

The Sellout: Readers Sentiment Analysis of 2016 Man Booker Prize Winner

N Kadiresan

Tripti Gogoi

Mangkhollen Singson

Abstract

Sentiment analysis also sometimes term as opinion mining refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to the voice of the customer materials such as reviews and survey responses, online and social media, and materials for healthcare applications that range from marketing to customer service to clinical medicine. This paper mainly focuses on the sentiment analysis of the 2016 Man Booker prize winner book “The Sellout” in Goodreads website. An opinion is classified as a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder. This is a relevant problem in today’s world as the amount of user generated text on the web is increasing and sentiment analysis can be used to detect the mood of users on a forum. The paper examines the positive and negative sentiment polarity of the readers before and after winning the prize, the most occurred word in the reviews, word frequency and workload.

Keywords: Analysis, Goodreads, Man Booker, R Language, Reviews, Sellout, Sentiment

1. Introduction

Goodreads was launched in January 2007 by Otis Chandler, II, a software engineer and entrepreneur, and Elizabeth Chandler. It is an Amazon company and “social cataloguing” website founded in December 2006. The website allows individuals to freely search Goodreads’ database of books, annotations, and reviews. Users can sign up and register books to generate library catalogs and reading lists. They can also create their own groups of book suggestions, surveys/polls, blogs, and discussions.

Goodreads is known to be one of “the world’s largest sites for readers and book recommendations”, and

in December 2007, the site had over 6, 50,000 members and over 10,000,000 books had been added. By July 2012, the site reported 10 million members, 20 million monthly visits, and 30 employees. On July 23, 2013, it was announced on their website that the user base had grown to 20 million members, having doubled in close to 11 months. The website’s offices are in San Francisco.

1.1 Background

The Chandlers created Goodreads in 2006. Goodreads’ mission is “to help people find and share books they love... [and] to improve the process of reading and learning throughout the world.” Goodreads also addressed “what publishers call the ‘discoverability’ problem” by guiding consumers in the digital age to find books they might want to read. During its first year of business, the



company was run without any formal funding. In December 2007, the site received funding estimated at \$750,000 from angel investors. In 2011, Goodreads acquired Discovereads, a book recommendation engine that employs “machine learning algorithms to analyse which books people might like, based on books they’ve liked in the past and books that people with similar tastes have liked.” After a user has rated 20 books on its five-star scale, the site will begin making recommendations. Otis Chandler believed this rating system would be superior to Amazon’s, as Amazon’s includes books a user has browsed or purchased as gifts when determining its recommendations. Later that year, Goodreads introduced an algorithm to suggest books to registered users and had over five million members. In January 2016, Amazon announced on Shelfari.com that it would be merging Shelfari with Goodreads and closing down Shelfari. To prepare Shelfari members for the move, Amazon posted on Shelfari.com a prominent announcement.

1.2 Features of Goodreads

Some of the features of Goodreads are as follows:-

1.2.1 Book Discovery: While going through the Goodreads website, users can add books to their personal bookshelves, rate and review books, see what their friends and authors are reading, participate in discussion boards and groups on a variety of topics, and get suggestions for future reading choices based on their reviews of previously read books. Once users have added friends to their

profile, they will see their friends’ shelves and reviews and can comment on friends’ pages. Goodreads features a rating system of one to five stars, with the option of accompanying the rating with a written review. The site provides default bookshelves - read, currently-reading, to-read and the opportunity to create customized shelves to categorize a user’s book.

1.2.2 Content Access: Goodreads have the content access option in it, as the users can read or listen to a preview of a book on the website using Kindle Cloud Reader and Audible. It also offers quizzes and trivia, quotations, book lists, and free giveaways. Members can receive the regular newsletter featuring new books, suggestions, author interviews, and poetry. If a user has written a work, the work can be linked to the author’s profile page, which also includes an author’s blog. Goodreads organizes offline opportunities as well, such as IRL book exchanges and “literary pub crawls”.

1.2.3 User Interaction: The website facilitates reader interactions with authors through the interviews, giveaways, authors’ blogs, and profile information. There is also a special section for authors with suggestions for promoting their works on Goodreads.com, aimed at helping them reach their target audience.

