LDA-Based Study of Content Themes in Reading Promotion on WeChat Platforms of Chinese Public Libraries (Postprint)

Authors: Li Qian, Wang Shuai

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] To identify the content themes and evolutionary characteristics of reading promotion conducted by domestic public libraries via the WeChat Official Platform. [Method/Process] We collected posts published by the WeChat accounts of 31 provincial-level public libraries in China, defined and selected reading promotion content, and applied the LDA topic model to achieve theme identification, evolution analysis, and view count analysis of the reading promotion content. [Results/Conclusion] It was found that the reading promotion content themes currently conducted by public libraries via WeChat can be divided into nine major categories, covering a relatively broad range. Each theme exhibits different evolutionary trends over time, and the view counts for each theme also reflect different levels of popularity. Recommendations are made for the future WeChat reading promotion efforts of public libraries from the perspectives of scientific planning and proportioning of promotion content, reversal of the subordinate status of promotion content, setting up promotion content for special groups, innovative design of promotion content and formats, and building brand promotion content.

Full Text

A Study on the Thematic Content of Reading Promotion on WeChat Platforms of Public Libraries in China Using the LDA Model

Li Qian[1],[2], **Wang Shuai**[2] [1]School of Economics and Management, Shanxi University, Taiyuan 030006 [2]Institute of Scientific and Technical Information of China, Beijing 100038

Abstract: [Purpose/Significance] This study aims to identify the content themes and evolutionary characteristics of reading promotion activi-

ties conducted by domestic public libraries via WeChat public platforms. [Method/Process] We collected articles posted by 31 provincial public libraries on their WeChat accounts, defined and selected reading promotion content, and employed the LDA topic model to identify themes, track evolution, and analyze reading volumes. [Result/Conclusion] The findings reveal that current reading promotion content on WeChat by public libraries can be categorized into nine major themes with broad coverage. Each theme exhibits distinct evolutionary trends over time, and reading volumes vary significantly across themes, reflecting different levels of popularity. Based on these results, we propose recommendations for future WeChat-based reading promotion efforts, including scientific planning of content 配比, reversal of content subordination, targeted content for special populations, innovative design of content and formats, and development of branded promotional content.

Keywords: LDA model; WeChat public platform; reading promotion; theme mining **Classification Number:** G250 **DOI:** 10.13266/j.issn.0252-3116.2022.08.008

Libraries, as vital institutions for disseminating knowledge, transmitting culture, and preserving civilization, possess abundant literature resources and provide diverse access services for readers. Effective reading promotion by public libraries stimulates public interest in reading and fosters reading behaviors, thereby cultivating good reading habits and a positive reading atmosphere across society. As China's largest social media platform, WeChat reached 1.262 billion monthly active users in the third quarter of 2021. With its powerful interactive features, WeChat has become an indispensable social APP in people's lives. Public libraries' utilization of self-media platforms, applying their efficient, convenient, interactive, and sharing characteristics to library promotion and communication, plays an irreplaceable role in advancing reading promotion efforts. Conducting in-depth and detailed mining analysis of reading promotion articles published on WeChat public platforms is therefore essential. This study selected 31 provincial public libraries in China, collected and analyzed articles and related data on reading promotion posted on their WeChat platforms, and employed text mining methods to analyze the thematic content and evolutionary characteristics of reading promotion.

2 Definition of WeChat Reading Promotion Content

Numerous scholars have provided definitions of reading promotion. For instance, Fan Bingsi considers library reading promotion in China as a collective term for library marketing and new types of library reading services. Zhang Huaitao defines reading promotion as the promotion of reading—briefly, activities conducted by social organizations or individuals to encourage reading; more specifically, activities that expand the influence of reading and enhance people's willingness and conditions to participate in reading. Wang Bo describes reading promotion as activities aimed at fostering universal reading, improving human cultural quality, enhancing national soft power, and accelerating national pros-

perity, with the strategic goal of cultivating public reading interest, habits, quality, ability, and effectiveness. These definitions indicate that reading promotion encompasses a broad scope, including not only the promotion of directly or indirectly readable book resources but also any information that helps cultivate reading methods, enhance reading literacy, stimulate reading interest, develop reading habits, and provide reading enjoyment.

Based on these definitions and existing scholarship, this study defines WeChat platform reading promotion content by public libraries as any information that helps cultivate readers' reading methods, enhance reading literacy, stimulate reading interest, develop reading habits, and provide reading enjoyment, in addition to the promotion of directly or indirectly readable book resources. The identification of reading promotion information can be based on two aspects: (1) content that readers can directly read (micro-reading, books, periodicals), and (2) information that promotes users' reading literacy (interest, habits, methods, ability, ethics).

