

Article

Sports Information Needs in Chinese Online Q&A Community: Topic Mining Based on BERT

Chuanlin Ning ¹, Jian Xu ¹, Hao Gao ², Xi Yang ³ and Tianyi Wang ^{1,*}¹ School of Media and Communication, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Minhang District, Shanghai 200240, China; ningchuanlin@sjtu.edu.cn (C.N.); xujian@sjtu.edu.cn (J.X.)² School of Journalism and Communication, Nanjing Normal University, No. 122, Ninghai Road, Gulou District, Nanjing 210097, China; 42396@njnu.edu.cn³ School of Humanities, University of Southampton, Hartley Library B12, University Road, Highfield, Southampton SO17 1BJ, UK; xy6n20@soton.ac.uk

* Correspondence: tyty0410@126.com

Abstract: The online Question and Answering (Q&A) community has grown globally, allowing users to ask, discuss, and answer questions based on shared interests. As a gathering place for people's knowledge production, collaboration, and dissemination in the current Internet scene, the online Q&A community can intuitively reflect the public's information needs and behavior. It also collects many sports-related data and becomes an effective vehicle for comprehending mass sports information needs and disseminating sports knowledge. However, sports-related studies on the online Q&A community have rarely been reported. This study took the sports information in Zhihu, the largest Q&A community in China, as the research object to explore the public needs for sports information in China. We introduced the BERT model through a self-compiled python program and collected 391,092 sports-topic answers in the online Q&A community of Zhihu. Then, we explored the topic content, evolution trend, and user attributes of these answers. We found that the overall trend of sports information needs in Zhihu can be divided into three cycles: the London 2012 Olympic period, the Rio 2016 Olympic period, and the Tokyo 2020 Olympic period in general. The diversified content of information needs included 40 second-level themes and eight first-level themes. Male and female users had similarities and differences in sports information needs. The male and female users had the same information needs for fitness-related information. However, men were more concerned with confrontational solid sports such as basketball and football; women were more likely to care about weight loss, shape effect, and self-protection while doing sports activities. In addition, compared with men, women preferred to emphasize their gender attributes when expressing their needs for sports information to obtain more practical knowledge. In conclusion, our finding reveals that the sports community formed by the current online Q&A community in China is still a male-dominated information field.

Keywords: information needs; sports communication; sports in China; online Q&A community; BERT model; topic mining



Citation: Ning, C.; Xu, J.; Gao, H.; Yang, X.; Wang, T. Sports Information Needs in Chinese Online Q&A Community: Topic Mining Based on BERT. *Appl. Sci.* **2022**, *12*, 4784. <https://doi.org/10.3390/app12094784>

Academic Editor: Federico Divina

Received: 22 March 2022

Accepted: 7 May 2022

Published: 9 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The acquisition, sharing, and dissemination of social knowledge have changed dramatically. Online Question and Answering (Q&A) communities that support users to ask, discuss and answer questions based on shared interests have become more and more popular in recent years [1,2]. With the primary function of attracting and promoting users to ask questions online and obtain answers from other community members, online Q&A communities construct a knowledge community around different topics and play an essential role in meeting people's information needs and disseminating social knowledge [3]. Quora, the world's most famous online Q&A community with about 300 million users in 2018, claims it is a place to share knowledge and better understand the world [4]. In

China, Zhihu, launched in 2011, has become the most influential online Q&A community platform. By January 2019, the users of Zhihu exceeded 220 million, accumulated more than 130 million answers, and Zhihu had established a community-driven business model [5]. Online Q&A communities such as Zhihu have changed the traditional mode of people's knowledge exchange, and their knowledge production, information interaction, user characteristics and business models have gradually received widespread attention in the academic field [6–9].

As a platform for disseminating knowledge in the new media era, online Q&A communities can be seen as virtual knowledge aggregators that provide users with the flexibility to ask open-ended questions to a broad audience, the answers to which may be of great help to both the user and the community on specific topics such as health, sports or finance [10–12]. By 16 August 2021, the number of people following the sports topic in Zhihu had exceeded 10 million. Therefore, Zhihu has become a good channel for understanding Chinese public sports information needs. Exploring the sports information needs in online Q&A communities could deepen the understanding of sports ecology and development in online communities from users' perspectives. At the same time, it is of great significance to popularize scientific sports information knowledge and enhance public sports information literacy.

An essential goal of data analysis is to determine the shared characteristics of data points [13]. It usually means determining which events or concepts are discussed in the document in text analysis. As a popular statistical tool, topic modeling is well suited for use with text data to extract latent variables from large datasets [14]. Employing BERT, the pre-trained language model, the core of our research is an analysis of all the questions asked by users about sports in Zhihu. Specifically, the main objective of our study is: on the one hand, to investigate the sports information needs of people in the online Q&A community platform of Zhihu in China. For example, the general profile of users' questions, such as temporal trends, user characteristics and other features. On the other hand, to explore the topic classification of sports information needs in Zhihu and the time and user characteristics of these topics through topic mining of these information needs.

There are three main contributions of the paper. First, it innovatively explores the sports information needs of the public in online Q&A communities through big data mining and analysis, which provides a valuable perspective on our overall understanding of people's sports information interactions within the social media era. To the best of our knowledge, only a few papers have mined and analyzed the topics of information needs in online Q&A communities, especially the identification of topics of health information needs [15,16]. However, not a single study elaborated on the characteristics of sports information needs in the Q&A community. For a decade, this study examined the Chinese public's sports information needs in online Q&A communities. Second, the study of the user attributes and topics supports not only the understanding of the Chinese public's sports participation but also the recognition of gender-specific and user-specific preferences for sports topics and thus the understanding of the power relations between male dominance and female subordination in the online fields from a sociological perspective of sports. Finally, given that the generation of these sports topics objectively reflects the level of concern of the Chinese public about sports issues and the demands for sports information, this study would also provide useful guidelines for sports administrations and commercial companies to optimize public sports policies or business solutions.

2. Literature Review

As a social media platform focusing on users' knowledge exchange, the online Q&A community accumulates tens of thousands of questions that people ask or answer every day. These questions form a diverse whole, supporting users searching for information. With the in-depth development of the Internet and new media technology, the online Q&A community has gradually shifted from search-engine-based interactive information service platforms such as Baidu Know and Answers to user-centered social network Q&A

communities such as Zhihu and Quora. The user-centered communities tend to focus more on social interaction, with a well-established social network and feedback mechanism [17]. Therefore, it has also triggered the focus of researchers, and the online Q&A communities discussed below are such socialized communities.

Previous research on the social network Q&A community focuses on studies of users and their behavior and studies of the questions and answers of the communities. There are two main approaches of user-platform and platform-user in the user and behavior studies. The user-platform approach focused on exploring the motivation of user participation and platform content production by quality users. Bao et al. studied the drivers of user participation in online Q&A communities based on social cognitive theory [18]. They found that outcome-based expectations were positively correlated with user participation and that users' self-efficacy could positively influence their participation behavior. In the research approach of platform-user, Wang et al. [19] proposed an improved method for identifying key users of online Q&A communities from knowledge dissemination. Guo et al. [20] found that platform anonymity has an essential impact on user participation and content production and suggested that user anonymity be viewed in two ways for online Q&A community platforms. In the studies of the questions and answers of the communities, both commenting on others and receiving comments were significant motivating factors in users' continued use of online question and answer communities [21]. Shi et al. [22] collected answer data through a crawler, established three evaluation dimensions of textual, rhetorical, and emotional content, and then identified nine features that might affect the quality of answer content in Zhihu.