Additionally, Goodreads has a presence on Facebook, Pinterest, Twitter, and other social networking sites. Linking a Goodreads account with a social networking account like Facebook enables the ability to import contacts from the social networking account to Goodreads, expanding one’s Goodreads “Friends” list. There are settings available, as well, to allow Goodreads to post straight to a social networking account, which informs, e.g.,

Facebook friends, what one is reading or how one rated a book. This constant linkage from Goodreads to other social networking sites keeps information flowing and connectivity continuous.

1.3 Goodreads Choice Awards: Goodreads had introduced Goodreads Choice Awards. It is a yearly award program, first launched in 2009. Through it users are able to vote for the books that Goodreads has nominated and are also able to nominate books of their choosing, released in the given year. The majority of books that Goodreads itself nominates are from Goodreads authors. The final voting round collects the top 10 books from 20 different categories.

1.4 Booker Prize: The Sellout

The Man Booker Prize for Fiction (formerly known as the Booker-McConnell Prize and commonly known simply as the Booker Prize) is a literary prize awarded each year for the best original novel, written in English language and published in the UK. The winner of the Man Booker Prize is generally assured international renown and success; therefore, the prize is of great significance in the book trade. From its inception, only Commonwealth, Irish, and South African (later Zimbabwean) citizens were eligible to receive the prize; in 2014, however, this eligibility was widened to any English-language novel.

The Sellout is a 2015 novel by Paul Beatty published by Farrar, Straus and Giroux, and in the UK by Oneworld Publications in 2016 won the Man Booker prize, becoming the first US writer to win that award. The novel takes place in and around Los Angeles, California, and concerns a protagonist who grows artisanal marijuana and watermelons. Beatty

has stated his motivation for writing the novel was that “[he] was broke”.

1.5 The Sellout: An Overview

Paul Beatty’s The Sellout showcases a comic genius at the top of his game. A biting satire about a young man’s isolated upbringing and the race trial that sends him to the Supreme Court, it challenges the sacred tenets of the United States Constitution, urban life, the civil rights movement, the father-son relationship, and the holy grail of racial equality - the black Chinese restaurant.

This book had been written in English. It had won Man Booker Prize in 2016, National Book Critics Circle Award for fiction in 2015, Hurston/Wright Legacy Award Nominee for Fiction in 2016, The Rooster - The Morning News Tournament of Books award in 2016. In Goodreads it has got 3.81 star, 21.974 ratings and 3136 Reviews by the users.

2. Literature Review

There has been an increasing research on sentiment analysis in the past few decades. For instance, Turney (2002) applied different methods for detecting the polarity of product reviews and movie reviews respectively. Through this research, it is less clear how sentiment analysis techniques can be employed in the context of social website analysis where the language tends to be more free form and informal. Similarly, Trott (2005) discussed the implementation of RA in the digital world, noting that online RA sites typically involved “online book recommendations and had a live chat and email based readers’ advisory service.”

Book reviews have been considered an important ingredient both by customers and the publishers. Chevalier (2006) found a difference between

Goodreads' and Amazon's e-commerce site is that Goodreads does not link the review-writing process for purchasing decisions in any overt way (the interested Goodreads user can find a link to booksellers, but this information is very discretely positioned on the book's web page). It is worth noting that while Goodreads is owned by Amazon, there appears to be little or no integration between the two services (other than a more prominent link to Amazon.com than to other booksellers). By studying Goodreads and Amazon reviewing behavior at large-scale, they have the opportunity to observe how overt review-commercialization impacts review writing and review construction. Notably, while work has been done on Amazon reviews, this is the first published study to perform large-scale analysis of any kind on Goodreads data. Moreover, Kuzyk (2006) discussed strategies and tools that would allow librarians to offer RA services in the online world, noting that a library's website can be a hub for a wider range of RA features such as roadblocks, annotated lists, staff written reviews and discussion forums. Others have discussed how traditional library resources, such as online catalogs, can be modified to provide better support for RA services. Further, Kuzyk (2006) noted that the Santa Fe Public library used their blog to draw attention to "wall flower" titles. This was done by using cover images from the catalog in the blog posts to add visual appeal to the blog post.

Considering the significance of the sentiment, Zhang (2006) extract both subjective adjectives from SentiWordNet in order to estimate the probability that a document contains opinion-bearing expressions. Similarly, Blitzer (2007) extended a learning model for sentiment classification of a product review data set in cross-domain settings.

