RGSA – Revista de Gestão Social e Ambiental
ISSN: 1981-982X
Submission date: 02/02/2023
Acceptance date: 04/02/2024
DOI: https://doi.org/10.24857/rgsa.v18n9-068
Organization: Interinstitutional Scientific Committee
Chief Editor: Christian Luiz da Silva
Assessment: Double Blind Review pelo SEER/OJS

THE TRENDS OF POTENTIAL USER RESEARCH FROM 2014-2023 BASED ON BIBLIOMETRIC AND BERTOPIC

Liu Kun [1]
Hassan Alli [2]
Khairul Aidil Azlin Abd Rahman[3]

ABSTRACT

Objective: Despite the increasing importance of lead generation research in increasing product or market share, cost and resource constraints have become a challenge for SMEs.Therefore, this study aims to explore and reveal research themes and market trends hidden in articles on lead generation over the past 10 years.
.
Theoretical Framework: In this study, qualitative and quantitative methods are combined, and three methods of bibliometrics, network analysis and BERTopic topic modeling are used to analyze the literature.

Method: A total of 7446 articles were analysed using bibliometrics, network analysis and BERTopic thematic modelling as the basis of a mixed method approach.

Results and Discussion: The study found that the field is currently experiencing a downward trend after a phase of rapid growth. During this period, the United States and China were the countries with the highest number of articles accounting for 77% of the total; the Journal of Cleaner Journal of Cleaner Production was the most cited journal. In addition, the potential user studies cover 43 mainstream topics, focusing on 6 aspects . In the in-depth analysis of the theme evolution, it was found that the potential user study gradually evolved from the initial multidimensional application to focus on open service, and was more oriented towards the public service field.

Research Implications: This provides a strong theoretical basis and practical guidance for identifying potential customers and increasing conversion rates and revenues.

Originality/Value: To our knowledge, this is the first study to use a mixed-methods approach to lead generation, which will help researchers to tackle more complex challenges and changes in the future.

Keywords: Prospective Users, Bibliometric Analysis, Topic Evolution, BERTopic Models, Text Data Mining.

AS TENDÊNCIAS DA PESQUISA DE USUÁRIOS POTENCIAIS DE 2014-2023 COM BASE EM BIBLIOMÉTRICA E BERTOPIC

RESUMO

Objectivo: Apesar da importância crescente da investigação sobre geração de leads no aumento da quota de produto ou de mercado, as restrições de custos e recursos tornaram-se um desafio para as PME. Portanto, este estudo visa explorar e revelar temas de investigação e tendências de mercado escondidas em artigos sobre geração de leads ao longo do últimos 10 anos.

Referencial Teórico: Neste estudo, métodos qualitativos e quantitativos são combinados, e três métodos de bibliometria, análise de redes e modelagem de tópicos BERTopic são utilizados para analisar a literatura.

[1] Faculty of Design and Architecture, Universiti Putra Malaysia, Malaysia. E-mail: gs60973@student.upm.edu.my
Orcid: https://orcid.org/0009-0007-8698-4950
[2] Faculty of Design and Architecture, Universiti Putra Malaysia, Malaysia. E-mail: halli@upm.edu.my
Orcid: https://orcid.org/0000-0003-1727-4561
[3]Faculty of Design and Architecture, Universiti Putra Malaysia, Malaysia. E-mail: drkhairulazlin@upm.edu.my
Orcid: https://orcid.org/0000-0002-4091-7353

Método: Foram analisados 7.446 artigos utilizando bibliometria, análise de rede e modelagem temática BERTopic como base de uma abordagem de método misto.

Resultados e Discussão: O estudo constatou que o campo está atualmente passando por uma tendência de queda após uma fase de rápido crescimento. Nesse período, os Estados Unidos e a China foram os países com maior número de artigos, representando 77% do total; o Journal of Cleaner Journal of Cleaner Production foi o periódico mais citado. Além disso, os estudos de potenciais utilizadores abrangem 43 tópicos principais, centrando-se em 6 aspectos. Na análise aprofundada da evolução do tema, constatou-se que o estudo do potencial utilizador evoluiu gradualmente da aplicação multidimensional inicial para se concentrar no serviço aberto, e foi mais orientado para o campo do serviço público.

Implicações de pesquisa: Isto fornece uma base teórica sólida e orientação prática para identificar clientes potenciais e aumentar as taxas de conversão e receitas.

Originalidade/Valor: Até onde sabemos, este é o primeiro estudo a utilizar uma abordagem de métodos mistos para geração de leads, o que ajudará os pesquisadores a enfrentar desafios e mudanças mais complexos no futuro.

Palavras-chave: Potenciais Usuários, Análise Bibliométrica, Evolução de Tópicos, Modelos BERTópicos, Mineração de Dados De Texto.

LAS TENDENCIAS DE LA INVESTIGACIÓN DE USUARIOS POTENCIALES DEL 2014-2023 BASADA EN BIBLIOMETRÍA Y BERTOPIC

RESUMEN

Objetivo: A pesar de la creciente importancia de la investigación sobre generación de leads para aumentar la participación de productos o mercados, las limitaciones de costos y recursos se han convertido en un desafío para las PYMES. Por lo tanto, este estudio tiene como objetivo explorar y revelar temas de investigación y tendencias de mercado ocultas en artículos sobre generación de leads a lo largo del tiempo. últimos 10 años.
.
Marco teórico: En este estudio, se combinan métodos cualitativos y cuantitativos, y se utilizan tres métodos de bibliometría, análisis de redes y modelado de temas BERTopic para analizar la literatura.

Método: Se analizaron un total de 7446 artículos utilizando bibliometría, análisis de redes y modelado temático BERTopic como base de un enfoque de método mixto.

Resultados y Discusión: El estudio encontró que el campo está experimentando actualmente una tendencia a la baja después de una fase de rápido crecimiento. Durante este período, Estados Unidos y China fueron los países con mayor número de artículos representando el 77% del total; el Journal of Cleaner Journal of Cleaner Production fue la revista más citada. Además, los estudios de usuarios potenciales cubren 43 temas principales, centrándose en 6 aspectos. En el análisis en profundidad de la evolución del tema, se encontró que el estudio de usuarios potenciales evolucionó gradualmente desde la aplicación multidimensional inicial para centrarse en el servicio abierto, y se orientó más hacia el campo del servicio público.

Implicaciones de la investigación: esto proporciona una base teórica sólida y una guía práctica para identificar clientes potenciales y aumentar las tasas de conversión y los ingresos.

Originalidad/Valor: Hasta donde sabemos, este es el primer estudio que utiliza un enfoque de métodos mixtos para la generación de leads, lo que ayudará a los investigadores a abordar desafíos y cambios más complejos en el futuro.

Palabras clave: Usuarios Potenciales, Análisis Bibliométrico, Evolución del Tema, Modelos BERTopic, Minería de Datos Textuales.