Not all content published by libraries on WeChat platforms constitutes reading promotion material; it also includes library announcements, closure notices, regional news, etc. Therefore, reading promotion content must be identified and filtered. Previous research has addressed this issue. For example, Wan Muchen et al. argue that reading promotion should not be limited to book recommendations, reading contests, or lectures but should be expanded to include bookmark design, photography exhibitions, and quiz activities. They consider all content on library WeChat platforms—except library announcements, work notices, information queries, services, and other difficult-to-categorize information—as reading promotion-related. Cai Liping et al. similarly define reading promotion content on library WeChat platforms as everything except library announcements, work notices, conference reports, information queries, service instructions, holiday greetings, and other difficult-to-categorize information. Wang Baoying et al. categorize reading promotion into dynamic activity promotion and static content promotion, focusing their analysis of university library WeChat platforms on static resources such as book recommendations, borrowing rankings, reading lists, book reviews, new book recommendations, journal recommendations, e-book lists, beautiful essays or inspirational content, activity achievements, reading anecdotes or methods, and knowledge recommendations.

3 Research Design

3.1 Research Subjects

This study selected the WeChat public platforms of 31 provincial-level public libraries in China as research subjects. Data collection began on August 25, 2020, covering the period from November 1, 2019, to July 31, 2020 (a total of 273 days). Data were collected using the Qingbo Intelligent Platform and the WeChat mobile client. The Qingbo Intelligent Database contains most WeChat public platform data and provides monitoring functions. Table 1 shows the

names and launch dates of the 31 provincial public library WeChat accounts.

Table 1 Basic Information of Provincial Public Library WeChat Accounts (Sorted by Launch Date)

Library Name	Launch Date	Library Name	Launch Date
Shanghai Library	2013-11-23	Guangdong Provincial Sun Yat-sen Library	2014-03-18
Chongqing Library	2013-05-07	Hainan Library	2014-03-28
Xinjiang Library	2013-05-25	Shanxi Library	2014-03-31
Hubei Provincial Library	2013-12-20	Hebei Library	2014-08-07
Nanjing Library	2014-05-20	Jilin Library	2014-09-17
Shaanxi Library	2014-01-26	Gansu Library	2014-10-15
Hunan Library	2014-02-12	Guizhou Library	2014-10-24
Liaoning Provincial Library	2014-02-15	Zhejiang Library	2014-10-28
Henan Library	2014-06-09	Tianjin Library	2015-03-04
Shandong Library	2014-06-10	Jiangxi Library	2015-04-20
Heilongjiang Provincial Library	2014-03-13	Qinghai Library	2015-04-23
Guangxi Zhuang Autonomous Region Library	2014-04-18	Yunnan Library	2015-04-28
Inner Mongolia Library	2014-04-22	Anhui Library	2015-07-09
Fujian Provincial Library	2014-04-23	Sichuan Library	2015-07-16
Ningxia Library	2016-03-15		

3.2 Research Methods

This study employs the LDA (Latent Dirichlet Allocation) topic model for thematic analysis of reading promotion content. LDA is an unsupervised machine learning technique that uses a three-layer Bayesian probability model to identify latent thematic information in large-scale documents. It is based on the following assumptions: (1) there are K topics in the document collection, and topics are independent of each other; (2) each document is randomly mixed from K topics, with topic parameters following a Dirichlet distribution; (3) each topic is a multinomial distribution over feature words, with parameters also following a Dirichlet distribution. The core idea is that a document selects a topic with

a certain probability and then selects a word from that topic with a certain probability. Thus, a document represents a probability distribution over topics, and each topic represents a probability distribution over words. LDA calculations yield two probability distributions: document-topic and topic-word. The document-topic distribution provides the support weight of each document for each topic, where a larger weight indicates stronger association. The topic-word distribution is represented by a series of feature words and their probability values in each topic, reflecting the internal structure of topics. A higher probability value indicates greater contribution of the word to the topic.

In LDA, the generation process of a document's "document-term" structure is illustrated in Figure 1 [Figure 1: see original paper]. First, the topic distribution for document m is generated from a Dirichlet distribution with hyperparameter α. Then, for each word position n in the document, a topic z , is sampled from . Next, the word distribution ϕ for topic z , is generated from a Dirichlet distribution β. Finally, word W , is sampled from ϕ . Repeating this process n times generates a complete document.

In LDA experiments, the number of topics must be predetermined. Perplexity is commonly used to determine the optimal number of topics. In information theory, perplexity measures the goodness of a probability distribution or probability prediction sample and has become an important indicator for determining the optimal topic number. The calculation formulas are shown in (1) and (2):

$$\text{Perplexity}(D) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(w_d)}{N_d}\right) \quad (1)$$

$$p(w_d) = \prod_{i=1}^{N_d} \sum_{z} p(w_{d,i}|z)p(z|d) \quad (2)$$

where N_d represents the vocabulary collection and document length, M is the number of documents, p(w_d) is the document generation probability, w_{d,i} is the i-th word in document d, and z is a specific topic. Based on formula (1), the "exp()" function uses natural constant e as its base. Following Qiu Junping and Shen Chao's research and Python code conventions where "log()" defaults to base e, the log in formula (1) is base e. In formula (2), $z\ p(w$\{d,i\}|z)p(z|d) represents the generation probability of the i-th word in document d. Due to the bag-of-words model, each word's probability is conditionally independent, so document d's generation probability p(w_d) is the product of each word's generation probability.