Researchers in reported high accuracy in English movie reviews and 72%-83% accuracy in non-English reviews. Subsequently, Devitt (2007) has exploited this resource in combination with WordNet's semantic content for sentiment polarity detection in financial news.

Spiteri (2007) suggested that allowing users to add their own subject tags to catalog records could lead to the formation of informal reading communities, which could be the basis of "user-driven readers' advisory services." Her subsequent examination of sixteen social cataloguing sites identified other features, such as client reviews or ratings or online book clubs, which could be used to add a social dimension to library catalogs.

Librarian application of opinion mining was also observed in the work by Wyatt (2007). According to Wyatt (2007) traditional annotations, or brief notes about the content of a title, are being adapted for use on the web, using Library Thing as an example.¹⁰ Book-oriented social networking sites are often recommended as good resources for RA. Porter and King suggested the use of Amazon.com reviews, Library Thing, and GoodReads as sources of book recommendations for RA librarians, and in a similar vein. Similarly, Zellers(2007) discussed a similar concept, where "each week a different librarian volunteers to compose reviews of any five items in the library, be it new, popular, books or movies." Fahrni (2008) focus on the target specific polarity determination of adjectives. A prior-polarity lexicon of adjectives is derived from SentiWordNet.

Gosling (2009) noted that Australian RA blogs used a range of techniques; including staff reviews, authors month, and Australian author promotions. Further, Stover (2009) examined the expansion of

reader-centered social networking sites and what she called “the wild appeal factor”. Stover looked at several then new sites that might be of interest to readers’ advisors, particularly in terms of how readers talk about books and reading in their own words. As Stover pointed out, the conversation about books is taking place on the web in a variety of forms, and as reflective practitioners, we need to be aware of those conversations happening outside the library walls.

Ross (2011) examines how readers in one online community Goodreads.com discuss books and assist each other in finding “good books” to read in organic, natural, and sometimes messy online discussions and comments. In the course of their interactions on Goodreads, they intuitively employ what they think of as RA tools and strategies, such as those discussed by Catherine Ross, Lynne McKechnie, and Paulette Rothbauer, and the appeal elements, as defined by Joyce Saricks, but in an informal way, what Stover terms readers’ advisory in the wild. Trolley (2011) noted that this type of catalog enhancement is likely to require collaboration between technical and public services staff. She suggested that one possible side effect of this might be increased awareness of RA needs by catalogers, which could encourage them to supplement catalog records with additional genre or other descriptive terms. Tarulli and Spiteri extended this concept by suggesting that catalog enhancements could also include recorded book discussions, perhaps led by an RA specialist. Some of the discussion about using online tools for RA has focused on non-library sites.

Examining Goodreads features that could potentially be useful for RA work, Naik (2012) concluded that this type of service provided opportunities for

readers to be more involved in discussions about books and reading. This aspect appears to be an important difference between RA services that use digital services and the traditional face-to-face approach. Further, Vnuk (2012) identified sites containing useful resources for RA librarians, such as Book Group Buzz, Children’s Literature Network, Fantastic Fiction, and Reading Group Guides. Zuccala (2014) in her study, she mentioned that in coupling the data to a classification of publishers shows that commercial presses have published 42 percent of the books that have achieved both academic and social impact.

3. Methodology

The data on book reviews was collected manually for the study from the Goodreads website (www.goodreads.com). Booker prize award 2016, *The Sellout* by Paul Beatty was considered for the study, which had a total of 3,150 reviews. However, for the present study, the first 10 pages of the reviews had been considered for the study. Each page consists of thirty reviews. Data was collected for the period of January 09, 2015 to April 25, 2017. Total of 292 reviews was collected from the Goodreads website. Reviews in English language were only selected for analysis, hence 8 reviews have been excluded from the study as it was not in English language. Out of 292 reviews, 189 reviews occurred before the announcement of the result, and 103 reviews occurred after the announcement of the result. The extracted text preparation is clean by converting plural to singular words, removing punctuations and numbers, URL, symbols, spaces etc. before any analysis. Packages such as, plyr, ggplot, stringr, tm and wordcloud were installed into R programming. Finally the accuracy of Naive Bayes sentiment score was at 81 percent.