1 INTRODUCTION

At the beginning of the 20th century, market research and consumer surveys, enterprises recognized the importance of potential consumer demand for the successful marketing of products [1].In the mid-to-late 20th century, the field of consumer research and marketing continued to develop, and more in-depth analysis of potential users became the mainstream, and began to use a more systematic approach to understand consumer psychology, behaviour and preferences, as a way to develop product marketing strategies [2]. At the beginning of the 21st century, with the application of big data and the Internet of Things, although it is possible to depict potential users more comprehensively, some enterprises rely too much on big data analysis, while ignoring the in-depth understanding of the real user needs, resulting in prediction work faced with uncertainty bias [3].Gleanster report shows that 56% of respondents said Potential user cultivation performance is poor/inadequate or needs improvement, while 84% said Potential user nurturing was challenging or very challenging[4]. . According to Forrester Research, companies that excel at user nurturing can increase the number of effective demands by 50% while reducing customer acquisition costs by 33% [5] . The role of potential user research in promoting enterprise sales conversion is self-evident.

In recent years, the unfolding of the construction trend of the digital economic system, the increasing awareness of user consumption, product and service demand is increasingly diverse, cross-disciplinary, multi-dimensional disciplines discipline value is also inevitable in the cross-fertilisation and mutual influence with other disciplines highlighted [6]. Due to the characteristics of the purchase decision cycle is uncertain, the decision-making process is complicated, and most small and medium-sized enterprises (SMEs) are forced to cost and resource constraints, which leads to the enterprise can not judge the real willingness of potential users [7]. So far, SMEs have been increasing their attempts to explore the techniques and methods of potential users, and the study of judging the true intentions of potential users has become the focus of many enterprises [8]. A number of professors and scholars around the world have made great efforts to explore potential users, and published a large number of papers, although the academic background, angle of entry, research content, and for the scenarios and so on are different, about the concept of potential users and characteristics of their own insights, although they use different terminology, but essentially about the concept of the potential user to define and analyse the characteristics of the enterprise's product innovation practice has an important role to play in promoting [9]. Then from a multidisciplinary, interdisciplinary perspective to discover the field of research on the characteristics of potential

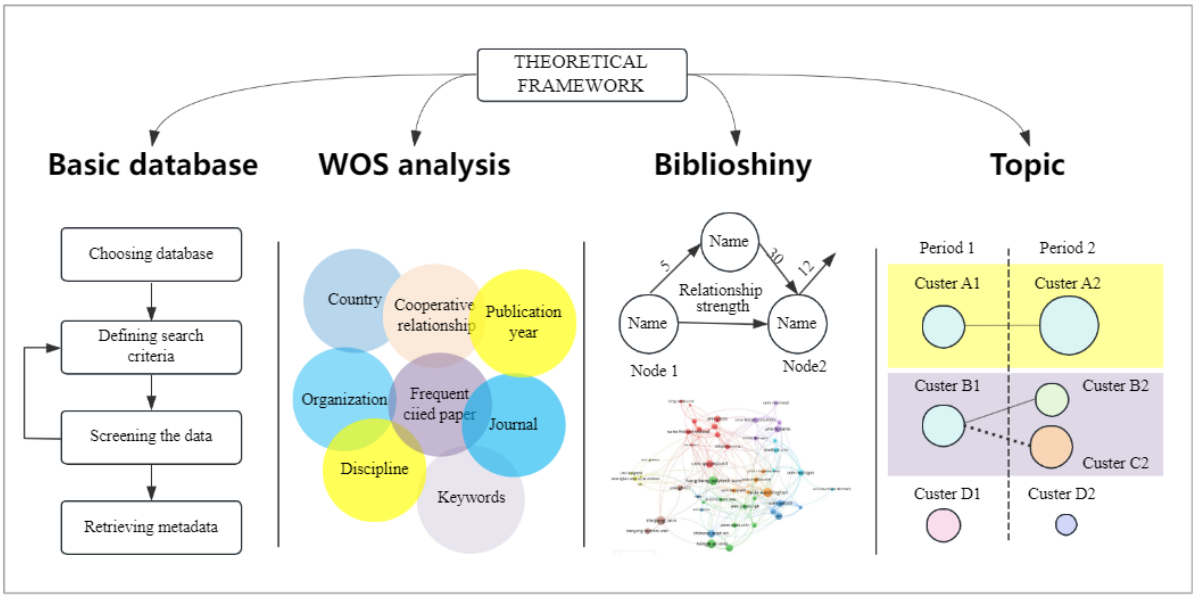
users, for some small and medium-sized enterprises market mining to provide valuable direction is imminent, however, to summarise the progress of these studies to provide help for new and old researchers of the relevant research has not yet been reported.

This paper examines the acquisition of potential user data through Web of Science (WoS) for a variety of data situations using mature bibliometrics. Bibliometrics is the automated quantitative study of literature output, citation, collaboration, and other factors to reveal the rules and trends of research activity and to assess the quality and impact of scientific research results [10]. Simultaneously, it is important to consider the global text data in a semantically complex manner to gain a deeper understanding of the text content [11]. To handle more complex contextual meanings, the BERTopic deep learning model can be used to learn the semantic information of the text [12]. This approach provides a more comprehensive reflection of the research Topic and evolutionary trends compared to traditional topic models [12-13].

2 THEORETICAL FRAMEWORK

Figure 1

Potential User theortical Framework by author



This paper carries out the following analysis methods framework (Figure 1): Firstly, the data of 7446 SCI papers were used as the basic database. Secondly, (1) Use the analysis function of WoS website to analyse the papers in terms of disciplines and journals, countries and institutions, and international cooperation [14-15]. (2) Import CiteSpace, VOSviewer through

the parameter settings for reference co-citation, keyword focus, time hotspot trend keyword clustering, timeline drawing reflecting the potential users of the research pattern of the knowledge map [16]. Thirdly, (1) BERTopic topic modelling based on sentence transformers model sentence vectors, constructing title and abstract document-level information, transforming sentence vectors into indexes in word lists, downscaling high dimensional embedding spaces using UMAP methods, and then using HDBSCAN clustering algorithm to create dense clusters and retaining important words in the topic [17-18]. (2) Capture the semantic relationship between words by Word2Vec embedding, obtain the topic similarity and generate the evolution trend map [19-20]. Finally, the literature information mined by the contemplation and argumentation map is used to root the research focus and hotspot and explore the research trend and innovation direction.

3 METHODOLOGY

The WOS (Web of Science) database is a commonly used tool for bibliometric analysis [21]. It includes over 12,000 authoritative and high-impact academic journals worldwide, covering natural sciences, engineering and technology, biomedicine, social sciences, arts, and humanities, among others. The database dates to 1900 and is widely recognised as the evaluation system used by the academic community [22]. On 7 December 2023, we conducted a search on the Science and Technology Citation Index (STCI) database's Web of Science Core Collection (WSC) using the search term 'potential users'. The search method employed was (TS=('potential customers*') OR 'potential customers*' OR 'potential consumers*' OR 'potential users*' OR 'hidden customers*' OR 'hidden customers*' OR '2022' OR '2021' OR '2020' OR '2019' OR '2018' OR '2017' OR '2016'). Following data checking and de-duplication, a total of 7,446 documents were included in the in-scope study of prospective trends.

