societi	age
kid	interest	left	adult	best
doesnt	great	loi	rule	certain
special	seri	main	yet	color
creat	love	decid	young	everi
reader	real	clear	middl	past
author	felt	power	control	read
individu	develop	given	perfect	friend
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
giver	year	seem	book	world
star	giver	mani	review	live
give	old	just	next	life
need	question	dystopia	love	pain
messag	societi	write	great	without
communiti	experi	reader	long	perfect
rememb	almost	back	word	person
tale	start	find	amaz	choic
sad	although	time	wasnt	chang
quick	page	pretti	younger	given
Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
jona	read	realli	novel	feel
communiti	dystopian	book	societi	time
receiv	school	read	life	giver
memori	idea	know	peopl	emot
everyon	genr	made	believ	love
assign	year	movi	happi	read
giver	must	still	ive	young
everyth	alway	mayb	free	releas
job	classic	just	better	perhap
differ	find	whole	freedom	learn

Figure 3. Latent Dirichlet allocation topic model results for *The Giver*

Categories Discovered

As shown in the topic modeling results, different aspects of book content and readers' evaluation, responses, and feeling were observed in the Goodreads review posts. We reviewed topic terms qualitatively and reorganized those terms to identify common facets that can be used to describe the books. To be more specific, we used the facet analytical technique, which is used in faceted classification and categorized the topic terms into a certain facet. Facets are defined as "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or specific subject" (Taylor 1992, 394). Faceted classification is an "analytico-synthetic" scheme that derives from two processes: analysis and synthesis. These definitions of facets help us understand the definition of faceted classification: "the sorting of terms in a given field of knowledge into homogeneous, mutually exclusive facets, each derived from the parent universe by a single characteristic of division" (Vickery 1960, 12). Faceted classification uses a "bottom-up" (Batty 1989) design technique where each class is analyzed and broken down into its basic elements, grouped together by common attributes, called facets. Using the facet analysis, we examined high-probability topic terms and then determined an appropriate facet to express those topic terms. For example, in figure 2, topic 7 includes terms like "kid," "school," "young," "adult," and "children," and we chose a category name as "audience" to express those words. Once the facets (the names of categories) were defined, we then enumerated terms under the facet. For example, we viewed that the terms of "children," "newbery," "poetry," "recommend," and "enjoy" can form mutually exclusive facets respectively, such as "audience," "award notes," "genre/form," "evaluation," and "emotion." In this way, we came up with 11 facets as shown in the third column of table 1, including "author," "characters," "physical description," "subject," "title," "genre/form," and others.

In addition, because online book reviews provide descriptions about a book, that is, bibliographic data, we examined to what extent the facets derived from the book reviews could map the KOS including DC metadata elements and MARC fields. The DC metadata was designed to describe a broad range of resources. We started with Simple (unqualified) DC metadata elements, which consist of 15 core metadata elements. We then reviewed MARC standards (Library of Congress 2020), which are currently used in most libraries for the description of items. We used MARC fields by mapping DC to MARC (Library of Congress 2004).

We found that readers' reviews would cover common metadata elements describing books, such as contributor, description, format, and subject. Interestingly, certain review terms indicated target audience information, such as "child," "children," and "kid." In addition, the unique facets contributed by the review terms include the emotional and evaluation facets, which represent readers' feelings and opinions, respectively. We observed a certain amount of terms expressing the readers' emotions, such as "sad," "hate," "happy," and "cry." Also, some portions of the terms were used to judge the positiveness or negativeness about the books, such as "good," "bad," "interesting," "amazing," or "boring." These terms form the facets of "emotion" and "evaluation," which are not found in DC core elements and MARC fields.

Table 1. Categories of Book Review Terms

Dublin Core Element	Machine-Readable Cataloging Fields	Category of Topic Terms	Exemplary Terms
Contributor	100, 110, 111, 700, 710, 711	Author	(Vary) author
Description	500–599, except 506, 530, 540, 546	Characters	Boy, girl, character (names of characters)
Format	340; 856\$ <i>q</i> 300	Physical description	Illustration, figure, picture
Subject	050, 060, 080, 082, 600, 610, 611, 630, 650, 653	Subject	(Vary)
Title	245, 246	Title	(Vary)
Type	655 Genre/form	Genre/form	Fiction, novel, poetry, narrative
...	586	Award	Award, medal, Newbery, winner, win
...	521 Target audience	Audience	Child, adult, kid, middle, audience
...	490 Series	Series	Series
...	...	Emotion	Favorite, sad, enjoy, like, excite, hate, glad, happy, cry, surprise
...	...	Evaluation (opinion)	Good, bad, funny, humor, great, complex, interesting, amazing, sweet, unique, boring, reread, perfect, recommend

On the other hand, it is interesting to note that some topic terms are associated with coherence. For instance, regarding book titles such as *The Crossover*, *Moon over manifest*, and *When you reach me*, the comments on the “audience” were more likely to co-occur with comments related to the “genre/form,” especially when books use the unique styles of writing elements such as “time travel,” “different periods,” and “mystery” (see table 2).