LDA is a commonly used classification model for text topic classification. Multiple software tools can implement topic modeling and mining. This study primarily uses the LatentDirichletAllocation library in Python's Scikit-learn toolkit to

build the model, obtaining topic-word distributions and document-topic distributions for reading promotion content analysis. The data collection and analysis process is shown in Figure 2 [Figure 2: see original paper].

3.3 Data Collection and Selection

Due to the high technical threshold of crawler systems and difficulties in implementation, this study combined manual and tool-based collection methods. The collection steps were: (1) Using the WeChat mobile client to follow the 31 provincial public library accounts. Through the "Add Friends" function, we searched for each library's name. To ensure authenticity and avoid fake accounts, we verified each account's registration information through the Qingbo Intelligent Platform before following. (2) Manually obtaining article URLs, read counts, and "like" counts using the WeChat mobile client. In each account's "Historical Messages" list, we opened articles, copied links via the hidden menu, and recorded read counts and "like" counts at the bottom of each article. These were saved in Excel files, resulting in 31 files with 11,131 article records. (3) Using the Octopus Collector to obtain article text content and related information. We imported the article URLs into Octopus Collector, set up custom collection modules and processes to batch-collect information including titles, publication times, and main text. All collected data were integrated using Excel, with each provincial library as a separate file. Figure 3 [Figure 3: see original paper] shows sample data from Hunan Library's WeChat account.

During collection, some data could not be successfully obtained because WeChat articles contain not only text but also images, audio, and video, and text formatting also affected collection. Therefore, the first step in data selection was removing records with empty main text content, totaling 246 articles. Since not all information published by libraries constitutes reading promotion content (including news announcements and work notices), we manually screened reading promotion information according to the conceptual definition above. By reviewing article titles, we filtered out 1,009 non-reading promotion articles, leaving 9,876 reading promotion articles.

3.4 LDA Topic Discovery for Reading Promotion Content

3.4.1 Text Preprocessing

(1) Removing special symbols and formatting. Original text data often contains English, numbers, special symbols, and emojis that contribute little to text analysis and interfere with machine understanding. We used Excel's intelligent toolkit to batch-remove these elements, creating clean text.

(2) Building a stopword list. Removing stopwords is crucial in machine learning text classification. Too many irrelevant words degrade algorithm performance. LDA modeling requires extracting document feature words, but unstructured documents often contain high-frequency yet meaningless words such as prepositions, conjunctions, and pronouns like "his,"

"my," and "they." While specialized stopword lists exist, directly applying them to different document collections often yields suboptimal results. Therefore, we manually constructed a stopword list based on the "Baidu Stopword List," combined with part-of-speech tagging and word frequency statistics. We removed high-frequency words like "library," "reading," and "book," and included conjunctions, interjections, numerals, prepositions, and pronouns, resulting in a final stopword list containing 2,318 words.

3.4.2 Feature Extraction Text feature extraction is essential in LDA modeling. Computers cannot understand text meaning directly, so text must be converted into quantifiable feature words through weight calculation for further mining. This study uses TF-IDF (Term Frequency-Inverse Document Frequency) for feature extraction. TF-IDF is an information retrieval weighting technique where TF represents term frequency and IDF represents inverse document frequency. It uses the ratio of a feature word's occurrence in a document to the number of documents containing that word as the weight, commonly used for text classification. We used the TfidfVectorizer parameter in Python's Scikit-learn module. Given the large volume of text data containing numerous words, many of which are insignificant for topic extraction and time-consuming to process, we set n_{features} = 3000 to extract the 3,000 most important feature keywords, resulting in a 9,876 × 1,136 feature word sparse matrix.

3.4.3 LDA Topic Modeling Based on the stopword list and topic feature extraction, we used the LatentDirichletAllocation library in Scikit-learn to build the topic model, with hyperparameters α and set to default values and maximum iterations max_{iter} = 1000. LDA requires pre-setting the number of topics. The authoritative method uses perplexity calculation, an important indicator for determining optimal topic number. Using lda.perplexity, we identified the optimal topic count. As shown in Figure 4 [Figure 4: see original paper], when the topic number is 9, the model's perplexity is relatively low. Combined with pyLDAvis visualization results (Figure 5 left panel), when K = 9, topic overlap is minimal, indicating good classification results. Therefore, the final topic number K was determined to be 9. The right panel of Figure 5 [Figure 5: see original paper] shows the top 30 words most correlated with Topic 3. Different topic numbers can be entered in the "Selected Topic" text box to display top-ranked words for each topic.