4 RESULTS AND DISCUSSIONS

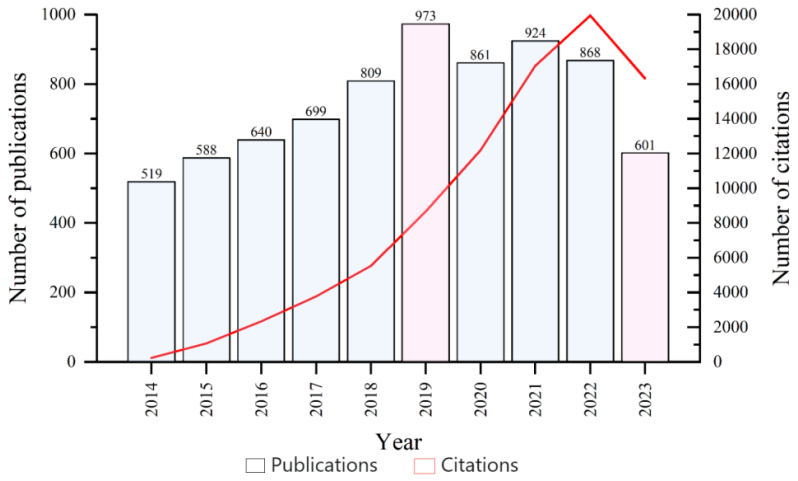
4.1 POTENTIAL USERS STUDY ANNUAL STATUS

According to Figure 2, the research area experienced rapid growth between 2014 and 2019. Specifically, there was a small increase in publications in 2018, followed by a sharp rise in 2019 which peaked. This is the peak of research interest and prosperity in the field. The number of publications fluctuates irregularly between 2020 and 2022, possibly due to

turbulence caused by changes in the external environment and shifts in research hotspots. The significant decline in the number of publications in 2023 suggests that research in the field is facing a challenge or change. In-depth study and interpretation are required to understand the various factors that affect researchers' attention and research focus in the field.

Figure 2

Publication and citation distribution



4.2 POTENTIAL USER STUDY STATUS

4.2.1 Countries and Institutions

By analysing the countries of publication, we can identify the core countries in the research field and reflect academic exchanges and cooperation between countries [23]. According to Table 1, A total of 125 countries or regions have published relevant literature , indicating global interest in the topic. The five countries or regions with the highest number of publications are the USA (3304 articles), China (2412 articles), the UK (1135 articles), Germany (1101 articles), and Spain (690 articles). Between 2014 and 2023, 7,188 institutions were involved in the research of potential users in the Web of Science core ensemble database. The issuing institutions were mainly concentrated in colleges, universities, and research institutes. Citation frequency is an important measure of the value of academic literature. It provides a more objective evaluation of the overall quality of academic literature in each research institution and can also reflect the academic influence of each institution to some extent [24]. According to statistics, the University of Washington (USA) has the highest number of

publications and citations. The University of Cambridge (United Kingdom) ranks 12th in the number of publications and 2nd in the number of citations.

Figure 3 shows the countries with three or more articles visualised by VOSviewer. The size of the nodes represents the number of articles issued, while the thickness of the connecting lines indicates the strength of association between two countries. The node colours represent different clusters [25]. Figure 4 illustrates the uneven distribution of issuing countries in this field, with the USA, China, the UK, Germany, and Spain having a significant impact. The map of national cooperation, shown in Figure 4, highlights that China and the USA have the largest number of papers issued and a more extensive and close cooperation compared to other countries. The number of publications in small countries is often limited due to the size of their research community. It is important to conduct user research on a global scale to gain insights and guidance for business and technological innovation, enabling adaptation to diverse user needs and global market competition.

Table 1

Top 20 country and Organization of potential user field paper output

Rank	Country	NP	TC	Rank	Institutions(Country)	NP	TC
1	USA	3304	15089	1	University Washington (USA)	43	2754
2	CHINA	2412	12612	2	Hong Kong Polytechnic University (Hong Kong)	41	520
3	UK	1135	8473	3	Tsinghua University (China)	41	706
4	GERMANY	1101	7723	4	Zhejiang University (China)	41	384
5	SPAIN	690	3051	5	Chinese Academy of Sciences (China)	36	935
6	CANADA	657	3241	6	University of Queensland (Australia)	36	765
7	AUSTRALIA	574	3384	7	Rhein-Westfälische Technische Hochschule Aachen (Germany)	36	430
8	ITALY	560	2540	8	University College London (United Kingdom)	34	488
9	INDIA	505	1407	9	University of Illinois (USA)	34	971
10	FRANCE	477	1576	10	University of Oxford (United Kingdom)	32	1485
11	NETHERLANDS	400	2985	11	Delft University of Technology (Netherlands)	29	697
12	SOUTH KOREA	323	2498	12	University of Cambridge (United Kingdom)	28	2500
13	BRAZIL	318	909	13	Ghent University (Belgium)	28	837
14	POLAND	301	1063	14	Nanyang Technological University (Singapore)	28	434
15	JAPAN	297	553	15	University of British Columbia (Canada)	27	646
16	PORTUGAL	295	587	16	Swiss Federal Institute of Technology (Switzerland)	27	284
17	RUSSIA	269	261	17	University of Sydney (Australia)	27	520
18	SWITZERLAND	242	584	18	Polytechnic University of Valencia (Spain)	27	146
19	SWEDEN	222	1242	19	University of Sao Paulo (Brazil)	26	161
20	INDONESIA	219	224	20	University of Chinese Academy of Sciences (China)	26	482

Note: NP, number of publications; TC, total citations; PC, citations per article

Figure 3

Country Collaboration World Map

4.2.2 Disciplines and Journals

Table 2

Top 20 categories and journals in terms of potential user field paper output

Rank	Subject Categories	NP	Rank	Journal	NP	TC
1	Computer Science Information Systems	1006	1	Sustainability	117	1071
2	Computer ScienceTheoryMethods	794	2	PLOS ONE	50	1021
3	Engineering Electrical Electronic	712	3	Journal of Cleaner Production	40	1258
4	Business	665	4	IEEE Access	39	577
5	Computer Science Artificial Intelligence	653	5	Journal of Medical Internet Research	26	362
6	Management	513	6	Journal of Business Research	23	417
7	Computer Science Interdisciplinary Applications	465	7	Technological Forecasting and Social Change	21	678
8	Environmental Sciences	400	8	Journal of Retailing and Consumer Services	21	455
9	Economics	375	9	Transportation Research Part A-Policy and Practice	20	662
10	Telecommunication	350	10	Expert Systems with Applications	20	434
11	Computer Science Software Engineering	302	11	Decision Support Systems	18	430
12	Green Sustainable Science Technology	281	12	Computers in Human Behavior	15	738
13	Environmental Studies	263	13	JMIR Mhealth and Uhealth	15	308
14	Operations Research Management Science	255	14	Electronic Commerce Research and Applications	14	507
15	Public Environmental Occupational Health	233	15	International Journal of Medical Informatics	13	906
16	Information Science Library Science	215	16	Remote Sensing	13	368
17	Health care sciences Services	210	17	International Journal of Bank Marketing	12	522
18	Computer Science Cybernetics	192	18	International Journal of Information Management	11	984
19	Transportation	182	19	Energy Policy	11	506
20	Transportation Science	179	20	Industrial Management & Data Systems	9	524