Discussion and Conclusion

This study examined the online book reviews of children’s books in the context of a social bookmarking site. We employed LDA topic modeling to investigate which aspects or facets

Table 2. Book Titles That Use the Unique Styles of Writing Elements

Title	Genre/Format	Topic Words	Notes
<i>The Crossover</i>	Poetry format	Poetry, verse, style, form, rhythm, novel	The book is written in verse.
<i>Moon Over Manifest</i>	Time period	Time, period, narrative, mystery	Stories in this book are set in two historical time periods between 1917 and 1936.
<i>When You Reach Me</i>	Time travel	Time, travel, mystery, fiction, narrative, novel	The story has time travel elements and deals with complex mystery characters.

were mentioned in online reviews of children's books. From the qualitative analysis of topic terms generated from the LDA topic model, we identified 11 facets of the review content.

The topic modeling results revealed that book reviews cover different aspects of the books, ranging from audience, subject, format, genre, evaluation, emotion, and others. This suggests that book reviews can serve as a rich source of vocabulary to be used for various information access points that might be useful for finding and selecting books for children. Then, we used the facet analytical technique to categorize the topic terms into a certain facet. We conducted a mapping of the facets with the DC and MARC and discovered 11 categories. Utilizing KOS such as DC and MARC in the process of developing facet categories facilitates creating descriptive metadata for discovering or selecting books for children.

The facets identified in this study represent various features of book reviews. The categories mostly belong to descriptive metadata. The descriptive metadata describe a resource for the purpose of its finding and understanding, for example, title, author, subject, genre, and publication date (Riley 2017). The terms categorized into the Author category include book authors' names, or reviewers just used a general term "author" instead. The Character category was defined to indicate the names of characters in a book. The terms assigned to the Physical Description category include "illustration," "figure," "picture," and others. The Subject category and the Title category terms vary by book. The Genre/Form category represents the genre of a book, such as "fiction," "novel," "poetry," and "narrative." The remaining three categories, namely the Award, Audience, and Series categories, do not have equivalent elements to be mapped with Simple DC. These three categories are uniquely identified from the analysis of the review terms. Because the selected books analyzed in this study are all Newbery Award-winning books, terms such as "award," "medal," "Newbery," "winner," and "win" appeared in the reviews. Moreover, in some reviews, those "award notes" were often observed. The Audience category consists of several terms indicating the levels of audience, such as "child," "adult," "kid," and "middle." Some reviewers mentioned the general term "audience" while discussing the reading levels of the book. Some book reviews include the term "series," which indicates the book is one of a book series.

The findings of this study suggest that emotion and evaluation can be considered important dimensions of book categorization from the perspective of readers. We observed that the book reviews contain a substantial number of emotional terms. These terms are not mapped with DC elements and MARC fields. This observation implies that emotional terms could be a candidate for a facet in the organization of children's books from the perspective of readers. In addition, many terms we observed reflected reviewers' evaluation or nuanced judgment about the book, for example, "good," "bad," "great," and "recommend."

On the other hand, some review posts contain terms related to the genre or format of books. The most common term in this category was "novel," as most of the Newbery selections are fictional novels. In table 2, for the book *The Crossover*, that takes the form of a verse

novel. The verse novel is defined as “a novel in poetry. A hybrid form, the verse novel filters the devices of fiction through the medium of poetry” (Academy of American Poets 1996). The book is a story about basketball, family, and the rules of life, which is written in verse. We observed a distinctively higher number of the terms “poem” and “poetri” in the reviews of this book. Some specific genre terms occur for those books that belong to that genre. For example, the term “mysteri” was observed much more in the book *When You Reach Me*. This book is a mystery novel about a girl who lives with a single mother. The story has time travel elements and deals with complex mystery characters. In addition, the book *Moon Over Manifest* is about a story that is set in two historical periods, 1917 and 1936. These findings may imply that book reviewers tend to recommend the audience for the books regarding the unique style of writing. Examples of the reviews include:

The Crossover

The Crossover, Newbery Medalist 2015, confirms **the growing trend of novels in verse form, telling basic coming-of-age stories in short, free verse**. This one follows exactly the same pattern as for example “The Weight of Water” or “Cloud Busting,” slightly different from “Love That Dog” and “Hate That Cat,” or “To Be Perfectly Honest,” yet another one of those seemingly endless novels which claim to be deep because the language is broken into stanzas, sometimes rhyming, but most of the time just stretched. One sentence per page doesn’t make poetry, though. A sentence that drips and falls to look like a ball on a basketball field is not that special. **It is an exercise for higher Primary School students**. Rubric: Make the words act out the content!