4 Data Results and Analysis

4.1 Topic Induction and Analysis of Reading Promotion Content

LDA classification yielded nine themes for library WeChat reading promotion content. We compiled the top 15 high-probability feature words for each theme and described scenarios based on these words to summarize the most appropriate theme labels. For example, in Topic 0, high-probability feature words like "e-book," "sharing," "search," "pandemic," and "website" fit the scenario of rec-

ommending and accessing digital resources online during the pandemic when people stayed home. Thus, Topic 0 was labeled “Digital Resource Promotion and Access.” In Topic 1, words like “children,” “kids,” “parent-child,” “parents,” and “co-reading” closely relate to family learning, reflecting parent-child education scenarios. Therefore, Topic 1 was labeled “Family and Parent-Child Education.” In Topic 4, words like “poetry,” “cultural heritage,” “folk customs,” and “traditions” describe Chinese traditional poetry culture and folk festivals, so Topic 4 was labeled “Traditional Culture.” Following this process, provincial public library reading promotion themes were summarized into nine categories, as shown in Table 2 .

Table 2 Topic-High Probability Feature Word Distribution

Topic ID	Theme Label	Top 15 High-Probability Feature Words
Topic 0	Digital Resource Promotion and Access	ancient books, universal, e-books, calligraphy, pandemic, awards, launched, sharing, search, special topics, massive, children, citizens, website, masters
Topic 1	Family and Parent-Child Education	children, kids, parent-child, parents, co-reading, children’s library, education, cultivation, growth, sister, interest, English, childhood, storytelling, teenagers
Topic 2	Life and Emotions	father, growth, son, women, love, youth, thinking, animals, daughter, booklist, young people, travel, growing up, emotions, loneliness
Topic 3	Science Popularization	children, challenge, animation, health, elementary school, universal, arrangement, Q&A, links, friends, staying home, lottery, science popularization, pandemic, universe
Topic 4	Traditional Culture	poetry, museum, cultural heritage, folklore, folk, customs, archaeology, cultural relics, crafts, food, festivals, ancient times, seasons, plants, traditions
Topic 5	Pandemic Prevention and Health	pandemic, prevention, pneumonia, coronavirus, virus, anti-pandemic, masks, protection, infection, health, cheer, people, frontline, salute, prevention
Topic 6	Poetry and Art Appreciation	music, poetry, poet, Chinese classics, poetry, famous teachers, France, classical, aesthetics, ancient, comedy, style, masters, artists, performance

Topic ID	Theme Label	Top 15 High-Probability Feature Words
Topic 7	Patriotic Education Theme	Communist Party of China, people, original aspiration, celebration, comrades, peace, patriotism, youth, mission, victory, commemoration, war, socialism, struggle, belief
Topic 8	Economics, Politics, Science, and Philosophy	construction, education, economy, theory, experts, Xi Jinping, politics, fields, mentors, philosophy, law, general secretary, socialism, system, technology

As Table 2 shows, Topic 0 (Digital Resource Promotion and Access) reflects libraries' active embrace of new media trends, leveraging the internet and social media's interactive features and rapid information transmission to strengthen digital resource promotion and services. Topic 1 (Family and Parent-Child Education) shows public libraries utilizing their resources and environmental advantages to conduct family education and guide parental involvement in children's reading, fostering parent-child relationships. Topic 2 (Life and Emotions) primarily targets young people, with content related to youth life, emotions, struggle, and inspiration that resonates with contemporary youth. Topic 3 (Science Popularization) focuses on health and science, using interactive Q&A and lectures for reading promotion. Topics 4, 6, 7, and 8 demonstrate provincial libraries' efforts to inherit Chinese traditional culture and classics, strengthen patriotic education, and promote content on socialism, economics, politics, philosophy, science, and law. Topic 5 (Pandemic Prevention and Health) emerged from the major public health event in early 2020, with libraries actively fulfilling social responsibilities by providing digital resources and online lectures.

4.2 Analysis of Reading Promotion Content Theme Intensity and Evolution

4.2.1 Theme Intensity Analysis Reading promotion theme intensity distribution is shown in Table 3 , where , represents article i's probability distribution across Topic j. Articles 1, 2, 3, and 9875 show maximum probability distribution in Topic 2, article 9874 in Topic 8, and article 9876 in Topic 6.