Potential user research promotes interdisciplinary development across social sciences, natural sciences, and engineering. The top five disciplines with the most influential potential

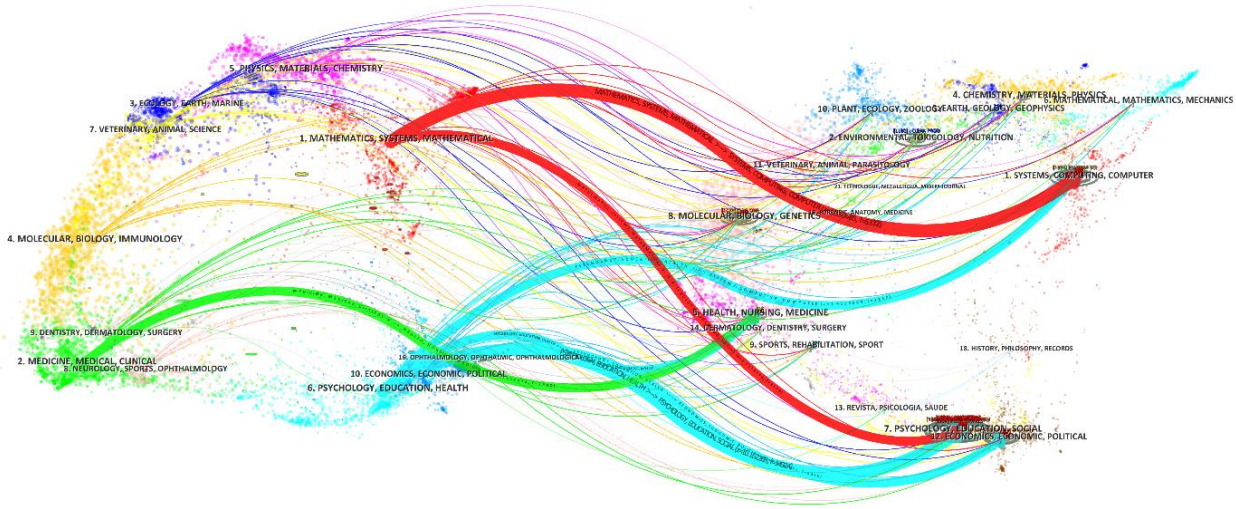
user field outputs, in order, are Computer Science Information Systems, Computer Science Theory Methods, Engineering Electrical Electronic, Business, and

Computer Science Artificial Intelligence (Table 2). Publication platforms such as Sustainability, PLOS ONE, and Journal of Cleaner Production are important in the potential user areas. The high citation of these journals suggests that their research has a wide impact in the academic community. The interdisciplinary research demonstrates the complexity and diversity of the potential user problem, emphasizing the integrative and interdisciplinary nature of potential user research.

The Brundle algorithm [26] is used to form journal clusters, and CiteSpace is used to obtain a double overlay mapping of the discipline distribution of potential users. The double-layer mapping overlay comprises of two sides: the left side shows the distribution of journals where the articles are applied, while the right side shows the distribution of journals of the cited literature. The right side represents the disciplines that are mainly cited in the research of potential users. The left side can be used as an application area, and the right side as a basis for research [27].

Figure 4

Dual map overlay of journals



The Z-Score algorithm is used to standardise citation links and present citation relationships between left and right journals more concisely. Three distinct sets of citation trajectories are obtained based on four main disciplines (right side of Figure 5):Psychology, Education, Social; Economics, Economic, Political; and Systems, Computing, Computer; and Health, Nursing, Medicine. The four most frequently published journals in the disciplines of

transportation, economics, management strategy, and energy economics are Transportation Research Part A: Policy and Practice, Journal of Economics & Management Strategy, Energy Economics, and Research in Transportation Business and Management. It is important to note that the source text already meets the desired characteristics and is free from errors. The four most frequently published journals in the disciplines of transportation, economics, management strategy, and energy economics are Transportation Research Part A: Policy and Practice, Journal of Economics & Management Strategy, Energy Economics, and Research in Transportation Business and Management. Their relevant findings have been applied to the fields of psychology, education, health, mathematics, systems, mathematical, and medicine. (right side of Figure 5) The journals cited are 'COMPUTERS IN HUMAN BEHAVIOR' and 'FRONTIERS IN PSYCHOLOGY'. This finding suggests that these journals are also cited in disciplines such as psychology, education, transport, and medicine, in addition to the fields of environment, politics, economics, and computing. Thus, this demonstrates that potential user research is not only closely linked to the environment and economy, but also increasingly intertwined with human survival, psychological well-being, and social development.

4.2.3 Literature Citation

A total of 257,799 references were cited in the prospect research. Figure 5 shows the spectrum of references over the years, the earliest reference dates back to 1703, indicating the long history of potential user research. The number of references between 2008 and 2020 exceeds 10,000, demonstrating the sustained interest in potential users. The highest number of references, 18,493, was recorded in 2015, when cloud computing, Internet of Things, and mobile payment drove innovation and digital transformation, leading to increased enthusiasm and vigour for in-depth research [28].