Information need is the basis for information behavior, which Dervin defines as an urge to understand the current situation when an individual is faced with a problem or concern or when there is a need to understand or make a choice [23]. The studies of user information behavior have found that a person only recognizes their knowledge needs before the corresponding information behavior occurs [24], such as information search, screening, and avoidance. Therefore, information need as a motivating factor for information behavior generation has also received academic attention [25]. Users' information needs are evolving and limited by time and space [26]. In the Internet era, social media has gradually become a diversified place for people to solve their information needs as an information field. As a result, scholars in different fields have conducted relatively affluent research on this subject. For example, Xing et al. [27] analyzed the type, content, and motivation of users' information needs through library microblogging data, which provided informative suggestions for libraries to understand users' needs and serve them better. Jia et al. [28] used a national survey during the COVID-19 outbreak to find that information needs influenced media use and media trust through the information matching mechanism, which helped people better understand the relationship among public information needs, media use, and media trust during emergencies. Some scholars have also explored the user needs of online Q&A communities. Wang et al. [29] analyzed the answers to weight loss topics in online Q&A communities to reflect the characteristics and gender differences of the current public needs for weight loss information. Huang et al. [30] studied the process of topic identification and analysis in online Q&A communities by improving the technical method and analyzed the method's effectiveness by using the topic of the elderly as an example.

Topic modeling has been widely used in studying topics related to Q&A communities. On the one hand, studies extracted hot topics of the Q&A community through topic modeling to detect the trending topics of the Q&A community and provide references for the recommendation of questions and answers from the Q&A community [31]. On the other hand, studies used topic modeling to detect the trending topics in a particular field in the Q&A community to understand users' focus and information needs, mainly focusing on the areas of health information communication [32–34], science communication [35], data science [36], product user preference [37], tourism [38] and library user services [39]. Junghwa Bahng et al. [32] detected the topics of hearing loss on the Naver Knowledge-iN

through topic modeling to identify patients' perceptions, concerns, and needs regarding hearing loss. Zhang et al. [38] extracted the topics of tourism information in Zhihu through topic modeling to explore the tourism information needs of the users in the context of COVID-19 [38].

In terms of the use of topic mining models, previous research on Q&A communities applied the models including LDA, STM, BERT, and so on [32–39]. Zhao et al. [34] applied LDA on health topic information in Zhihu to explore internet users' needs for health information. Jiang et al. [35] used the STM model to extract the topics about climate change in Quora to investigate public opinion on science communication. Luo et al. [40] applied two BERT-based models on the Q&A site to detect pregnancy-related topics and concluded that the BERT-based models were better than the traditional models. In terms of the data processing, it was mainly divided into three main steps: First was data collection, which generally crawls data from Q&A communities. The second was data preprocessing, including Text Cleansing, Word Tokenization, POS Filtering, Stop-word Filtering and Word Stemming. The third was Topic modeling, including topic clustering and topic validation [32–40].

From the perspective of the Q&A communities studied, Quora, Zhihu, Reddit and Naver Knowledge-In were mainly concentrated, with multiple languages such as Chinese, English and Korean. Karbasian et al. [36] applied the LDA model to extract data science topics in two English Q&A communities called Stack Exchange and Reddit to provide a path for detecting the trending topics of data science research. Han et al. [41] used the KoBERT model based on Korean to detect the topics of the course teaching evaluation on a university online Q&A site. Qian et al. [42] used the Chinese BERT topic model released by Google to extract the health-related topics in Chinese Q&A sites to understand the health information needs of the Chinese elderly. The research mentioned above shows that the topic modeling tools, including BERT, are effective in multi-language environments and available for topic modeling in multi-language texts.

3. Methods and Data

The language model pre-training brings a breakthrough to natural language processing (NLP) technology. Latent Dirichlet Allocation (LDA) [43] is a popular and major model in textual topic mining, whose main feature is that all documents in a collection may contain the same set of topics. Still, each document contains a different number of topics. In this iterative process, documents are observed one after another, while the hidden structure-available topics, topic distribution per document and topic assignment per word-remains. Despite the popularity of this approach, it has many limitations, such as the need to identify a certain number of topics before modeling; the high possibility of generating irrelevant topics; the fact that the identified topics are static and do not change over time; and the fact that semantic relevance is lost because the algorithm uses a bag-of-words model, which also faces serious performance problems for small text data [44].

The transformer model based on the self-attention mechanism is the foundation of the language model pre-training. GPT, BERT, XLNet, and other large-scale language model pre-training are stacked and optimized on the basis of the transformer model, which relies on powerful arithmetic to obtain a general language model and representation based on easily accessible, non-manual data, and then fine-tunes the pre-trained model with the task corpus on the target NLP task to rapidly converge to improve the accuracy in various downstream NLP tasks. Therefore, pre-trained language models have been rapidly developed and widely used since their inception and have become the core technology for various NLP tasks. The effectiveness of pre-trained language models in various NLP projects is evident. As the parameter size increases and the training data increases, pre-trained language models can improve accuracy and generalization [44,45].

3.1. BERT Model

The BERT (Bidirectional Encoder Representations from Transformers) model is a natural language model pre-trained on a large-scale corpus based on pre-training-fine-tuning, which Google AI proposed in October 2018. It is an actual recent research result in NLP, as it has significantly improved accuracy in several natural language processing tasks [46] and provides a good feature representation for word learning. BERT, BERT-like, and fusion-based models outperform traditional machine learning and deep learning models. BERT models include two training tasks: a masked language model (MLM) and a Next Sentence Prediction (NSP). MLM is a good solution to the problem of inverse order information leakage in bidirectional modeling, while NSP is good for understanding the relationship between two texts, which is suitable for reading comprehension or textual entailment tasks [43,47]. Many studies have shown that BERT can have good results in topic mining of texts in Chinese, English, Russian, Arabic and other languages. This paper adopted the BERT model for topic mining [36,42,48,49].

3.2. Data Sources

This study crawled the total raw data of 391,092 questions (as of 2 June 2021) under the sports topic of Zhihu by a self-coded python program. The data included the question, question time, questioner gender, questioner authentication status, and whether the question was anonymous, forming a question collection. The coding is performed by the BERT model based on preprocessing.

3.3. Data Processing Process

The Sentence_transformer module was used to load the pre-trained BERT cross-linguistic model [50,51] and encode the preprocessed sentences into a 512-dimensional vector representing the sentences. The dimensionality of the feature vector should be reduced before clustering to avoid the effect of the curse of dimensionality. UMAP constructs a high-dimensional representation of the data and optimizes the layout of the data in the low-dimensional graph. Thus, it is an effective means of dimensionality reduction. Compared to t-SNE, the UMAP method is more efficient in processing [52]. Using UMAP, the 512-dimensional feature vector was dimensioned down to 20 dimensions. K-means is one of the most used clustering algorithms based on Euclidean distance, which considers that the closer the distance between two targets, the greater the similarity. It was used for clustering in this study, and the results are visualized in Figure 1.