When You Reach Me

Another award winner for all ages. Time travel is a difficult subject and the author describes it well. She builds a wonderful mystery. I enjoyed that *A Wrinkle in Time* was an inspiration for this book.

We also found that the reviewers were aware of the age range of the book. Reviewers were likely to discuss for which age group a book would be appropriate to read. The examples of reviews include:

Moon Over Manifest

When first assigned to read this book, I was slightly skeptical and hesitant to read it; mainly because it didnt seem like something i would enjoy. However, after reading *Moon Over Manifest*, I came to the conclusion that I really enjoyed it. I really loved how there was two different eras in the book, even though it took me a little bit to get used to the layout. . . . **Overall though, I loved this book and would [recommend] it to any child around age 10+.**

The Bronze Bow

This book is not one I would have picked up off the shelves. I read it only because it was assigned to my daughter for her history class, and I wanted to be able to discuss it with her, as reviews of the book stated that it touched on sensitive themes. **I'd recommend this book for 6th graders through adult, though sensitive individuals should wait a bit longer than 6th grade to read it.** The author writes with nuance and inferences, so strong reading skills are required to truly understand the book's many relationships and messages. *The Bronze Bow* is intended as historical fiction, and is published by Houghton Mifflin, a secular publisher.

This finding highlights the uniqueness of online reviews of children's books, particularly represented in the "emotion" and "evaluation" facets. Online reviewers tend to use "natural" language, which is close to our daily spoken language, rather than well-written, structured writings. Therefore, it is likely to involve many emotional terms that we use in our daily life when evaluating something, such as "love," "enjoy," "happy," and "excite." Also, the selected Newbery books are novels that have stories entailing the changes of the characters' feelings across the storyline or emotional touching scenes or moments, so the reactions of reviewers consequently would become emotional as well.

Although our study was limited to online reviews of only 20 books, the results of this study showed that reader-generated book reviews involve rich information about the book contributed by a reviewer community. Reader communities could benefit from different sources of information about books, and online reviews can be a compelling source, which can be easily accessible via social media such as Goodreads. The contribution of this study lies in that we tried to understand the nature of readers' language used in book reviews and attempted to identify the facets of online book reviews that can be compared with traditional KOS. We also identified additional facets such as "emotion" and "evaluation (opinion)" that are uniquely applied to online reviews. This implies that user-generated book reviews might be influential on potential users' behavior or decision making on book selection by complementing formal bibliographic data with additional access points. A further study might be meaningful if it investigates the emotional and evaluative aspects of reader-generated reviews. Furthermore, in the future study, facets will be extended by using Qualified DC metadata (Dublin Core Metadata Initiative 2005). For example, in the Qualified DC, the Audience element has refinements. These refinements are the extensions to the "Simple DC" 15 elements and have additional elements including "Mediator" and "Education Level." The examples of the "Mediator" element are reading specialist and teacher. The "Education Level" element can be used for a grade level such as elementary school students or fourth to fifth grades. In the future study, the facets will be further extended by mapping with the DC element refinements.