Figure 5

Reference Publication Year Spectroscopy

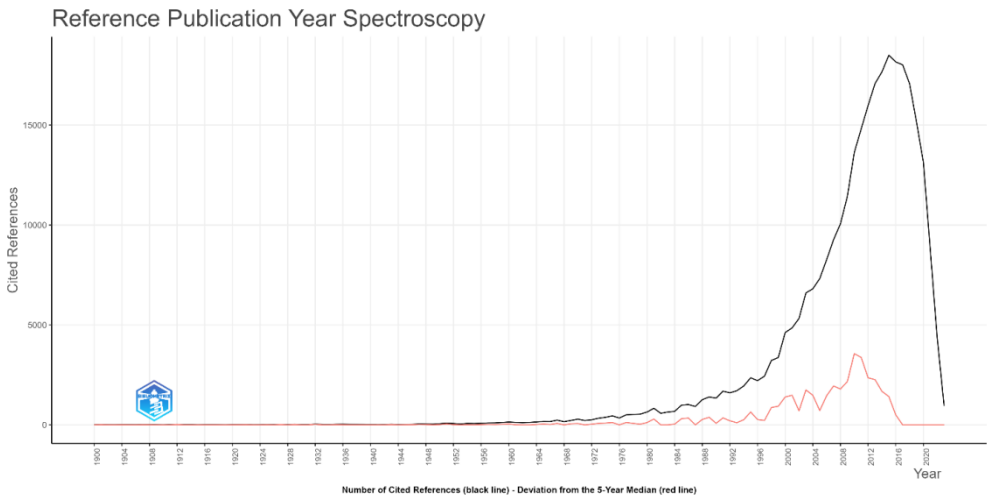
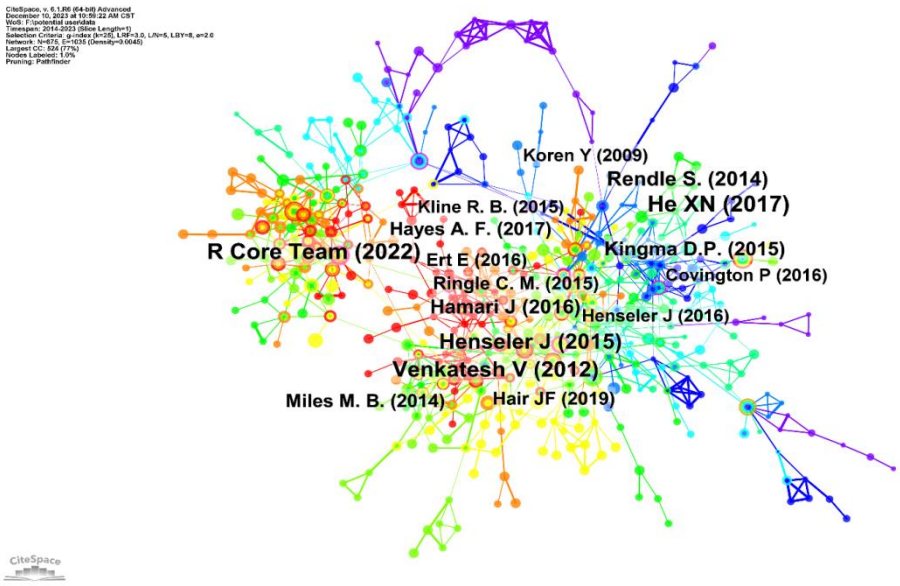


Figure 6

Co-citation of cited references



In 1973, Henry Small, an American bibliographer and information scientist, introduced the concept of co-citation. Co-citation refers to a group of documents or a collection of scholarly papers that are all cited in the same or interrelated literature [29]. A collection of co-cited literature reflects a network of knowledge with similar influence in scholarly research. This, in turn, provides an understanding of the knowledge relationships and research trends in a particular field. Figure 6 shows the final presentation of the co-citation relationship mapping through Citespace. Nodes with higher frequency appear more frequently in other literature.

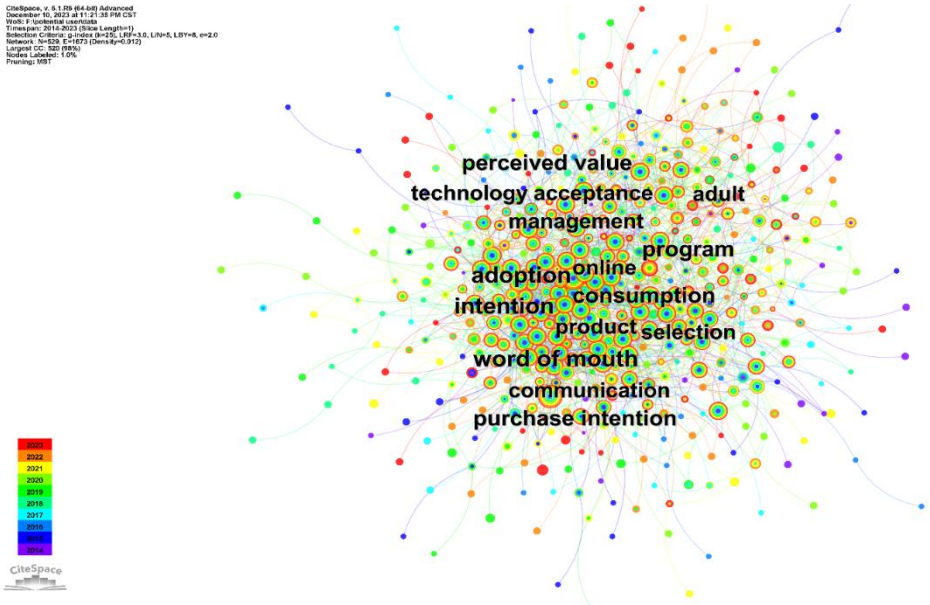
Closely connected nodes are grouped into the same clusters, which usually share similarities in terms of research topics [30]. As a result of our analysis, we identified three additional notable clusters. Cluster 1 includes the R Core Team (2022), which conducts applied research using R, a language and environment for statistical computing. Cluster 2 has the most transformative potential, as demonstrated by Henseler J's (2015) new criterion for assessing discriminant validity in variance-based structural equation modelling and Venkatesh V's (2012) research on consumer acceptance and use of information technology, which extends the unified theory of acceptance and use of technology. Kingma D P.'s (2015) 'Variational Dropout and the Local Reparameterisation Trick' in Cluster 3 is of great interest in applications. Innovative literature can have a significant impact in a short period of time, pushing the frontiers of the field, but its impact may need to be observed on longer time scales.

4.2.4 Hot Spot Analysis

A hotspot is an item of widespread interest that becomes the focus of attention [31]. Keyword co-occurrence analysis can reveal the research direction and characteristics of the research field [32]. To perform keyword analysis, 7446 documents were imported into Citespace, with a time zone segmentation set as '2014-2023', a time slice of 1, and a threshold of K=25. After merging synonyms, the keyword co-occurrence map shown in Figure 8 was obtained. The keyword mapping comprises 529 nodes and 1673 links. The size of each circle in which a keyword is located indicates its frequency of occurrence. The top 10 keywords are word of mouth, intention, adoption, perceived value, consumption, programme, purchase intention, adult, management, and online.