Table 3 Document-Topic Distribution

Article	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
1	0.0231	0.0230	0.8156	0.0231	0.0230	0.0230	0.0230	0.0231	0.0230
2	0.0129	0.0129	0.8965	0.0129	0.0129	0.0129	0.0129	0.0129	0.0129
3	0.0201	0.0201	0.8389	0.0201	0.0201	0.0201	0.0201	0.0201	0.0201
9874	0.0176	0.0175	0.0309	0.0175	0.0175	0.0176	0.0175	0.0175	0.8462
9875	0.0236	0.0236	0.8110	0.0236	0.0237	0.0236	0.0236	0.0236	0.0236

Article	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
9876	0.0279	0.0278	0.0601	0.0282	0.0279	0.0278	0.7447	0.0278	0.0278

Article 1's title, "Zhejiang Library Koi 2.0 Fully Upgraded! Massive Books, Kindle Reading Device, Travel Packages… Take Poetry on Your Journey," mainly describes book recommendation activities and "culture-tourism integration" travel route promotions, aligning with Topic 2 keywords like "booklist" and "travel." Article 2's title, "If Not for an Accident, the Top Poet of the Tang Dynasty Might Not Have Been Li Bai," introduces Wang Bo's brief but brilliant life and works, aligning with Topic 2 keywords like "growth," "youth," "growing up," and "loneliness." Article 3's title, "China's First Whale Fall Discovered in the South China Sea: One Whale Falls, All Things Live! This is the World's Most Romantic Death," describes how a dead whale sustains 10,000 organisms for about a century, forming a thriving biological community. While inspiring wisdom and interest in "whale fall" content, its poetic artistic expression and lack of explicit scientific keywords led to its classification under Topic 2 rather than Topic 3 (Science Popularization). Article 9875's title, "A Great Talent Created a Rumor That Deceived Us for Hundreds of Years," discusses whether Zhang Fei was rough or handsome in appearance, using historical texts as evidence, aligning with Topic 2 keywords like "father" and "daughter." Article 9874's title, "Four Lectures, Four Loves: Stimulating Tibet's Poverty Alleviation Momentum," introduces an educational campaign, stimulating interest in political literature during the poverty alleviation period, aligning with Topic 8 keywords like "education," "Xi Jinping," "politics," and "technology." Article 9876, "One Recommendation, Heartfelt | Understanding Classical Poetry," focuses on poetry appreciation and recommendation, aligning with Topic 6 keywords like "poetry," "poet," "classical," and "Chinese classics."

Based on the document-topic probability distribution, we calculated the number of reading promotion articles for each theme, as shown in Figure 6 [Figure 6: see original paper]. The distribution shows the following ranking from most to least articles: Topic 2 (3,098), Topic 0 (1,683), Topic 8 (1,289), Topic 5 (1,177), Topic 1 (914), Topic 4 (661), Topic 6 (517), Topic 3 (276), Topic 7 (261).

4.2.2 Theme Intensity Evolution Analysis Theme intensity evolution refers to changes in the number of reading promotion posts (theme intensity) over time. Figure 7 [Figure 7: see original paper] shows the evolution trends of the nine themes from November 2019 to July 2020.

Overall, each theme shows different fluctuations. Topic 5 (Pandemic Prevention and Health) began rising significantly in January 2020 due to the pandemic, with rapid growth from January to February, peaking at 434 posts in February. Similarly affected by the pandemic, Topics 1, 2, 4, 6, 7, and 8 all showed significant declines in February. Topics 0 and 3, however, showed upward trends during this period.

As pandemic prevention measures took effect and results became significant, all themes except Topic 5 gradually recovered and reached small peaks in April. Topic 5 showed an accelerated decline. The recovery in April was partly due to effective pandemic control and partly because April 23 is World Book Day, with libraries designating April as Reading Month and organizing various reading-themed activities. For example, Nanjing Library held "4.23 Cloud Reading: Without Poetry, No Speech—Fragrance of Books Fights the Pandemic and Nourishes the Soul" and online knowledge competitions about the Four Great Classical Novels. Topic 4 (Traditional Culture) showed another small peak in January and a major peak in June, corresponding to traditional holidays like New Year's Day, Spring Festival, and Dragon Boat Festival. Topic 7 (Patriotic Education Theme) remained relatively stable from November 2019 to June 2020, with a significant increase in July, as libraries designated July as Red Culture Month, promoting patriotism through red literature recommendations, classic recitation contests, local history introductions, and revolutionary culture propaganda.

4.3 Analysis of Reading Volume Distribution by Theme

To analyze reader engagement with each theme, we calculated average reading volume per theme (total reading volume divided by number of articles) and examined the distribution of top 100, 200, 500, and 1000 most-read articles across the nine themes, as shown in Table 4 .