After clustering, texts were classified into different topics. This study used the TF-IDF algorithm to derive the importance scores of the words in each topic and thus determine each category. The TF-IDF algorithm, a common textual keyword mining method, considers that if a word appears more frequently in a topic and less frequently in other topics, it has a more significant influence on the core content of that topic [53]. TF-IDF involves two components, Term Frequency (TF) and Inversely Document Frequency (IDF), and is calculated as follows:

$$TF - IDF = TF_{i,j} * IDF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} * \log\left(\frac{N}{d_i + 1}\right)$$

where $n_{i,j}$ is the number of occurrences of the word i in document j . $\sum_k n_{k,j}$ represents the total number of occurrences of all words in document j . N is the total number of corpus documents, d_i is the number of documents containing the word i in the corpus, and d_i is calculated as " $d_i + 1$ " to prevent zero in the denominator. $TF_{i,j}$ is the normalization result of word frequency i in document j . $IDF_{i,j}$ is a measure of the word i 's ability to distinguish document j , and it is also an adjustment of $TF_{i,j}$ weight to suppress words with high frequency in documents such as "的(of)," "和(and)" [54].

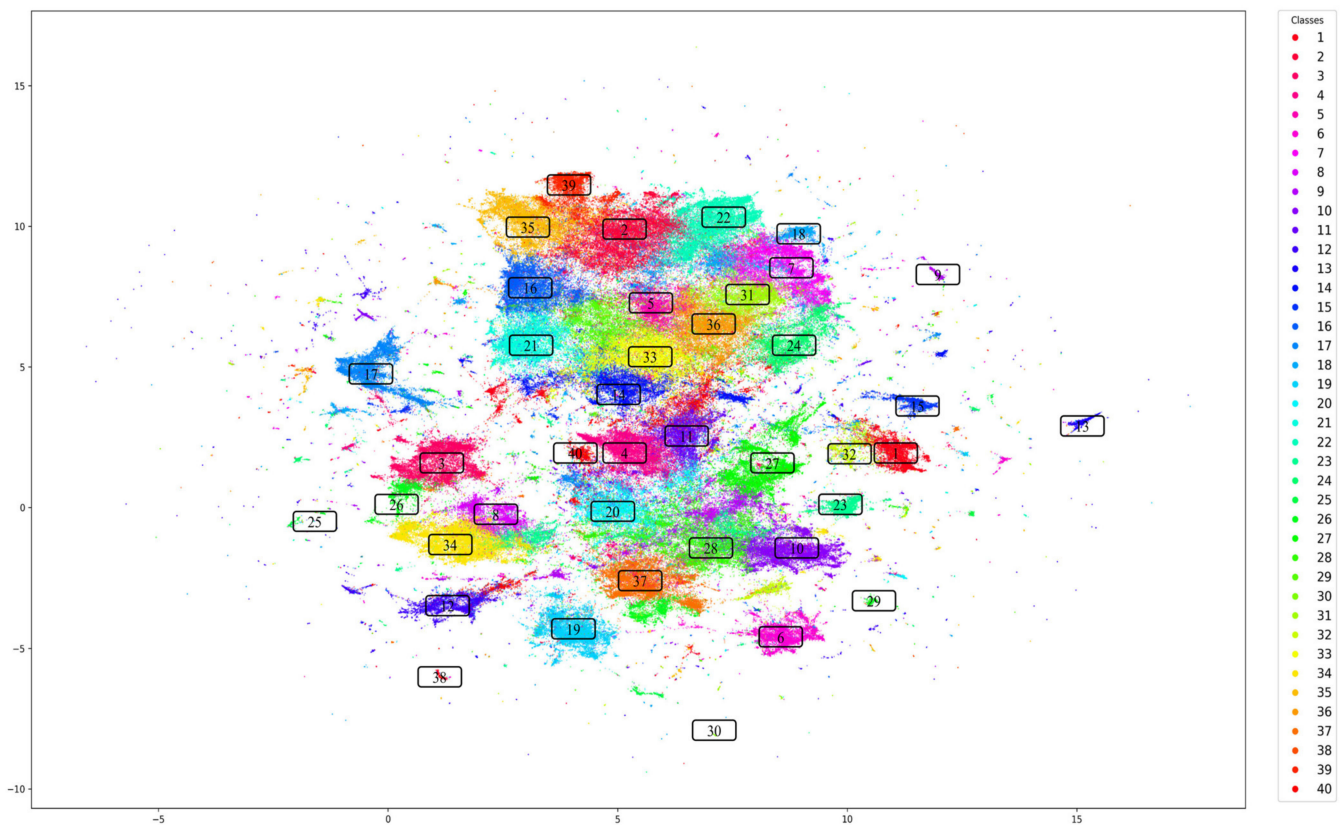


Figure 1. Results of topic clustering.

The overall research idea is shown in Figure 2, in which a rhombus represents a process, and a box represents an outcome.

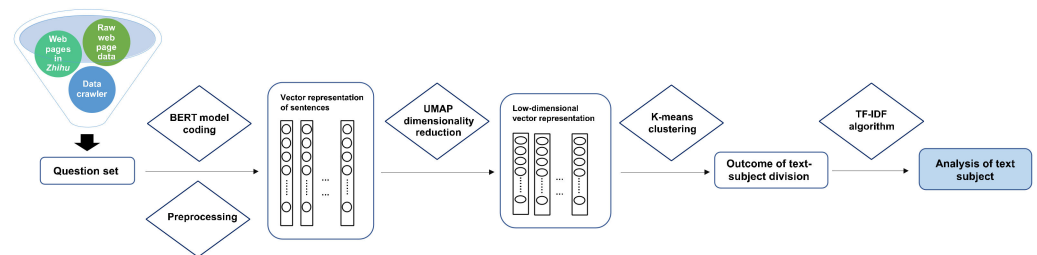


Figure 2. Technology roadmap in this research.

4. Results and Analysis

A total of 391,092 questions have been asked under the ‘Sports’ topic since 2010. Of these, 3020 questions were asked by authenticated users, 279,590 by non-authenticated users, and 1457 by users who have logged out of their Zhihu accounts. Of the total, 46,909 questions were asked by female users, 133,981 by male users, and 210,202 by users whose gender is unknown (no gender set, anonymous, or logged out).

Overall, most of the sports information needs in Zhihu are raised by non-certified users, accounting for 71.50%, while certified professional users account for only 0.77%. Among the information needs of the known gender of the questioners, 74.07% were male, reflecting the popular trend of sports information needs and the male-dominated information characteristics in online Q&A communities, which further confirms the related study by Vasilescu et al. [55], who found that men use online Q&A communities more frequently than women, posting both questions and answers more than women. In addition, users of online virtual Q&A communities have a high level of personal information concealment.

4.1. Time Trend

In terms of time (excluding the six data from 2010 when Zhihu was not yet online; the 2021 data are as of 2 June of that year), the highest number of sports information requests in Zhihu was in 2016, with 93,959 questions, followed by 2017 and 2015, with 84,848 and 69,382 questions, respectively. The lowest number of questions was in 2011, with 1410, followed by 2012 and 2013, with 5569 and 8612 questions. After 2013, the number of questions asked each year was above 10,000, indicating that the development of Zhihu as an online Q&A community has continued to mature (Figure 3).