Figure 7

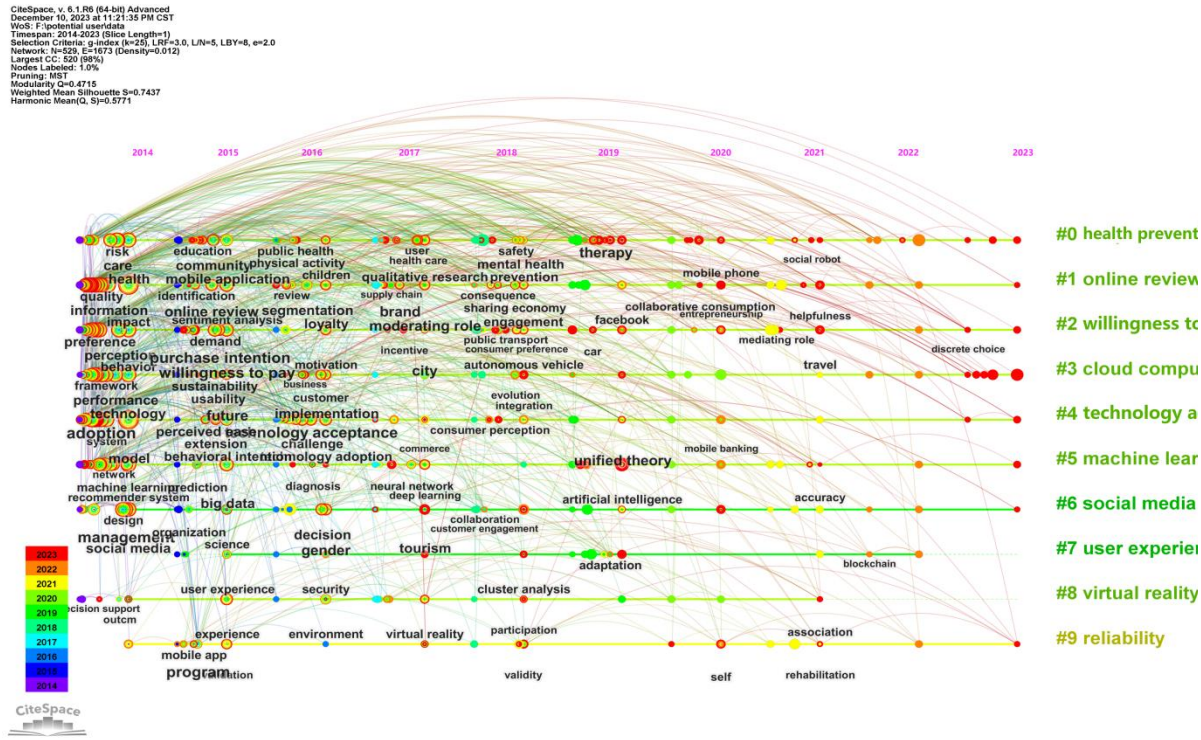
Co-occurrence network



The y-axis represents the number of clusters, and the x-axis represents the year of publication. The visualisation of the keyword co-occurrence network (Figure 8) [33] shows the trend of keyword clustering graph evolution and the time when different keywords appeared. The connecting lines indicate a correlation between the different clusters. The year of research clustering can indicate the maturity and stability of related research. Earlier clustering suggests more mature and stable research, while later clustering may indicate a field in the early stages of development or highlight hotspots. According to Figure 8, from 2014 to 2016, studies on quality, information impact, preference, perception behaviour, FRAME, ADOPTION, and other related potential users gradually increased, and high mediator centrality emerged. From 2017 to 2023, the rapid clustering of health prevention, online reviews, willingness to pay, and machine learning exploration led to the emergence of breakthrough hotspots such as therapy and moderating roles at different points in time. Research results that attracted attention include 'An overview of the features of chatbots in mental health: a scoping review' and 'An overview of the features of chatbots in mental health.' A scoping review investigating m-Health acceptance from a Protection Motivation Theory perspective: gender and age differences. The authors with high citation counts include YANG H, KUMAR P, LEE H, KUMAR A, LIU B, and ZHANG Y (Appendix for details).

Figure 8

Keywords timeline graph



4.3 POTENTIAL USER TOPIC ANALYSIS

Topics typically refer to general concepts and can persist over time with various aspects and dimensions. A topic may consist of multiple related hotspots, which are parts of a topic that attract strong short-term attention. The analysis of text topic categorisation was approached from two perspectives: overall popularity (Global topics) and historical evolutionary trends（Periodic topics）[34]. Topic counts were used to determine the popularity of a topic. Topics ranked in the top 25% by median topic estimate were classified as 'high', those between the 25th and 75th percentile as 'medium', and those below 25% as 'low'. For historical trend classification, Topics were categorized based on their reversal of trend every three years, with their weighting increasing or decreasing over time.

4.3.1 Global Topic

Global topics are theme or subjects that are widely followed, discussed, or cared about as a whole, rather than local or subset topics.Bertopic's unsupervised learning method for extracting topics or Topics from textual data helps to reveal hidden information and

relationships in the text [35]. The embedding model 'sentence transformers' is used, and the number of topics can be determined by setting the 'tropics' parameter. BERTopic merges similar topics until the specified number of topics is reached. However, if the number of topics is too low, even topics with low similarity will be merged, resulting in the identification of too few topics and the exclusion of important ones. To address this issue, BERTopic offers an 'auto' option for the number of topics, which iteratively reduces the number of topics when pairs of topics with a similarity score above 0.9 are found. To maintain objectivity, this paper does not provide a specific value for tropics when training the model. Instead, it sets it to 'auto', which is generated automatically and iteratively by BERTopic. This ensures an appropriate number of topics without subjective evaluation.