Table 4 Theme-Reading Volume Distribution

Theme ID	Average Reading Volume	Top 100	Top 200	Top 500	Top 1000
Topic 0	953.9239453	1008.850109	2693.974822	666.442029	774.075643
Topic 1	1452.297114	633.6213592	376.605364	1017.04422	
Topic 2	2693.974822	666.442029	774.075643	1452.297114	633.6213592
Topic 3	376.605364	1017.04422			
Topic 4	953.9239453	1008.850109	2693.974822	666.442029	774.075643
Topic 5	1452.297114	633.6213592	376.605364	1017.04422	
Topic 6	376.605364	1017.04422			
Topic 7	376.605364	1017.04422			
Topic 8	953.9239453	1008.850109	2693.974822	666.442029	774.075643

Topic 2 (Life and Emotions) ranks first across all metrics, indicating it is the most popular overall. Topic 5 (Pandemic Prevention and Health) ranks second in average reading volume but not in the top rankings for the hottest articles (top 100 and 200), suggesting it is generally popular but lacks extremely viral content. Topic 8 (Economics, Politics, Science, and Philosophy) ranks third overall, with stable performance despite some fluctuations. Topic 1 (Family and Parent-Child Education) has some highly popular articles but ranks fourth

overall. Topic 0 (Digital Resource Promotion and Access) maintains a middle position. Topics 4, 3, 6, and 7 rank lower in overall reading volume.

5 Research Conclusions

5.1 Characteristics Reflected in Reading Promotion Content Categorization

5.1.1 Broad Content Scope Demonstrating Public Service Philosophy Provincial public libraries' reading promotion content covers extensive themes spanning life, emotions, culture, arts, health, politics, economics, science, and philosophy, touching multiple aspects of public learning, work, and daily life. This reflects a public service philosophy. Target audiences vary, sometimes focusing on children, sometimes youth, sometimes the general public. Content shows thematic characteristics for different target groups: Topic 0 targets all citizens, Topic 1 focuses on children and teenagers, and Topic 2 targets youth and women.

5.1.2 Emphasis on Disseminating Cultural Classics and Promoting Excellent Culture Libraries promote traditional culture through reading activities, helping youth understand classic Chinese culture and draw nourishment from excellent traditions. Promoted content includes classics like *The Book of Songs*, *Mencius*, and *The Analects*. Libraries use picture-based children's books to guide reading comprehension, organize classic recitation and poetry appreciation activities, and conduct online and offline cultural lectures to create strong learning atmospheres and lead the public in appreciating traditional cultural classics.

5.1.3 Closely Following Era Development and Actively Assuming Social Responsibility Reading promotion content is not limited to inheriting traditional culture but also aligns with contemporary trends. First, it reflects promotion of socialist economics, systems, and laws through lectures and interpretations of important Party meetings, conveying core values. Second, during the pandemic, libraries actively fulfilled social responsibilities by encouraging home reading through online lectures, training, themed reading, and exhibitions. Utilizing social media's convenience and interactivity, libraries led nationwide home reading initiatives.

5.1.4 Focus on Youth Reading Promotion Content Among target audiences, children and teenagers are the primary focus, with numerous youth-centered activities such as English programs, youth work collections, poetry recitation, coding workshops, classic reading guides, and youth reading clubs. Youth reading is the foundation and future of nationwide reading, requiring active parental guidance. Libraries have conducted family parent-child reading activities like parent-child communication sharing classes, picture book reading

demonstrations, and co-reading sessions to foster good parent-child relationships and reading habits.

5.1.5 Active Promotion of Digital Reading Services Libraries provide convenient resource access through rich digital resources and innovative service methods, using new social media to guide the transition toward combined traditional and digital reading. Especially during the pandemic, digital resource-themed content dominated, with libraries offering digital reading services through reading check-ins, QR codes, digital libraries, resource links, digital exhibitions, cultural training classes, and database recommendations to cultivate digital reading literacy.

5.2 Evolutionary Characteristics of Reading Promotion Content Themes

5.2.1 Adjusting Content Based on Current Events From January 23 to April 8, 2020, societal attention to Topic 5 (Pandemic Prevention and Health) reached unprecedented heights. Libraries reduced content related to Topics 1, 2, 4, 6, 7, and 8 while increasing Topic 5-related posts. Topics 0 and 3, which helped maintain social distance and disseminate pandemic knowledge, also saw increased posting. Topic 7 (Patriotic Education Theme) reflected society's pandemic response, with libraries slightly increasing related content. This demonstrates libraries' social responsibility during public emergencies.

5.2.2 Using Commemorative Days and Holidays as Content Opportunities Figure 7 shows that peaks in reading promotion themes align with World Book Day (April 23), New Year's Day, Spring Festival, Dragon Boat Festival, and Party Founding Day. This indicates that provincial libraries consider holidays in content selection, helping readers develop historical and cultural awareness, enhancing national pride, and securing promotional funding for more impactful online and offline activities.

5.3 Relationship Between Posting Volume and Reading Volume by Theme

Figures 6 and Table 2 show that the ranking of posting volume by theme largely matches the ranking of reading volume, indicating that libraries' content generally meets reader demand. The posting volume ranking is: Topic 2, Topic 0, Topic 8, Topic 5, Topic 1, Topic 4, Topic 6, Topic 3, Topic 7. The reading volume ranking is: Topic 2, Topic 5, Topic 8, Topic 1, Topic 0, Topic 4, Topic 3, Topic 6, Topic 7. Topics 2, 0, 8, 5, and 1 rank in the first tier in both metrics, while Topics 4, 6, 3, and 7 rank in the second tier.