4. Results

4.1 Sentiment Polarities

Opinions and feelings are expressed in different ways, including the amount of details given, type of vocabulary used, context of writing, slangs and linguistics variations are just a few examples. This makes manual analysis tedious, and almost impossible. But, with sentiment analysis, innovative text analytics and natural language processing is employed to extract and classify data.

Based upon the objectives formed the results has been explained below.

Each review has been analysed with the sentiment function in R and the score for each review has been derived. If the score < 0 , the review is said to have negative sentiment, if the score > 0 , the review is said to have positive sentiment and if score = 0, the review is said to be neutral.

Based upon the sentiment analysis, scores ranging from -1 to -48 are considered to be the negative sentiment, while scores ranging from 1 to 23 are considered to be the positive sentiment.

Table 1: Negative scores

Negative scores	Total number
-48	1
-26	1
-20	2
-19	1
-18	4
-16	1
-15	5
-14	2
-12	3

-11	4
-10	4
-9	4
-8	2
-7	4
-6	10
-5	11
-4	13
-3	15
-2	20
-1	29
Total	137

Table 2: Positive scores

Positive scores	Total number
1	30
2	25
3	17
4	15
5	7
6	6
7	7
8	4
9	1
10	5
11	1
12	1
13	1
14	1
18	1
23	1
Total	123

3: Table Neutral scores

Neutral score	Total number
0	32

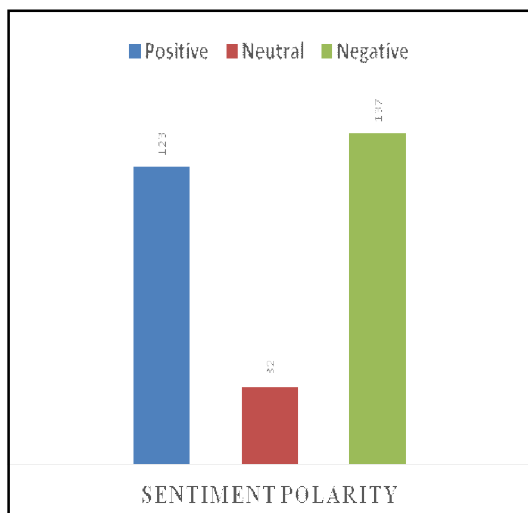


Figure 1: Sentiment polarity of the book reviews

Table 4: Sentiment Polarity of the reviews in percentage

Sl. No.	Sentiment Polarity	Total Number of Reviews	Percentage
1.	Positive	123	42.1%
2.	Neutral	32	10.9 %
3.	Negative	137	47%
	Total	292	100 %

From the figure 1, it is evident that the book “The Sellout” has more negative reviews. Maximum (47%) of the users have critically reviewed the novel, while 10.9% of the users have reported it with neutral reviews. However, 42.1% of the users have reviewed it with positive reviews.

From the above table 4 and figure 1 it is evident that the users have both negative and positive review

on the book “The Sellout”. Negative sentiment is on the higher side when compared to positive sentiment.

4.2 Sentiment analysis before and after the result

4.2.1 Sentiment analysis before the result

Sentiment polarity of the reviews are analysed to understand whether the reviews are of positive or negative sentiment. Out of 292 reviews, 189 reviews occurred before the announcement of the result.

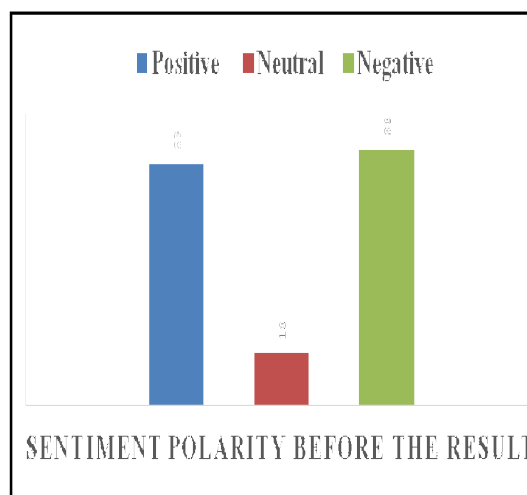


Figure 2: Sentiment polarity before the announcement of the result

Figure 2 shows that majority of the user’s reviews before the announcement of the results was negative reviews indicating that the readers are not satisfied with theme of the book.