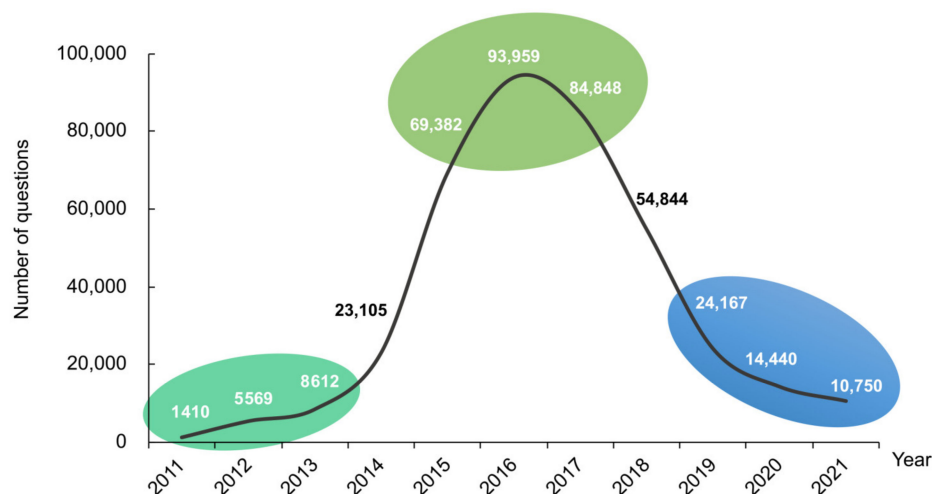


Figure 3. The trend of sports information needs on Zhihu by year.

The results show that the overall trend of sports information needs in Zhihu can be divided into three cycles: the London 2012 Olympic period, the Rio 2016 Olympic period, and the Tokyo 2020 Olympic period. Between 2011 and 2021, the number of sports information demanded in Zhihu shows the distribution with increases then decreases. The Rio Olympics period was the highest number of sports information required, exceeding the others.

4.2. Contents of Information Needs

According to the results of preliminary data processing, the sports information needs of Zhihu users were divided into 40 different categories of topics. However, the topic model can only show the keywords and their contribution to each category of topics and cannot automatically generate the name of each topic. Referring to the way previous studies determined topic names, these 40 topic categories were named manually by reading the keywords with important contribution degrees. In contrast, this study randomly selected 100 original questions in each category of topics to be read manually to confirm the topics and name them more accurately. For the case of topic number 38, which had only four available subject terms, the questions were found to be highly consistent with the inquiry form “What were the highlights of the NBA regular-season game on X?” Therefore, the relevant subject line can summarize the topic. We have grouped the 40 topics to have a more concentrated theme. The 40 topic categories of information needs can be combined into eight primary categories (Table 1). In descending order of percentages, the primary topics are sports skills, sports events, sports shaping, weight loss, professional athletes and teams, Chinese sports and physical education, sports health, sports equipment, and sports experience. From the 40 secondary topic categories, the most demanded information in Zhihu is sports and slimming, followed by NBA player performance, European soccer league, and fitness consultation.

Table 1. Classification of topics and content of sports information needs in Zhihu.

Primary Subject Category	Secondary Subject Category	Subject Heading (Partial)	Ratio	Topic Number
Sports skills (25.65%)	Fitness consultation	Fitness, Gym, Coach, Plan, Personal Training, Exercise, Great God, Equipment	3.53%	10
	Middle-distance running and Marathon performance improvement	Running, Marathon, Improve, Performance, Kilometers, Treadmill, Long Distance Running, Training, Speed, Minutes	3.48%	27
	Personal football training and planning	Soccer, Soccer games, Soccer field, Soccer playing, Like, Professional, Become, Goal	3.40%	33
	Learning in wrestle and Wushu	Boxing, Fighting, Taekwondo, Combat, Sparring, Learning, Practical, Wrestling, Muay Thai, Karate	3.13%	19
	Bicycle and hiking	Bike, Riding, Road Mountain Bike, Needs, Routes, Outdoor, Experience, Cycling, A bike	3.13%	17
	Basketball skills and tactics	Basketball, Play basketball, Shooting, Basketball game, Defense training, Foul dunk, Improve, Boys	2.82%	21
	E-sports game skills	League of Legends, King's Glory, Dota, Lol, Game, Players, Assist, Mid-single, Operation	2.42%	12
	Swimming learning	Swimming, Swimming pool, Breaststroke, Freestyle, Need, Myopia, Experience, Ask	2.25%	6
	Snow and ice sports	Skateboarding, Skiing, Longboarding, Snowboarding, Beginners, Recommend, Skating, Compare, Beginners	0.91%	15
	Yoga study	Yoga Practice, Training, Coach, Learning, Recommend, Moves, Asana, Ask, Pilates	0.58%	13
Sports events (19.78%)	UEFA	Evaluation, Real Madrid, Barcelona, Manchester, United Champions League, Season, Bayern, Chelsea, Liverpool, Premier League	4.09%	22
	Chinese Super League and Korea-Japan World Cup	Evaluation, Looking at, Evergrande, China, Match, Final, Korea, Chinese Super League, Winning, World Cup	3.29%	7
	Chinese Football and World Cup	World Cup, China, Football, Fans, Match, Europe, Country, Level, National Team, National Football Team	3.01%	36
	NBA events	NBA, Teams, History, Basketball, Season, Games, Current, Playoffs, Level, USA	2.94%	16
	E-sports matches	E-sports, games, Sports, Matches, Live, China, Professional, Lol, Tournaments	2.18%	40
	Major basketball and football matches	Soccer, NBA, Players, Athletes, Games, Like, China, CBA, Soccer games, Live	2.12%	29
	World Cup performance of European and American teams	World cup, Evaluation, Brazil, Germany, Argentina, Watch, France, Spain, Finals, National teams	2.10%	31
	NBA Regular Season highlights	Highlights, Regular, Season, NBA, Opener	0.03%	38
	NBA highlights and roasts	Highlights, NBA, Regular season, All-star game, Big game, Tweet, Season Dunk, Tips, Three points	0.02%	30

Table 1. Cont.

Primary Subject Category	Secondary Subject Category	Subject Heading (Partial)	Ratio	Topic Number
Sports shaping and weight loss (14.67%)	Exercise to lose weight	Loss weight, Weight, Loss fat, Fitness, Girls, Gym, Body fat, Plan, Fat, Exercise Muscle, Exercise, Pectoral Muscle,	4.44%	34
	Fitness and body training	Push-ups arms, Shoulders, Strength, Back, Chest	3.28%	37
	Grow taller through exercise	Fitness, Workout, Bodybuilding, Training, Growth, Stick, Women, Men, Movements, Effect	3.27%	28
	Running to lose weight	Running, Weight loss, Aerobic, Exercise, Bodyweight, Daily, Fat loss, Consumption, Effect, Jogging	1.94%	8
	Exercise for body shaping	Abs, Belly, Vest, Calves, Thighs, Exercise, Girls, Train, Lean legs, Buttocks	1.75%	26
Professional athletes and teams (14.07%)	NBA Player Performance	Kobe, Curry, Players, Looked at, Peak, Career, Yao Ming, Harden, Lin Shuhao, Ability	4.41%	2
	NBA Teams	Warriors, Cavaliers, Evaluation, Season, Spurs, Rockets, Thunder, Lakers, Watch, Celtics	3.48%	35
	Famous Chinese athletes	Watch, Evaluate, Sun Yang, Players, Liu Xiang, Performance, Competition, Events, Ning Zetao, Lin Dan	2.77%	18
	NBA Star comments	James, Jordan, Durant, Paul, Kobe, Evaluation, Wade, History, Duncan, Status	1.96%	39
	Soccer teams and players	Players, Football, History, Teams, Sports stars, Players, League, Clubs, Top	1.44%	5
Sports and physical education in China (8.28%)	Physical education and examination	Sports, University, Professional, School, Results, Culture, Graduate exams, Students, Education, Postgraduate	3.20%	4
	Olympics	Olympic Games, China, Seeing, Winter Olympics, Chinese Women's Volleyball, Evaluation, Gold Medal, Country, Beijing	2.62%	24
	Development of China's sports industry	Sports, Athletes, Development, Projects, Sports games, Companies, China, Professional, Sports industry	2.46%	11
Sports health (7.66%)	Sports protection and rehabilitation	Knee, Running, Calf, Muscle, Injury, Correction, Recovery, Meniscus, Pain, Soreness	3.40%	3
	Fitness diet choices	Protein Powder, Fitness, Protein, Diet, Fat Loss, Food, Intake, Calories, Eggs, Supplement	2.30%	23
	Running for health	Running, Exercise, Sweating, Heart Rate, Breathing, Night, Winter, Cardio, Skin, Sleep	1.96%	9
Sports equipment (7.08%)	Ball sports equipment selection	Badminton, Tennis, Table Tennis, Golf, Playing, Rackets, Recommend, Ask, Tennis Rackets	3.07%	14
	Sports shoes recommendation	Running Shoes, Basketball Shoes, Recommend, Shoes, Sneakers, Nike, Pair, Suitable, Brand	2.21%	1
	Sports apparel and equipment	Sports, Brand, Clothes, Underwear, Recommend, Pants, Watch, Ask, Bracelet, Taobao	1.37%	32