Appendix

Table A1. A List of Selected Books

No.	Title	Author(s)	Year	Award Year
1	<i>Caddie Woodlawn</i>	Carol Ryrie Brink	1935	1936
2	<i>Criss Cross</i>	Lynne Rae Perkins	2005	2006
3	<i>Dead End in Norvelt</i>	Jack Gantos	2011	2012
4	<i>From the Mixed-Up Files of Mrs. Basil E. Frankweiler</i>	E. L. Konigsburg	1967	1968
5	<i>Julie of the Wolves</i>	Jean Craighead George	1972	1973
6	<i>Moon Over Manifest</i>	Clare Vanderpool	2010	2011
7	<i>Out of the Dust</i>	Karen Hesse	1997	1998
8	<i>Roll of Thunder, Hear My Cry</i>	Mildred D. Taylor	1976	1976
9	<i>Sarah, Plain and Tall</i>	Patricia MacLachlan	1985	1986
10	<i>The Bronze Bow</i>	Elizabeth George Speare	1961	1962
11	<i>The Cat Who Went to Heaven</i>	Elizabeth Coatsworth	1930	1931
12	<i>The Crossover</i>	Kwame Alexander	2014	2015
13	<i>The Girl Who Drank the Moon</i>	Kelly Barnhill	2016	2016
14	<i>The Giver</i>	Lois Lowry	1993	1994
15	<i>The One and Only Ivan</i>	K. A. Applegate	2012	2013
16	<i>The Tale of Despereaux</i>	Kate DiCamillo	2003	2004
17	<i>The View from Saturday</i>	E. L. Konigsburg	1996	1997
18	<i>Walk Two Moons</i>	Sharon Creech	1994	1995
19	<i>When You Reach Me</i>	Rebecca Stead	2009	2010
20	<i>Wonder</i>	R. J. Palacio	2012	2013

References

- Academy of American Poets. 1996. *Verse Novel: From A Poet's Glossary*. <https://www.poets.org/poetsorg/text/verse-novel-poets-glossary>.
- Batty, David. 1989. "Thesaurus Construction and Maintenance: A Survival Kit." *Database* 12 (1): 13–20.
- Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55 (4): 77–84.
- Blei, David M., Andrew Y. Ng, and Michale I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (January): 993–1022.
- Bus, Adriana G., Maria T. de Jong, and Marinus H. van Ijzendoorn. 2007. "Social Aspects in Language and Literacy Learning: Progress, Problems, and Interventions." In *Contemporary Perspectives on Social Learning in Early Childhood Education*, edited by Olivia N. Saracho and Bernard Spodek. Charlotte, NC: Information Age.
- Chen, Jia, Gang Kou, and Yi Peng. 2018. "The Dynamic Effects of Online Product Reviews on Purchase Decisions." *Technological and Economic Development of Economy* 24 (5): 2045–64.
- Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43 (3): 345–54.
- Choi, Yunseon, and Soohyung Joo. 2019. "Topic Detection of Online Book Reviews: Preliminary Results." *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. <https://doi.org/10.1109/JCDL.2019.00098>.

- Dimitrov, Stefan, Faiyaz Zamal, Andrew Piper, and Derek Ruths. 2015. "Goodreads versus Amazon: The Effect of Decoupling Book Reviewing and Book Selling." *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/viewFile/10557/10452>.
- Driscoll, Beth, and DeNel Rehberg Sedo. 2019. "Faraway, So Close: Seeing the Intimacy in Goodreads Reviews." *Qualitative Inquiry* 25 (3): 248–59.
- Dublin Core Metadata Initiative. 2005. *Using Dublin Core—Dublin Core Qualifiers*. <http://www.dublincore.org/specifications/dublin-core/usaguide/qualifiers/>.
- Gaikar, Dipak Damodar, Bijith Marakarkandy, and Chandan Dasgupta. 2015. "Using Twitter Data to Predict the Performance of Bollywood Movies." *Industrial Management and Data Systems* 115 (9): 1604–21.
- Ganu, Gayatree, Yogesh Kakodkar, and Amélie Marian. 2013. "Improving the Quality of Predictions Using Textual Information in Online User Reviews." *Information Systems* 38 (1): 1–15.
- Guo, Yue, Sturt J. Barnes, and Quong Jia. 2017. "Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation." *Tourism Management* 59 (April): 467–83.
- Hajli, M. Nick. 2014. "A Study of the Impact of Social Media on Consumers." *International Journal of Market Research* 56 (3): 387–404.
- Hu, Mingqing, and Bing Liu. 2004. "Mining and Summarizing Customer Reviews." In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. New York: ACM.
- Huang, Liqiang, Chun-Hoo Tan, Weiling Ke, and Kwok Kee Wei. 2018. "Helpfulness of Online Review Content: The Moderating Effects of Temporal and Social Cues." *Journal of the Association for Information Systems* 19 (6): Article 3.
- Jenkins, Paul O. 1996. "Faculty Priorities: Where Does Material Selection Stand?" *Collection Building* 15 (1): 19–20.
- Kiefer, Barbara. 2004. "Children's Literature and Children's Literacy: Preparing Early Literacy Teachers to Understand the Aesthetic Values of Children's Literature." In *Contemporary Perspectives on Language Policy and Literacy Instruction in Early Childhood Education*. Vol. V. Edited by Olivia N. Saracho and Bernard Spodek. Greenwich, CT: Information Age.
- Kim, Junyong, and Pranjal Gupta. 2012. "Emotional Expressions in Online User Reviews: How They Influence Consumers' Product Evaluations." *Journal of Business Research* 65 (7): 985–92.
- Library of Congress. 2004. *MARC to Dublin Core Crosswalk*. <https://www.loc.gov/marc/marc2dc.html>.
- Library of Congress. 2020. *MARC Standards*. <http://www.loc.gov/marc/>.
- Liu, Yong. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue." *Journal of Marketing* 70 (7): 74–89.
- Ma, Baizhang, Dongsong Zhang, Zhijun Yan, and Taeha Kim. 2013. "An LDA and Synonym Lexicon Based Approach to Product Feature Extraction From Online Consumer Product Reviews." *Journal of Electronic Commerce Research* 14 (4): 304–14.
- Mate, Chetan. 2015. "Product Aspect Ranking Using Sentiment Analysis: A Survey." *International Research Journal of Engineering and Technology* 3 (1): 126–27.
- McAuley, Julian, and Jure Leskovec. 2013. "Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text." *Proceedings of the 7th ACM Conference on Recommender Systems*: 165–72.
- Mudambi, Susan M., and David Schuff. 2010. "What Makes a Helpful Review? A Study of Customer Reviews on amazon.com." *MIS Quarterly* 34 (1): 185–200.
- National Information Standards Organization. 2007. *ANSI/NISO Z39.85 - 2001 Dublin Core Metadata Element Set*. https://groups.niso.org/apps/group_public/download.php/6572/z39-85-2007_dublin_core.pdf.