4.2.2 Sentiment analysis after the result

Out of 292 reviews, 103 reviews occurred after the announcement of the result.

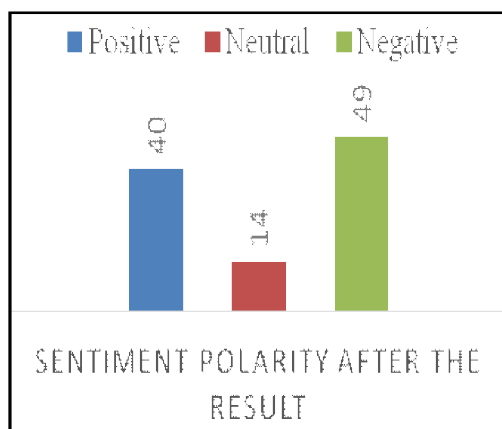


Figure 3: Sentiment polarity after the announcement of the result

Figure 3 shows result of reviewer's sentiment polarity after the result of the award. It was found that there was no change in the reviewer's criticism, a large number of reviewers continue to review the novel critically with evidences of negative sentiment being more than the positive sentiment. Before the result of the award, there were 83 positive sentiments, 18 neutral and 88 negative. Subsequently, the number reduces to 40 positive sentiments, 14 neutral and 49 negative. Though the sample of review before and after the results were different the difference between positive and negative after the result indicates an inclination towards more of negative sentiments.

Table 5: Sentiment polarity of before and after result announcement

	Positive	Neutral	Negative
Before result	83	18	88
After result	40	14	49

4.3 Most frequent word in the reviews

The Most frequent words used in book reviews have been shown in the Table 6. The first 20 words with maximum frequency have been listed below.

Table 6: Most frequent words occurred in reviews

Sl. No.	Word	Frequency
1.	book	531
2.	read	279
3.	black	251
4.	beatty	214
5.	satire	176
6.	novel	164
7.	race	147
8.	sellout	139
9.	funny	135
10.	people	126
11.	white	125
12.	racism	99
13.	reading	99
14.	america	91
15.	dickens	89
16.	time	86
17.	narrator	82
18.	little	81
19.	world	78
20.	father	76

The word "book" occurred 531 times, the word "read" occurred 279 times, "black" occurred 251 times, "beatty" occurred 214 times, "satire" occurred

176 times, “novel” occurred 164 times, “race” occurred 147 times, “sellout” occurred 139 times, “funny” occurred 135 times, people occurred 126 times, “white” occurred 125 times, both “racism” and “reading” occurred 99 times, “america” occurred 91 times, “dickens” occurred 89 times, “time” occurred 86 times, “narrator” occurred 82 times, “little” occurred 81 times, “world” occurred 78 times, “father” occurred 76 times.

4.4 WordCloud

WordCloud of frequently used words in the reviews is considered for visual understanding/analysing the data. The most occurred word in the reviews has been made as wordcloud for the easy visualization.

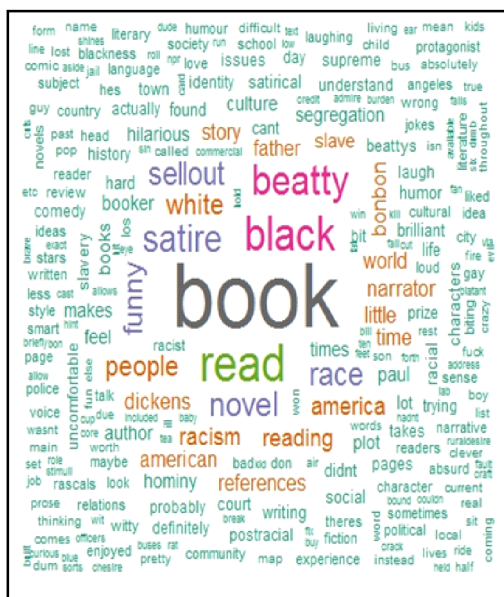


Figure 4: Wordcloud

From the figure 4, it is clear that the word “book” has been used more in the reviews, which is followed by the word “read”, “black”, “beatty”, “satire”, “novel” etc.