Figure 9

Prediction of customer churn by three models

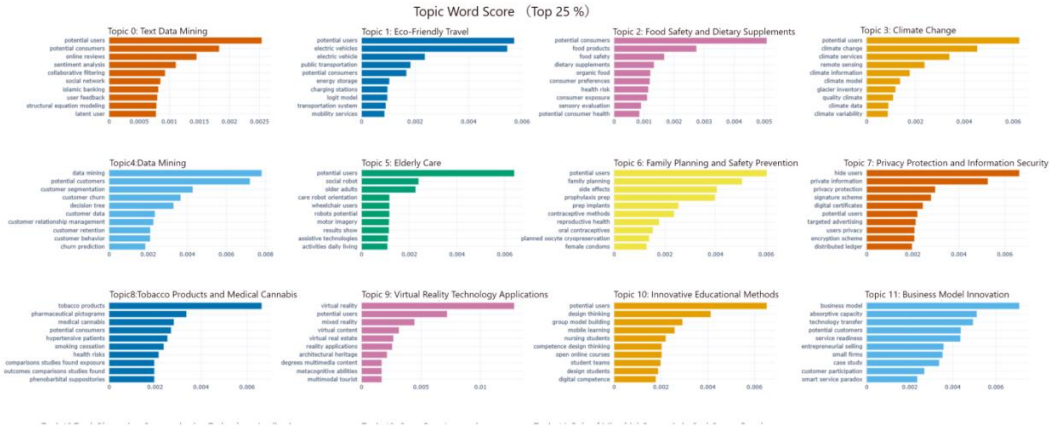


Figure 10

Prediction of customer churn by three models

Figure 11

Prediction of customer churn by three models

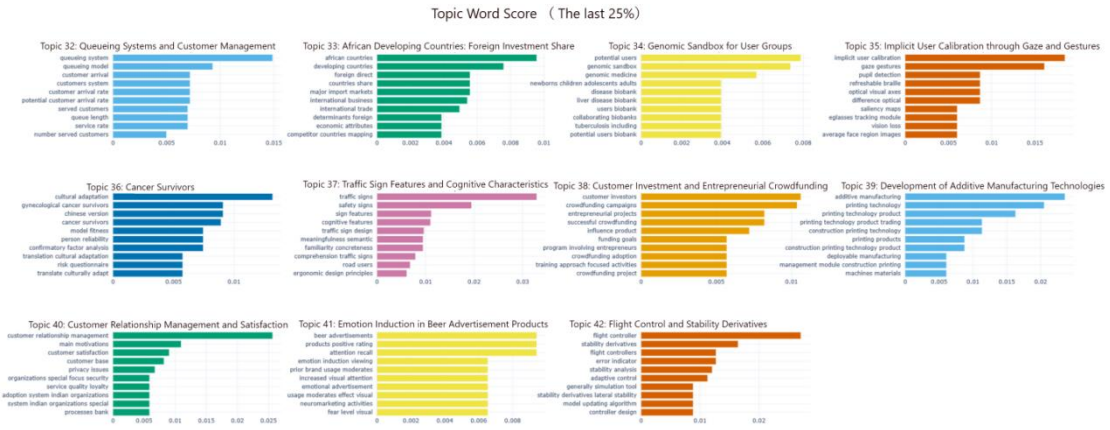


Figure 09-11 displays the 43 topics generated by the BERTopic model along with the Top10 topic terms for each topic. The topic term score indicates the relevance or importance of a keyword or term in each topic, with higher scores indicating greater significance and a closer connection to the topic [36]. To enhance comprehension, we present a visualisation of the weight of each keyword in the topic, ordered from highest to lowest. Additionally, we name the topic based on the keywords to facilitate hierarchical analysis. This process improves

understanding and interpretation of the Topics extracted by the model and provides a reference for subsequent analyses and applications.

Let's analyse the Top 25% （Figure 10）topic results in more detail:

Topic 0 Text Data Mining, which mines and analyses text data, discusses the role of potential users and potential consumers in online reviews and sentiment analysis to extract valuable information and patterns from them. Topic 1 Eco-Friendly Travel, which explores potential users' interest in electric vehicles and public transport, which may be relevant to sustainable transport modes. Topic 2 Food Safety and Dietary Supplements, focuses on potential consumers' questions about food safety and dietary supplements to ensure the safety of the food supply chain with refined health services. Topic 3Climate Change, explores potential user interest in climate change, climate services and remote sensing to address sustainable climate change services. Topic 4 Data Mining discusses lead generation, customer segmentation and churn, looking for hidden patterns and insights to support decision making and innovation. Topic 5 Elderly Care explores potential user interest in social robots, robots for the elderly and caregiving. Topic 6 Family Planning and Safety Prevention, explores potential user issues in family planning, side effects and prevention to improve the quality of family life. Topic 7 Privacy Protection and Information Security, highlights the importance of hidden users, private information and privacy protection and mentions signature schemes. Topic 8 Tobacco Products and Medical Cannabis, discusses the impact of tobacco products, pharmaceutical pictograms, and medical marijuana on potential consumers. Topic 9 Virtual Reality Technology Applications explores the appeal of virtual reality, mixed reality and virtual content to potential users. Topic 10 Innovative Educational Methods explores how to better integrate design thinking and digital technologies into education to enhance student learning experiences and innovation. Topic 11 Business Model Innovation, examines and discusses the absorptive capacity of business models in different sectors and the impact of technology transfer on potential customers to adapt to market and technological changes.

4.3.2 Global Topic Cluster Analysis

The process of dimensionality reduction using UMAP allows the Topics from the Topic model to be embedded in a low-dimensional space, as shown in Figure 12 [37]. Such a visualisation helps to intuitively understand the relationships between Topics, making the Topic analysis more interpretable. Each point on the map represents a Topic and the distribution of

points reflects the relative positions between Topics. Points that are closer together indicate greater similarity between Topics, while points that are further apart indicate greater differences between Topics. The closer the distance between points on the map, the stronger the correlation between the corresponding Topics; conversely, the further the distance, the weaker the correlation. This helps to identify potential clusters of Topics or interrelated Topics [38]. Intersections on the map indicate complex relationships or interesting interactions between Topics, allowing the search for potential cross-cutting Topics.

Figure 12

Prediction of customer churn by three models

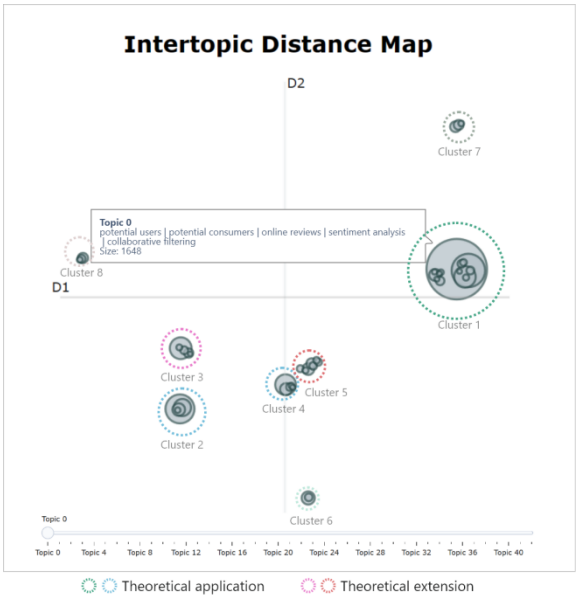


Figure 12 shows that there are three new potential user theory applications (cluster 1, cluster 3, cluster 5) and two theory extensions (cluster 2, cluster 4). Among them, cluster 1 is the most significant, cluster 5 is closely related to cluster 4, and cluster 3 is more closely related to cluster 2. The detailed analysis is as follows

(1) Cluster 1: Digital innovation: Discover personalised, convenient, and innovative services through data analysis to expand profit opportunities.

Topic 0 "Text Data Mining" is the most significant feature, indicating that the advancement of text data mining technology enables enterprises to better understand the needs and feedback of potential users, and improve user satisfaction by mining user reviews, user feedback and social media data to achieve personalised services. Topic 4 "Data Mining" enables enterprises in business, can be used for customer behaviour analysis, market forecasting, to help enterprises better understand the market and consumer demand. This is mainly reflected in

Topic 1 "Eco-friendly travel", which finds that potential users may be more inclined to support eco-friendly travel; Topic 20 "Insurance consumption behaviour", which uses big data to conduct risk analysis and market forecasting. "Risk assessment and personalised pricing through big data, the use of personalised insurance products and online insurance services have changed the way consumers buy and use insurance"; Topic 21 "Library marketing" Digital technologies have enabled libraries to better engage with users, provide online resources and improve the quality of their services. Digital technologies enable libraries to better interact with users, provide online resources, and increase engagement with potential users; Topic 27 'Multilingual Translation Tools' Multilingual translation tools help expand the pool of potential users, facilitate cross-cultural communication, and promote global business and cultural exchange; Topic 17 'Material Durability' New materials for the future. "The durability of new materials has become an important consideration in product design and manufacturing, directly influencing consumer purchasing decisions, and is of great importance to both the manufacturing and consumer markets.