Table 1. Cont.

Primary Subject Category	Secondary Subject Category	Subject Heading (Partial)	Ratio	Topic Number
Sports experience (2.81%)	Sports music	Headphones, Music, Running, Songs, Recommend, Suitable, Theme Song, World Cup, Listen, Song, Name	0.44%	25
	Sports fun and experience sharing	Sports, Experience, Meditation, Life, Like, Things Stick, Feel, Work	2.81%	20

4.3. Characteristics of Information Needs

4.3.1. The Changing Trend of Information Needs

Regarding the evolution of the temporal trend of information needs on primary topics, the overall trend of information needs on primary topics and the needs of sports information in Zhihu is generally comparable, with slight differences among topics.

The information needs on sports skills first increase and decrease, reaching a peak in 2017 and showing an increasing trend in 2020. The information needs on sports events and sports and physical education in China had two significant peaks, the former in 2016 and 2018 and the latter in 2016 and 2019. The information needs on sports shaping and weight loss, professional athletes and teams, sports health, sports equipment, and sports experience showed a trend of increase first and then decreased, with the peak in 2016 or 2017 (Figure 4).

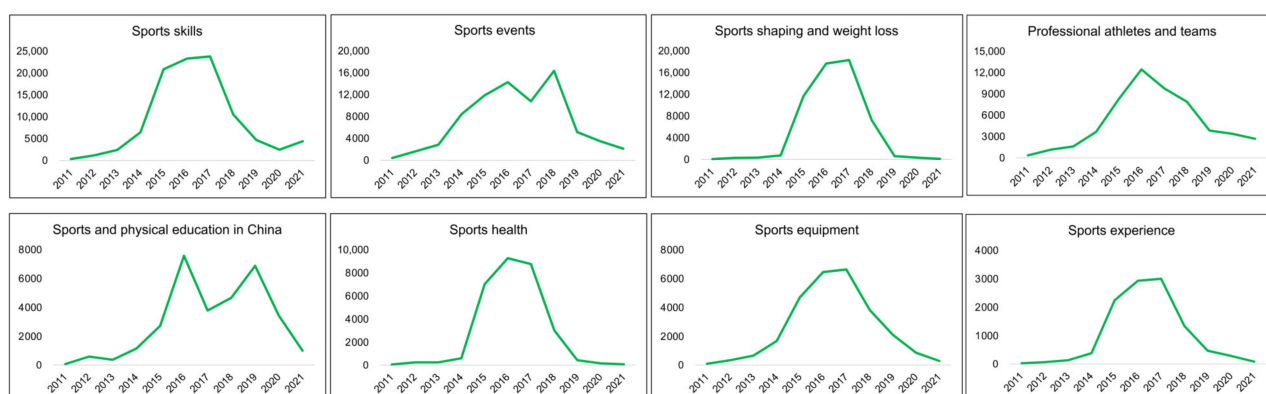


Figure 4. Trends of different primary topics by year.

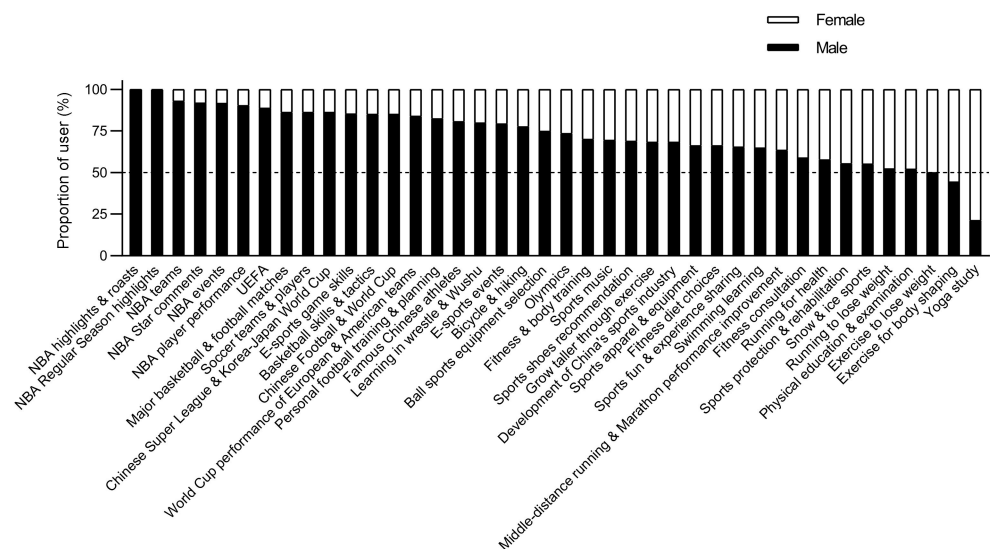
4.3.2. Sports Information Needs of Different Gender Users

Analysis of the user attributes of the topics, we found that men are most concerned about sports events regarding the eight primary topics, accounting for 27.71%. Women are most concerned about sports shaping and weight loss, accounting for 21.42%, about twice the proportion of men. Both men and women ranked second regarding information needs about sports skills. Men ranked third in terms of information needs about professional athletes and teams, nearly three times the proportion of women. Regarding sports health information, women's needs are nearly twice as high as men's (Table 2). Stice et al. [56] found that women tend to have more negative evaluations of body size and appearance than men, and the resulting pressure to lose weight makes women feel more strongly negative about their bodies.

Table 2. Gender-specific sports information needs of primary subject categories.

