

The Rise of Bibliosocial Networking Sites: Goodreads and the Repurposing of Data

Mackenna Swing

SLIS, KENT STATE UNIVERSITY

Abstract

This paper explores the ways in which Goodreads' utilizes and repurposes data by incorporating machine learning algorithms and use of the Semantic web to for optimal user consumption. Goodreads' open-source collection of 1.5 billion books is accessed by 55 million users who have reviewed and rated over 50 million books. A massive quantity of metadata about books is generated by users through their reviews, ratings, categorization, and discussion, which enriches Goodreads' information collection and creates a stronger recommendation algorithm for users. Information is also maintained and edited by Goodreads administrative boards and members with Librarian Status. This information collection can be incorporated into research and scholarly databases, like NoveList and OPAC databases used by public libraries. Goodreads' also helps boost author and publisher's book sales, and increases awareness within the Goodreads community.

The Rise of Bibliosocial Networking Sites: Goodreads and the Repurposing of Data

Goodreads was founded in 2007 as a place for book lovers to converge in discussion, to track personal progress with books, to organize past, current, and future reads, and to receive personalized book recommendations. Since 2007, Goodreads' has grown into a large-scale information service whose data is incorporated into larger research databases worldwide.

Overall Focus and Goals of the Project or Service

Goodreads is an Amazon company founded by Otis Chandler (Wood, 2013) and is the first social media platform to encourage users to interact with one another based on a shared love of books (Zuccacala et. Al., 2015). It is also the "largest website for readers and book recommendations worldwide," with 55 million members, 1.5 billion books added, and 50 million reviews (About Goodreads, 2016). Their mission is to connect people to books and those with similar interests by allowing users to create personal libraries with digital shelves while also providing users with tools to organize and track their past, present, and future reads and incorporating and tracking user input through the collection of quantitative data, qualitative data, and metadata to improve the service.

Sources of Data

The majority of data is obtained by Goodreads through user input, including ratings, reviews, categorization, and discussion. Goodreads uses open data sourcing, which creates

equitable access and contribution by and from users. After Amazon bought Goodreads, a more international perspective for information sharing became a priority, and Goodreads' director of community Patrick Brown advocates for "mak[ing] the site reflect wherever you are in the world," (Wood, 2013). This has been described by critics as "approaching the insurmountable" (Wood, 2013), by combining Goodreads' recommendation database with Amazon's own databases.

Data is also contributed heavily by users, through active collaboration by submitting book reviews and ratings, creating personal libraries to be shared with the community with categorized "virtual shelves," and sharing reading habits and progress. Book reviews take the form of a five-star rating scale, as well as posting comments. This information is visible to the public, so users may read others thoughts and opinions. Members of the public collaborate by becoming "Goodreads Librarians," who have the abilities to edit and maintain information about the author data and book data (Librarian Manual, 2016).

Tools and Technologies

The amount of consumer data shared between Goodreads and Amazon is a massive quantity, which is why a recommendations algorithm was put in place. Described by Brown as a "tremendous technological feat" (Wood, 2016), Goodreads gathers data that is highly valuable to publishers.

Goodreads acquired the start-up Discoverreads.com, which uses machine learning algorithms to analyze and predict book recommendations for users based on past reads, books

they are currently reading, and books they would like to read, as well as what other people who like similar books are interested in (Miller, 2011). This specialized algorithm which combines data from multiple fields creates what founder Chandler refers to as an “an online version of walking into a friend’s living room and scanning the bookshelves to get recommendations” (Miller, 2011). Goodreads combines the intelligent capabilities of technology with the personification of a social network to provide users with tailored recommendations based on their interests and the interests of “friends” online.

Users and Uses

Goodreads’ Users and Members

The majority of Goodreads users are readers and lovers of books, as well as people with interests in education, the humanities, literature, research, and library science. Prior to using Goodreads, users most likely have experience working on social media networks, and are familiar with features like profile pages and avatars, posting features and capabilities. This familiarity provides an easy in accessibility and navigation of the site, and therefore encourages engagement between users in a social context

Members of the public have the ability to contribute and access data, in terms of five-star ratings, posted reviews, sharing of content, and contributing to discussion boards and forums, as long as they have created a Goodreads account. Users also have the ability to become a “Goodreads Librarian,” which requires a minimum of 50 “shelved” books, as well as a submitted application through Goodreads website which is reviewed by a Goodreads administrator. If

selected for Librarian Status, members may “edit book and author data, add book covers, and combine different editions of books” (Librarian Manual). The Librarians are responsible for maintaining accuracy and consistency amongst Goodreads’ book collection.