6 Research Implications

6.1 Strengthen Scientific Planning and 配比 of WeChat Reading Promotion Content

Libraries should plan promotional content systematically with clear goals, selection criteria, targeting, and quality standards rather than arranging content randomly. Content selection should coordinate: (1) proportions among the nine themes based on library collections, reader needs, and core responsibilities; (2) balance between old and new book recommendations to maintain freshness while preserving classics and improving underutilized resources; (3) synergy between knowledge introduction and literature recommendations to provide both accessible knowledge and deep reading guidance; (4) appropriate 配比 between current events and regular recommendations to demonstrate social responsibility while highlighting library characteristics. Additionally, libraries should address the contradiction between systematic, in-depth content requirements and fragmented formats by using "series" or "special topics" to divide themes into multiple posts, solving the boredom of long articles while gradually guiding readers from superficial to comprehensive understanding.

6.2 Reversal of WeChat Reading Promotion Content's Subordinate Status During Home Isolation

Previously, offline activities were the main arena for reading promotion. During pandemic isolation, WeChat-based promotion allowed social distancing while keeping people engaged, highlighting its importance. Libraries need to adapt, adopting a mindset that prioritizes WeChat as the main platform. Beyond Topic 0 (Digital Resource Promotion and Access), libraries should carefully select and recommend excellent e-books, reading apps, and databases with greater intensity and depth. Previous offline activities should be transformed into online content to fill gaps and attract users to prefer online promotion methods. Operators need media knowledge and strong writing skills to innovate and operate WeChat accounts professionally, making them powerful tools for reading promotion.

6.3 Strengthen Content Settings for Special Populations

Current reading promotion themes primarily target children, teenagers, and the general public, with limited content for special populations. IFLA's Library Services to People with Special Needs Section defines special populations as "people who cannot use library resources normally," focusing on those unable to use existing services due to living conditions or physical, mental, or cognitive disabilities, including hospital and prison patients, homeless individuals, and nursing home residents. While these groups face difficulties accessing library services, WeChat-based reading promotion is non-location-dependent, universal, and open, breaking offline service limitations. Libraries should increase content for special populations, with professional librarians designing targeted content

considering group characteristics.

6.4 Innovative Design of Promotion Content and Presentation Formats

WeChat content operations follow the path: "attract attention—generate interest—active search—take action—experience sharing." Different content types receive varying user attention. Social hot topics attract more attention than classic literature, requiring innovative design, deconstruction, flexible use of online materials, and humorous language styles to engage users. Multiple expression methods (narration, argumentation, expression) should mobilize users' cognitive processing. Modern mobile reading characteristics include simplified, colloquial language and eye-catching titles, with integrated use of images and videos to highlight key content. Video streaming for "book talks" can also attract readers. Additionally, segmenting and targeting audiences enables more tailored content design that guides deep reading and thinking, improving reading quality.

6.5 Create Branded Content

While addressing diverse social needs, public libraries should combine their service characteristics to create distinctive and branded content. Branded content is an optimal expression of library reading promotion service systems, reflecting service advantages, personality, and philosophy while attracting readers and creating brand effects. WeChat platforms' features allow readers to easily access past articles and participate more readily than offline services. Libraries should determine branded content based on the nine themes' evolution characteristics and user acceptance, ensuring both universality and uniqueness.

7 Limitations and Future Directions

This study uses quantitative methods to comprehensively analyze the thematic categorization and characteristics of reading promotion content from 31 provincial public libraries' WeChat platforms, providing more intuitive and accurate results than experiential analysis. However, limitations exist: (1) Data collection was limited to nine months due to difficulties, representing a short cycle; (2) The analytical perspective is relatively single, constrained by the LDA model; (3) Only overall data were analyzed without individual library-level analysis due to space limitations; (4) Machine learning-based document classification has inherent limitations and cannot achieve perfect accuracy.

Future research can: (1) Track multi-year data (e.g., five years) for longitudinal analysis; (2) Combine promotion effectiveness analysis with multi-dimensional, comprehensive perspectives for richer conclusions; (3) Analyze individual libraries first before conducting comparative analyses.