Ranking	Primary Subject Category of Male	Ratio	Primary Subject Category of Female	Ratio
1	Sports events	27.71%	Sports shaping and weight loss	21.42%
2	Sports skills	22.77%	Sports skills	21.25%
3	Professional athletes and teams	17.43%	Sports events	18.28%
4	Sports shaping and weight loss	10.68%	Sports health	12.20%
5	Sports equipment	6.46%	Sports and physical education in China	9.65%
6	Sports and physical education in China	6.42%	Sports equipment	7.50%
7	Sports health	6.16%	Professional athletes and teams	6.13%
8	Sports experience	2.37%	Sports experience	3.58%

Given the relatively wide range of information demanded by each primary topic, to further explore the similarities and differences of information required by different genders in online Q&A communities, the information requested by different genders for secondary topics was analyzed. Figure 5 shows the proportion of male and female information needs in each secondary topic. Thirty-seven out of forty topics have more male than female information needs. In the topics of NBA highlights and roasts and NBA regular-season highlights, the gender of information seekers is all male.

**Figure 5.** User proportion in the gender of secondary topic information.

The most significant gap between men and women is basketball and soccer sports. The number of questions on exercise, body part shaping, and yoga learning is higher for women than men. The number of female questions on yoga learning is more than 50% higher than that of male questions.

Table 3 shows the proportion of male and female users in the secondary topic information needs categories and reveals more specific differences in sports information needs between men and women. The top 10 sports information needs topics are different. For male users, NBA player performance, European soccer league, and NBA team are the topics with the highest information needs. Sports slimming, sports protection and rehabilitation, and middle and long-distance running and marathon performance improvement are the most popular topics for women. In general, men are more concerned about sports with

intense confrontation, such as soccer, basketball, and fighting. At the same time, women are more concerned about sports that shape the body and sports with less confrontation, such as running, fitness, swimming, and self-protection during sports.

Table 3. Sports information needs by gender of secondary subject categories (Top 10).

Ranking	Secondary Subject Category of Male	Ratio	Secondary Subject Category of Female	Ratio
1	NBA player performance	5.53%	Exercise to lose weight	7.30%
2	UEFA	5.17%	Sports protection and rehabilitation	5.86%
3	NBA teams	5.13%	Middle-distance running and Marathon performance improvement	5.21%
4	Personal football training and planning	4.28%	Fitness consultation	4.93%
5	Chinese Football and World Cup	4.07%	Physical education and examination	3.99%
6	Bicycle and hiking	3.81%	Grow taller through exercise	3.84%
7	NBA events	3.75%	Fitness and body training	3.72%
8	Combat and Wushu learning	3.67%	Swimming learning	3.67%
9	Basketball skills and tactics	3.37%	Sports fun and experience sharing	3.58%
10	Chinese Super League and Korea-Japan World Cup	3.27%	Running to lose weight	3.51%

The keyword word frequency analysis of the information needs of male and female users found that the keywords appearing more frequently in the information needs of male users are basketball and soccer, which indicates that men are more concerned about basketball and soccer sports. Those appearing more frequently in female users are fitness and weight loss, reflecting women pay more attention to the weight loss effect of fitness or sports. The keyword that appears more often in both male and female users is fitness, which indicates that both women and men have a greater demand for fitness information. Compared with men, women are more likely to emphasize their gender attributes, such as girls, when asking questions to obtain more practical information.

4.3.3. Characteristics of Sports Information Needs of Users with Different Authentication Attributes

From the perspective of different authentication attributes (Authenticated, non-authenticated, and anonymous users), sports events and sports skills are the essential information content for all three types of users. For authenticated users, sports events, professional athletes and teams, and Chinese sports and sports education all account for higher demand than non-authenticated and anonymous users. At the same time, sports experience is the only topic that is more popular among anonymous users than authenticated and non-authenticated users (Table 4).

Table 4. Sports information needs with different authentication attributes of primary subject categories.

Primary Subject Categories	Authenticated User	Unauthenticated User	Anonymous User
Sports events	30.59%	22.77%	25.31%
Sports skills	21.68%	22.41%	20.59%
Professional athletes and teams	19.13%	12.72%	17.45%
Sports and Physical Education in China	9.80%	8.67%	7.21%
Sports shaping and weight loss	7.12%	14.89%	14.29%
Sports health	4.63%	8.07%	6.64%
Sports equipment	4.57%	7.68%	5.59%
Sports experience	2.48%	2.78%	2.91%

Of the top five secondary topics (Table 5), authenticated users are most concerned with the topic of Chinese Super League and Korea-Japan World Cup, non-authenticated users are most concerned with sports slimming, and anonymous users are most concerned with NBA players' performance. In addition, among the top five secondary topics, NBA

team and Olympic Games discussions are unique to authenticated users compared to non-authenticated and anonymous users. The topics of fitness consultation, middle distance running, and marathon performance improvement are special to non-authenticated users, and Esports tips are individual to anonymous users.

Table 5. Sports information needs of users with different authentication attributes in secondary subject categories (Top five).

Ranking	Secondary Subject Category	Authenticated User	Secondary Subject Category	Unauthenticated User	Secondary Subject Category	Anonymous User
1	Chinese Super League and Korea-Japan World Cup	5.36%	Exercise to lose weight	4.39%	NBA player performance	5.70%
2	NBA players' performance	5.36%	NBA player performance	3.91%	Chinese Super League and Korea-Japan World Cup	5.43%
3	UEFA	5.16%	Fitness consultation	3.85%	UEFA	5.00%
4	NBA team	4.80%	UEFA	3.73%	E-Sports game skills	4.81%
5	Olympics	4.73%	Middle-distance running and Marathon performance improvement	3.73%	Exercise to lose weight	4.64%

5. Limitations

There are three limitations of the study. First, we only studied one platform, Zhihu, which is the largest online Q&A community in China. It does not wholly represent Chinese netizens' sports information needs. Other online communities, such as Baidu Know, need further research to reflect their sports information needs more accurately. Second, in terms of the methodology, we chose the more mature BERT model, combining K-means and TF-IDF for topic analysis. With the development of NLP technology, more and more short text models suitable for texts of social networks have been created, and more accurate topic mining models could be used in future studies. Finally, the presentation of questions is just one side of the demand for sports information.

6. Discussion

The Internet is changing the way people disseminate and access information. Online Q&A communities, with their unique advantages of continuity, openness, timeliness, anonymity, and content diversity, have become an essential source of information and knowledge for the public nowadays, which is an online ecology that many disciplinary fields cannot ignore. In terms of quantitative trends in demand for sports information, the increase before decrease trend does not indicate a decrease in users' needs for sports information. Because the number of relevant questions is increasing under the sports topic, not all of them will be displayed. For many discussions, people ask a lot of repetitive questions. These repetitive questions can cause content and information to be scattered, causing trouble for the answerer, the reader, and the questioner. Zhihu employs a question redirection mechanism, which provides higher value by automatically jumping from question page to question page so that discussions and thoughts about an issue can be presented more centrally on a single page. This mechanism is enabled when two or more questions are duplicated. Sometimes the text may be different, but if the question is essentially about the same thing, this mechanism is also triggered so that the questions generated are new ones, which also encourages users to try searching before asking a

question, and if the question already exists, to stop asking the question and check the answer under the question [57]. After the peak in 2016, the number of sports information needs is above 10,000 every year, reflecting that the development of Zhihu continues to mature and the public sports information needs tend to be stable.

The differences in sports information needs between male and female users are also noticeable, reflecting different motivations for sports participation among users of different genders. From the differences in information needs, men pay more attention to solid and aggressive sports such as basketball and soccer, especially information about NBA, European and American soccer, and the Chinese Super League. On the other hand, women pay more attention to weak aggressive sports such as running, swimming, yoga and focus on weight loss and shaping effects brought by sports and sports protection issues. In addition, this study also found that women prefer to emphasize their gender attributes when expressing their needs for sports information to obtain more appropriate information. It suggests that women are disadvantaged in online Q&A communities, especially in sports topic discussions, where women tend to be marginalized and subconsciously view the discussion of sports information needs in Q&A communities as a male-dominated arena.