Uses of Goodreads

The majority of questions being asked on Goodreads’ pertain to literature, both fiction and non-fiction. Recommendations are the answers to these questions, as well as community dialogue between Goodreads’ members. By having users rate and review books, as well as connect to friends and readers with similar interests, Goodreads uses an algorithm to generate personalized reading recommendations. Goodreads functions in cycle by suggesting books for readers, then providing a space for discussion about those books.

Incorporation of Goodreads’ Into Larger Databases

Goodreads’ user reviews are incorporated amongst a variety of different databases, including research databases like NoveList, a division of EBSCO Publishing (EBSCO, 2011), and public library OPACs (Naik, 2012). Reading Advisory librarian Neal Wyatt (Naik, 2012) advocates for an embracing of “bibliosocial networking sites” by librarians to provide quality service to their patrons.

Impact

Research on the altmetric value of Goodreads, or the “study and use of non-traditional scholarly impact measures based on activity in web-based environments” (PLOS, 2016) has been

conducted to better understand Goodreads' impact. Studies by Zuccala et. Al. (2015) have concluded that although Goodreads' is most often associated with fictional literature, many texts of scholarly non-fiction, including research articles by university professors are also being reviewed and rated by users. This encourages the expansion of Goodreads' to move beyond recreational literature recommendations, and potentially serve as a research data base for scholars and professionals.

Goodreads' has also increased sales for authors, especially in the case of authors who have won a "Goodreads Choice Award," which is the only major book award chosen by readers (Impact, 2015). Winning a Goodreads Choice Award can "book a book—even a bestselling book—into higher levels of awareness and drive more sales" (Impact, 2016). The creation of an award dedicated to authors and publishers selected by readers has caused an escalation of sales for both print and e-book, as well as a greater awareness and presence to readers.

Conclusion

Goodreads' has exceeded its initial goal of serving as a bibliosocial platform for readers, and has quickly become the top website for book lovers and avid readers while compiling massive amounts of information about its users and their preferences. Research suggests that Goodreads is becoming more embedded in non-fiction literature, and being used to review scholarly articles and research. Goodreads' is also being incorporated into large-scale databases, in both leisure and academic settings, including University research databases and public library OPACs.

References

Altmetrics. (2016). Retrieved from <http://collections.plos.org/altmetrics>

About Goodreads. (2016). Retrieved from <https://www.goodreads.com/about/us>

EBSCO Publishing. (2011). *Reader Ratings and Book Reviews from Goodreads Being Added to NoveList Products* [Press Release]. Retrieved from <https://www.ebscohost.com/uploads/general/NoveList-Goodreads.pdf>

The Impact of Goodreads Choice Awards for Authors and Publishers. (2015). Retrieved from <http://www.goodreads.com/blog/show/594-the-impact-of-goodreads-choice-awards-for-authors-and-publishers>

Librarian's Manual. (2016). Retrieved from https://www.goodreads.com/librarian_manual

Miller, Claire Cain (March 10, 2011). "Need Advice on What to Read? Ask the Internet". *New York Times Bits*. Retrieved from <http://bits.blogs.nytimes.com/2011/03/10/need-advice-on-what-to-read-ask-the-internet/>

Naik, Yesha. (2012). Finding Good Reads on Goodreads. *Readers' Advisory* 51 (4) 1-6.

Wood, Felicity. Good vibrations: Felicity Wood meets Patrick Brown, Goodreads' direct of

community, to talk Amazon, analytics, and algorithms. *The Bookseller* (5573), 14-15.

Zuccala, Alesia A., Verleypsen Frederik T., Cornacchia, Roberto and Engels, Tim C.E. (2015).

Altmetrics for the humanities. *Aslib Journal of Information Management*, 67 (3) 320-336