5. Conclusion

The number of reviews for different book varies, but the reviews provide accessible and plentiful data for relatively easy analysis for a range of applications. This paper seeks to apply and extend the current work in the field of natural language processing and sentiment analysis to data retrieved from Goodreads. Sentiment analysis attempts to determine which features of text are indicative of its context (positive, negative, neutral, etc.) and build systems to take advantage of these features. The problem of classifying text as positive or negative is not the whole problem in and of itself, but it offers a simple enough premise to build upon further. Reviews were manually taken from the Goodreads website on the 2016 man Booker prize award winner “The Sellout”. Reviews were analysed in R. Text analysis and sentiment polarity were developed in R using various functions. Sentiment scores have been developed for each review using sentiment function. Positive, Neutral and Negative sentiment were calculated separately. On the whole, negative sentiment, is more than the positive sentiment. Word frequencies were also formed and the most occurred word in the reviews has been tabulated. Wordcloud was formed to visualise the most occurred word in the reviews. Public sentiments which are difficult to comprehend differs with the expert, at least in this case of Goodreads users. Popular reading differs from acclaimed or accomplish novel. Hence, user driven acquisition is a popular choice among Librarian and it is believed that sentiment analysis has an important role to play.

References

1. Blitzer, J., Dredze, M., & Pereira, F. (2007, June). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL* (Vol. 7, pp. 440-447).
2. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3), 345-354.
3. Devitt, A., & Ahmad, K. (2007, June). Sentiment polarity identification in financial news: A cohesion-based approach. In *ACL* (Vol. 7, pp. 1-8).
4. Fahrni, A., & Klenner, M. (2008, April). Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Proc. of the Symposium on Affective Language in Human and Machine, AISB* (pp. 60-63).
5. Gosling, M., Harper, G., & McLean, M. (2009). Public library 2.0: some Australian experiences. *The Electronic Library*, 27(5), 846-855.
6. Kuzyk, R. (2006). AREADER AT EVERY SHELF- Book blogs, readers' advisory newsletters pushed via email, personalized lists, and more! A series of new tech tools and library innovations are helping librarians take. *Library Journal*, 131(3), 32-37.
7. Naik, Y. (2012). Finding good reads on Goodreads: Readers take RA into their own hands. *Reference & User Services Quarterly*, 51(4), 319.
8. Ross, C. S., McKechnie, L. E., & Rothbauer, P. (2006). Reading matters: What the research reveals about reading, libraries and community.
9. Spiteri, L. F. (2007). The structure and form of folksonomy tags: The road to the public library catalog. *Information technology and libraries*, 26(3), 13.
10. Stover, K. M. (2009). Stalking the wild appeal factor: readers' advisory and social networking sites. *Reference & User Services Quarterly*, 48(3), 243-269.
11. Tarulli, L. (2011). Readers' services and the library catalog: coming of age fiction? Or non-fiction?. *Reference & User Services Quarterly*, 51(2), 115-119.
12. Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
13. Trott, B. (2005). Advising readers online: A look at Internet-based reading recommendation services. *Reference & User Services Quarterly*, 44(3), 210-215.
14. Vnuk, R. (2012). Readers' Advisory Website Roundup. *Public Libraries*, 51(3), 56.
15. Wyatt, N. (2007). 2.0 for Readers. *Library Journal*, 132(18), 30-33.
16. Zellers, J. (2007). In blog heaven: A painless new approach to readers' advisory. *Virginia Libraries*, 53(3), 23-24.
17. Zhang, E., & Zhang, Y. (2006, November). UCSC on REC 2006 Blog Opinion Mining. In *TREC*.

18. Zuccala, A., Someren, M., & Bellen, M. (2014). A machine-learning approach to coding book reviews as quality indicators: Toward a theory of megacitation. *Journal of the Association for Information Science and Technology*, 65(11), 2248-2260

About Authors

Mr. N Kadiresan, Research Scholar, Department of Library and Information Science, Pondicherry University
Email: nkadiresan1992@gmail.com

Ms. Tripti Gogoi, Research Scholar, Department of Library and Information Science, Pondicherry University
Email: tripti10aug@gmail.com

Dr. Mangkhollen Singson, Assistant Professor, Department of Library and Information Science, Pondicherry University
Email: manglien@gmail.com