The need for sports information somehow reflect people's concern about the sports industry or sports. This study can provide an intuitive understanding of the sports information needs of the Chinese. For the sports authorities, this study could help them understand the sports hotspots people are concerned about and the problems in the sports participation process in the online Q&A community. Therefore, based on the results of this study, sports authorities could provide solutions and strategies which could better optimize the quality of the answers, address the problems raised by the users in the online Q&A community, and enhance the scientific nature of the public sports participation. For example, sports coaches, players, and experts could be organized to answer the questions the users are concerned about in the community. Secondly, the results show people's concern about sports music, sports equipment, scientific fitness and other topics, which could provide reference and direction for the products and services of sports enterprises. In addition, the exploration of user attributes is an addition to the sociology of sports and sports communication.

7. Conclusions and Future Work

By taking the sports-related questions of Zhihu as a sample, this study found that the sports information needs in China's online Q&A community present the following three main characteristics. First, the number of sports information needs to be formed three distinct phases around the three Olympic periods, with the Rio Olympic period having the highest number of questions. Second, the information covers eight primary topics and 40 secondary topics, with rich content and a balanced proportion. The number of the topics is relatively balanced regarding the secondary topics, with most topics accounting for about 2% and 3% of the overall number. The topics of sport-related information needs cover sports skills, sports events, sports shaping and weight loss, professional athletes and teams, Chinese sports and physical education, sports health, sports equipment, and sports experience, reflecting the diversity and richness of sports information needs by users of the online Q&A community. Finally, the data based on the known gender of users showed that the percentage of male users was 74.07%. It reflects that male users in online question and answer communities show more attention to sports information needs than female users. Male users dominate sports information needs by number.

This article is an exploratory study, and it is a good attempt to study sports issues in online Q&A communities. There is much more work that can be completed in the future. First, more studies could focus on user issues, gender issues and power issues of sports topics in online Q&A communities to provide more academic exploration. User responses for Q&A community sports topics and quality content generation are also valuable research directions for sports development in the social media era. Second, in terms of information mining techniques, researchers can try more advanced and precise techniques to enhance the accuracy of their research. It is also worth noting that we anticipate seeing more

studies on the varied consequences of sports information requirements' features. What fundamental changes will these features bring to the promotion of mass sports, China's sports industry's future development, and the link between sports and society? What are the strategies for optimizing and responding to them? These will significantly impact how people, the media, and sports interact in the future.

Author Contributions: Conceptualization, C.N. and T.W.; methodology, C.N. and T.W.; validation, C.N. and H.G.; formal analysis, X.Y.; data curation, C.N.; writing—original draft preparation, C.N.; writing—review and editing, H.G.; supervision, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guan, T.; Wang, L.; Jin, J.; Song, X. Knowledge contribution behavior in online Q&A communities: An empirical investigation. *Comput. Human Behav.* **2018**, *81*, 137–147. [CrossRef]
2. Zhao, L.; Detlor, B.; Connelly, C. Sharing knowledge in social Q&A sites: The unintended consequences of extrinsic motivation. *J. Manag. Inform. Syst.* **2016**, *33*, 70–100. [CrossRef]
3. Yang, F. The production and dissemination of Confucian knowledge in international knowledge-sharing communities—An examination of Quora. *Dongyue Ser.* **2021**, *7*, 85–95.
4. Schleifer, T. The Question-and-Answer Quora Platform Is Now Worth \$2 Billion. Available online: <https://www.vox.com/recode/2019/5/16/18627157/quora-value-billion-question-answer> (accessed on 18 February 2022).
5. Yang, S.; Yan, Z.W. The Number of Users Exceeds 220 Million, Exploring Different Paths to Cash. Available online: <https://ishare.ifeng.com/c/s/7mF5LX9EPfs> (accessed on 18 February 2022).
6. Kuang, L.; Huang, N.; Hong, Y.; Yan, Z. Spillover effects of financial incentives on non-incentivized user engagement: Evidence from an online knowledge exchange platform. *J. Manag. Inf. Syst.* **2019**, *36*, 289–320. [CrossRef]
7. Jiao, Z.; Chen, J.; Kim, E. Modeling the Use of Online Knowledge Community: A Perspective of Needs-Affordances-Features. *Comput. Intell. Neurosci.* **2021**, *2021*, 3496807. [CrossRef] [PubMed]
8. Shunli, G.; Xiangxian, Z.; Xing, T.; Liman, Z. Research on Automated Evaluation of User Generated Answer Quality in Social Question and Answer Community—Taking “Zhihu” as an example. *Libr. Inf. Serv.* **2019**, *63*, 118. [CrossRef]
9. Wang, J.; Li, Z.; Feng, H.; Guo, Y.; Liang, Z.; Wang, L.; Wan, X.; Wang, Y.; Visvizi, A.; Lytras, M.D. A Research on the Development Trend of Knowledge Payment Based on Zhihu. In *The New Silk Road Leads through the Arab Peninsula: Mastering Global Business and Innovation*; Emerald Publishing Limited: Bingley, UK, 2019. [CrossRef]
10. Zhang, X.; Liu, S.; Chen, X. Social capital, motivations, and knowledge sharing intention in health Q&A communities. *Manag. Decis.* **2017**, *55*, 1536–1557. [CrossRef]
11. Rosenbaum, H.; Shachaf, P. A structuration approach to online communities of practice: The case of Q&A communities. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 1933–1944. [CrossRef]
12. Pu, J.; Liu, Y.; Chen, Y.; Qiu, L.; Cheng, H.K. What questions are you inclined to answer? Effects of hierarchy in corporate Q&A communities. *Inf. Syst. Res.* **2022**, *33*, 244–264.
13. Vayansky, I.; Kumar, S.A. A review of topic modeling methods. *Inf. Syst.* **2020**, *94*, 101582. [CrossRef]
14. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [CrossRef]
15. Alasmari, A.; Zhou, L. How multimorbid health information consumers interact in an online community Q&A platform. *Int. J. Med. Inform.* **2019**, *131*, 103958. [CrossRef]
16. Jin, J.; Yan, X.; Li, Y.; Li, Y. How users adopt healthcare information: An empirical study of an online Q&A community. *Int. J. Med. Inform.* **2016**, *86*, 91–103. [CrossRef]
17. Xie, X.; Huang, Y. From platform to the community: A study on the evolution of online community Q&A. *Publ. Sci.* **2018**, *26*, 14–19.
18. Bao, Z.; Han, Z. What drives users' participation in online social Q&A communities? An empirical study based on social cognitive theory. *ASLIB J. Inform. Manag.* **2019**, *71*, 637–656. [CrossRef]
19. Wang, J.; Luo, Y.; Hao, J.; Ding, L.; Zhang, R. Who are influential in Q&A communities? A measure of V-Constraint based on knowledge diffusion capability. *J. Inf. Sci. Eng.* **2018**, *45*, 488–501. [CrossRef]