- Park, Do-Hyung, Jumin Lee, and Ingoo Han. 2007. "The Effect of On-line Consumer Reviews on Consumer Purchasing Intentions. The Moderating Role of Involvement." *International Journal of Electronic Commerce* 11 (4): 57–80.
- Ridings, Catherine M., and David Gefen. 2004. "Virtual Community Attraction: Why People Hang Out Online." *Journal of Computer-Mediated Communication* 10 (1): 1–10.
- Riley, Jenn. 2017. "Understanding Metadata. What Is Metadata, and What Is It For?" *National Information Standards Organization*. http://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Met%E2%80%A6.
- Saracho, Olivia N., and Bernard Spodek. 2010. "Families' Selection of Children's Literature Books." *Early Childhood Education Journal* 37 (5): 401–9.
- Silge, Julia, and David Robinson. 2017. *Text Mining with R*. Sebastopol, CA: O'Reilly.
- Sparks, Beverley A., Helen E. Perkins, and Ralf Buckley. 2013. "Online Travel Reviews as Persuasive Communication: The Effects of Content Type, Source, and Certification Logos on Consumer Behavior." *Tourism Management* 39 (December): 1–9.
- Taylor, Arlene G. 1992. *Introduction to Cataloging and Classification*. 8th ed. Englewood, CO: Libraries Unlimited.
- Vickery, Brian Campbell. 1960. *Faceted Classification: A Guide to Construction and Use of Special Schemes*. London: Aslib.
- Xianghua, Fu, Liu Guo, Guo Yanyan, and Wang Zhiqiang. 2013. "Multi-Aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and HowNet Lexicon." *Knowledge-Based Systems* 37 (2013): 186–95.
- Yu, Xiaohui, Yang Liu, Xiangi Huang, and Aijun An. 2012. "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain." *IEEE Transactions on Knowledge and Data Engineering* 24 (4): 720–34.
- Zha, Zheng-Jun, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua. 2014. "Product Aspect Ranking and Its Applications." *IEEE Transactions on Knowledge and Data Engineering* 26 (5): 1211–24.
- Zhu, Feng, and Xiaoquan Michael Zhang. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics." *Journal of Marketing* 74 (2): 133–48.

Yunseon Choi: assistant professor in the Department of Library and Information, Valdosta State University. Choi earned her PhD in library and information science from the University of Illinois at Urbana–Champaign. Her research interests encompass diverse aspects of information organization in the context of information technologies, including extracting values from data in social media, linked data, metadata management, and curation, and health informatics. Email: yunchoi@valdosta.edu.

Soohyung Joo: assistant professor in the School of Information Science at the University of Kentucky. Joo received his PhD in information studies from the University of Wisconsin–Milwaukee. His areas of research interest have centered on user interaction with digital resources and information systems, applied data science, textual analysis, and social media. He has recently participated in an Institute of Museum and Library Services funded project that investigates language use and community needs in storytimes. Email: soohyung.joo@uky.edu.