References

[1] Zhou Yumeng. Tencent Releases Q3 2021 Financial Report [N]. Shenzhen Special Zone Daily, 2021-11-11 (A04). [2] Li Wu, Wang Dan, Huang Danyu, et al. Ten Years of Progress in Library Reading Promotion Research (2005-2015) [J]. Library Forum, 2016(12): 54-65. [3] Zhang Huaitao. The Concept and Implementation of Reading Promotion [J]. Henan Library Science, 2015(1): 2-5. [4] Wang Bo. The Definition of Reading Promotion and Library Reading Promotion—Also on How to Understand and Learn from Fashionable Library Reading Promotion Cases [J]. Library Forum, 2015(10): 1-10. [5] Wan Muchen, Ou Liang. Empirical Research on Reading Promotion Effectiveness of University Libraries Based on WeChat Public Platforms [J]. Library and Information Service, 2015, 59(22): 72-78. [6] Cai Liping, Kong Dechao. Research on WeChat Reading Promotion of Provincial Public Libraries Based on WCI [J]. Library Work and Research, 2016(10): 90-95. [7] Wang Baoying, Wang Baoling, Jin Yueqing. Research on University Library WeChat Public Platforms from the Perspective of Static Reading Promotion Resources [J]. Library Forum, 2018, 38(4): 111-118. [8] Chinese Government Website. Administrative Divisions of the People's Republic of China [EB/OL]. [2021-10-01]. http://www.gov.cn/guoqing/2005-09/13/content_{5043917}. [9] Liang Chenchen, Li Renjie. Analysis of Lijiang Ancient Town Image Perception Integrating LDA and Feature Dimensions [J]. Geographical Science Progress, 2020(4): 614-626. [10] Wang Yaru. Text Mining Based on Yunnan Tourism Online Website Complaint Data [D]. Kunming: Yunnan University, 2019. [11] Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003(3): 993-1022. [12] Qiu Junping, Shen Chao. Analysis of Hot Topics in Domestic Big Data Research Based on LDA Model [J]. Modern Information, 2021, 41(9): 22-31. [13] CSDN Community. Personal Understanding of LDA Perplexity [EB/OL]. [2021-12-06]. https://www.csdn.net/tags/NtTaYgzsNzY0NjQtYmxvZzY0O0O0O0O0.html. [14] Zhejiang Library. Zhejiang Library Koi 2.0 Fully Upgraded! Massive Books, Kindle Reading Device, Travel Packages… Take Poetry on Your Journey [EB/OL]. [2021-10-01]. https://mp.weixin.qq.com/s/gLcE6qFa-gdGkpXhYuP-. [15] Dasheng's Little Universe. If Not for an Accident, the Top Poet of the Tang Dynasty Might Not Have Been Li Bai [EB/OL]. [2021-10-01]. https://mp.weixin.qq.com/s/IGTFVpBaXqaA99Usenn7w. [16] Yifei. China's First Whale Fall Discovered in the South China Sea: One Whale Falls, All Things Live! This is the World's Most Romantic Death… [EB/OL]. [2021-10-01]. https://mp.weixin.qq.com/s/Q66gm4Vraag9O15EQOYZWg. [17] A Great Talent Created a Rumor That Deceived Us for Hundreds of Years [EB/OL]. [2021-10-01]. https://mp.weixin.qq.com/s/vzH01LzpEnF4bHL51jkBQ. [18] Tibet Autonomous Region Library. "Four Lectures, Four Loves" Stimulates Tibet's Poverty Alleviation Momentum [EB/OL]. [2021-10-01]. https://mp.weixin.qq.com/s/6exLFVvJadMJqC16qKOJtg. [19] Jilin Provincial Library. One Recommendation, Heartfelt | Understanding Classical Poetry [EB/OL]. [2021-10-01]. https://mp.weixin.qq.com/s/vcBF-

tL17GiK8uJOUVWiRw. [20] Fan Bingsi. Reading Promotion and Library Science: Analysis of Basic Theoretical Issues [J]. Journal of Library Science in China, 2014, 40(5): 4-13. [21] Cui Yongxin. Research on Content Marketing Strategies for First-Class University Library WeChat Reading Promotion [D]. Tianjin: Tianjin University of Technology, 2020.

Author Contributions: Li Qian: Topic selection, overall structure determination, paper revision Wang Shuai: Literature research and organization, paper conception, drafting

---

A Study on the Topic of WeChat Platform's Reading Promotion Content in Public Libraries by Using LDA Model in China

Li Qian[1],[2], Wang Shuai[2] [1]School of Economics and Management, Shanxi University, Taiyuan 030006 [2]Institute of Scientific and Technical Information of China, Beijing 100038

Abstract: [Purpose/Significance] This paper aims to identify the content themes and evolutionary characteristics of reading promotion conducted by domestic public libraries via WeChat public platforms. [Method/Process] We collected tweets from 31 provincial public libraries' WeChat accounts, defined and selected reading promotion content, and used the LDA topic model to identify themes, analyze evolution, and examine reading volumes. [Result/Conclusion] Reading promotion content on WeChat by public libraries can be divided into nine categories with broad coverage. Each theme shows different evolutionary trends over time, and reading volumes vary across themes, reflecting different popularity levels. Finally, suggestions are provided for future WeChat reading promotion work from perspectives including scientific planning of content 配比, reversal of content subordination, special population content settings, innovative design of content and forms, and creation of branded promotion content.

Keywords: LDA model; WeChat public platform; reading promotion; theme mining

Note: Figure translations are in progress. See original paper for figures.