20. Guo, C.; Caine, K. Anonymity, user engagement, quality, and trolling on Q&A sites. *Proc. ACM Hum. Comput. Interact.* **2021**, *5*, 1–27. [CrossRef]
21. Chen, L. The impact of content commenting on user continuance in online Q&A communities: An affordance perspective. *arXiv* **2020**, arXiv:2001.08927. [CrossRef]
22. Shi, J.; Shen, H.; Ma, Q. What kind of answer will be better: Exploring the features of high-quality answer contents in social Q&A community. In Proceedings of the 19th International Conference on Electronic Business (ICEB19), Newcastle upon Tyne, UK, 8–12 December 2019.
23. Dervin, B. *Information as a User Construct: The Relevance of Perceived Information Needs to Synthesis and Interpretation*; Temple University Press: Philadelphia, PA, USA, 1983; p. 170.
24. Fourie, I. A call for libraries to go green. *Libr. Hi-Tech.* **2012**, *30*, 428–435. [CrossRef]
25. Case, D.; Given, L. *Looking for Information*; Emerald Group Publishing: Bingley, UK, 2016; pp. xv–xvi.
26. Hu, C. *Information Services and Users*, 4th ed.; Wuhan University Press: Wuhan, China, 2015; p. 121.
27. Xing, W.; Hong, F. Research on the information needs of library users based on microblog interaction. *New Cent. Lib.* **2019**, *7*, 5.
28. Jia, Z.; Meng, T. Information as an axis: Media use, information demand, and media trust during the new coronavirus outbreak. *E-Government* **2020**, *209*, 20–33.
29. Wang, J.; Zhi, Y. A study on the thematic characteristics of weight loss information needs in online question and answer communities from the perspective of gender differences: The example of “Zhihu”. *Mod. Intell.* **2021**, *41*, 89–96, 131.
30. Huang, L.; Jiang, L.; Miao, H. Topic identification and analysis based on online question and answer communities: The example of Zhihu’s “elderly” topic. *Lib. Intell. Work* **2016**, *60*, 94–101.
31. Yue, H.; Yue, F.; Shupeng, Z.; Yufeng, M. Recommending Contents Based on Zhihu Q&A Community: Case Study of Logistics Topics. *Data Anal. Knowl. Discov.* **2018**, *2*, 42–49. [CrossRef]
32. Bahng, J.; Lee, C.H. Topic Modeling for Analyzing Patients’ Perceptions and Concerns of Hearing Loss on Social Q&A Sites: Incorporating Patients’ Perspective. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6209. [CrossRef]
33. Chen, Y.; Dong, T.; Ban, Q.; Li, Y. What Concerns Consumers about Hypertension? A Comparison between the Online Health Community and the Q&A Forum. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 734–743. [CrossRef]
34. Zhao, W.; Lu, P.; Yu, S.; Lu, L. Consumer health information needs in China: A case study of depression based on a Social Q&A community. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 130. [CrossRef]
35. Jiang, H.; Qiang, M.; Zhang, D.; Wen, Q.; Xia, B.; An, N. Climate change communication in an online Q&A community: A case study of quora. *Sustainability* **2018**, *10*, 1509. [CrossRef]
36. Karbasian, H.; Johri, A. Insights for curriculum development: Identifying emerging data science topics through analysis of Q&A communities. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education, New York, NY, USA, 11–14 April 2020; pp. 192–198. [CrossRef]
37. Liu, G.; Wei, Y.; Li, F. Understanding Consumer Preferences—Eliciting Topics from Online Q&A Community. In Proceedings of the 18th International Conference on Electronic Business, Guilin, China, 6 December 2018; pp. 2–6.
38. Zhang, W. Text Mining Applied in Evolution of Q&A Platforms users’ Information Demand on Tourism in COVID-19 Normalization. In Proceedings of the 2021 5th Annual International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China, 24–26 September 2021; pp. 29–40.
39. Chen, X.; Wang, H. Automated chat transcript analysis using topic modeling for library reference services. *Proc. Assoc. Inf. Sci. Technol.* **2019**, *56*, 368–371. [CrossRef]
40. Luo, X.; Ding, H.; Tang, M.; Gandhi, P.; Zhang, Z.; He, Z. Attention mechanism with bert for content annotation and categorization of pregnancy-related questions on a community Q&A site. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea, 16–19 December 2020; pp. 1077–1081. [CrossRef]
41. Han, J.Y.; Heo, G.E. Analyzing Students’ Non-face-to-face Course Evaluation by Topic Modeling and Developing Deep Learning-based Classification Model. *J. Korean Soc. Libr. Inf. Sci.* **2021**, *55*, 267–291. [CrossRef]
42. Qian, Y.; Gui, W. Identifying health information needs of senior online communities users: A text mining approach. *Aslib J. Inf. Manag.* **2020**, *73*, 5–24. [CrossRef]
43. Lu, J.; Plataniotis, K.N.; Venetsanopoulos, A.N. Face recognition using LDA-based algorithms. In *IEEE Transactions on Neural Networks*; IEEE: Piscataway, NJ, USA, 2003; pp. 195–200. [CrossRef]
44. Basmatkar, P.; Maurya, M. An Overview of Contextual Topic Modeling Using Bidirectional Encoder Representations from Transformers. In Proceedings of the Third International Conference on Communication, Computing and Electronics Systems, Coimbatore, India, 28–29 October 2021; Springer: Singapore, 2022; pp. 489–504.
45. Wang, H. Development of Natural Language Processing Technology. *ZTE Technol. J.* **2022**. Available online: <http://kns.cnki.net/kcms/detail/34.1228.TN.20220408.1420.004.html> (accessed on 28 April 2022).
46. Liu, H.; Zhang, Z.; Wang, Y. A review of the primary optimization improvement methods of the BERT model. *Data Anal. Knowl. Dis.* **2021**, *5*, 3–15.
47. Glazkova, A. Identifying topics of scientific articles with BERT-based approaches and topic modeling. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Delhi, India, 11 May 2021; pp. 98–105.
48. Abuzayed, A.; Al-Khalifa, H. BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique. *Procedia Comput. Sci.* **2021**, *189*, 191–194. [CrossRef]

49. Slapoguzov, A.; Malyuga, K.; Tsopa, E. Word sense induction for Russian texts using BERT. In Proceedings of the 28th Conference of Fruct Association, Moscow, Russia, 25–29 January 2021; pp. 621–627.
50. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv* **2019**, arXiv:1908.10084. [[CrossRef](#)]
51. Thakur, N.; Reimers, N.; Daxenberger, J.; Gurevych, I. Augmented sbert: Data augmentation method for improving Bi-encoders for pairwise sentence scoring tasks. *arXiv* **2020**, arXiv:2010.08240. [[CrossRef](#)]
52. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426. [[CrossRef](#)]
53. Salton, G.; Yu, C.T. On the construction of effective vocabularies for information retrieval. *Acm. Sigplan Not.* **1973**, *10*, 48–60. [[CrossRef](#)]
54. Ding, X.; Wang, L. Research on the optimization method of calculating the weight of text feature words in online forums. *Intell. Theor. Pract.* **2021**, *44*, 187–192.
55. Vasilescu, B.; Capiluppi, A.; Serebrenik, A. Gender, representation and online participation: A quantitative study. *Interact. Comput.* **2013**, *26*, 488–511. [[CrossRef](#)]
56. Stice, E.; Shaw, H. Role of body dissatisfaction in the onset and maintenance of eating pathology. *J. Psychosom. Res.* **2002**, *53*, 985–993. [[CrossRef](#)]
57. Zhihu. What Is Question Redirection? Available online: <https://www.zhihu.com/question/19570036> (accessed on 20 April 2022).