(2) Cluster 2 Health and safety: Creating more comprehensive health management services by focusing on individual and population health.

Topic 2 "Food Safety and Dietary Supplements" pays great attention to food health to improve food safety standards, attention to potential users of dietary supplements for nutritional choices to achieve personalised health care. Topic 6 "Family Planning and Safety Prevention" emphasises concern for the safety of family members, including not only medical solutions for family members, but also more scientific and safer solutions. Family Planning and Safety Prevention" focuses on the safety of family members, including not only medical solutions for family members, but also more scientific and safe birth plans to guide the healthy development of future users. Topic 8 "Tobacco Products and Medical Cannabis" features significant attention to food health to improve food safety standards, focus on potential users of dietary supplements to achieve personalised health care. Topic 8 'Tobacco products and medical cannabis' promotes the tobacco industry's shift towards healthier products, with medical cannabis playing a role in the treatment of specific medical conditions and influencing the health behaviour choices of potential users. This area aims to promote healthy lifestyles where irrational choices or lifestyles may have a negative impact on the health of potential users.

(3) Cluster 3 Low-energy environmental protection: Environment and sustainable development.

Topic3"Climate change" is the most significant, meaning that governments, environmental organisations, businesses, and individuals with a high level of environmental

concern are concerned about the impact of climate change on the environment and society, and are actively seeking solutions for sustainable development. For companies in the industrial, manufacturing and energy sectors, it provides advanced resource management technologies to minimise energy costs, in line with the sustainability objectives of this group of potential users.

(4) Cluster 4 Assistive Care: Assistive technologies for elderly and vulnerable care.

Topic 5 Elderly Care focuses on the potential need for comprehensive care for the elderly, including social support and medical services. The aspect of assistive services for the visually impaired discusses, for example, voice assistants, intelligent guidance systems and other assistive technologies to improve the quality of life and autonomy of the visually impaired. In the area of virtual reality applications, the group explored how this technology can be used to improve the quality of life of older people, covering areas such as medical treatment, entertainment, and social interaction. This group exemplifies the important role of technology in bridging the inequality gap in society, promoting inclusiveness, and improving quality of life, thereby increasing the confidence of potential users.

(5) Cluster 5: Open Services: Integrating open data to create experiential software systems.

Topic 10 Innovative Education Methods Emphasis on the ability to understand and use digital technologies, such as mobile technologies, conducting open online courses, focusing on student-formed teams, etc.Topic13 Methods Open Data Integration Emphasis on the importance of integration, which includes integration across different domains, sharing and using open data to drive innovation. This set of topics covers different domains such as education, data science, software engineering, computing technologies, etc., reflecting some of the key issues and trends in today's technology and education environments.

4.3.3 Periodic Topic Comparison

Based on the global topic clustering results, the prospect dataset was sliced every three years as a time slice, and each topic group was set to 12 to obtain four groups of map with topic details（Figure 13-15）.2014-2016 A total of 1745 literatures were clustered into 3 Cluster, with the largest contribution from topic marketing networks literatures, which form an integrated marketing foundation covering various aspects of prospecting, marketing communication, customer satisfaction, and concepts related to marketing and strategy.2017-2019 A total of 2408 literatures were clustered into 4 Clusters, where researchers understand

and respond to the complexity of the prospecting market more comprehensively for more effective decision making and strategic planning at different levels.2020-2022 A total of 2581 literatures were clustered into 4Clusters in 2020-2022, with Topic Purchase Behaviour Modeis and Topic Community Mental Health being the most significant.2023 A total of 708 literatures were clustered into 4Clusters, with Topic Media Marketing Research being the most significant.

Figure 13

2014-2016Topic Relationships and Descriptions

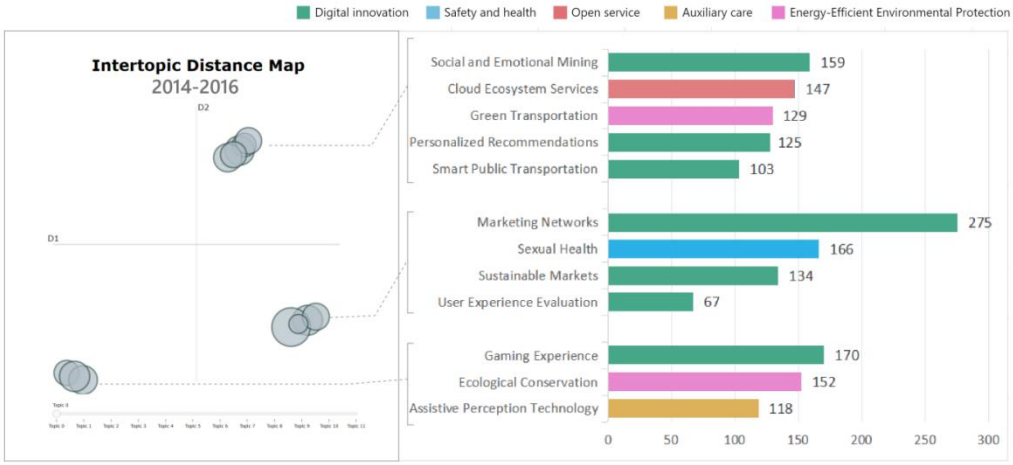


Figure 14

2017-2019Topic Relationships and Descriptions

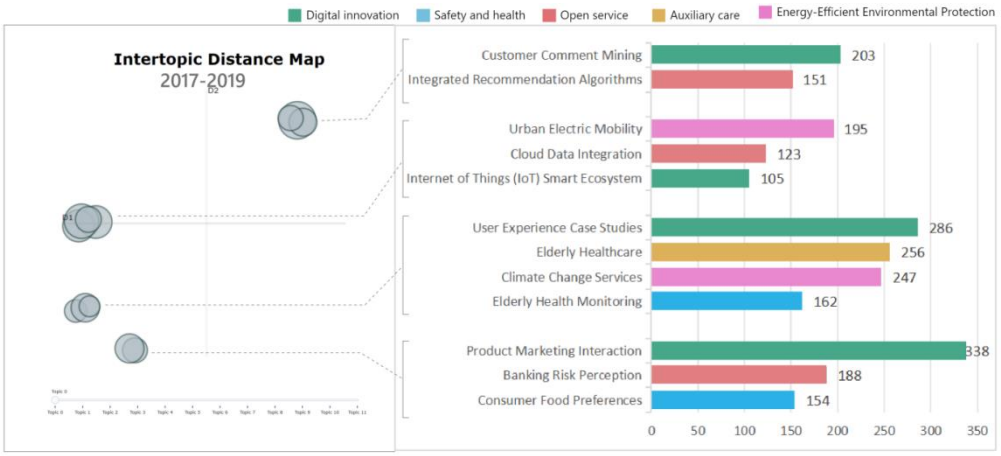


Figure 15

2020-2022Topic Relationships and Descriptions

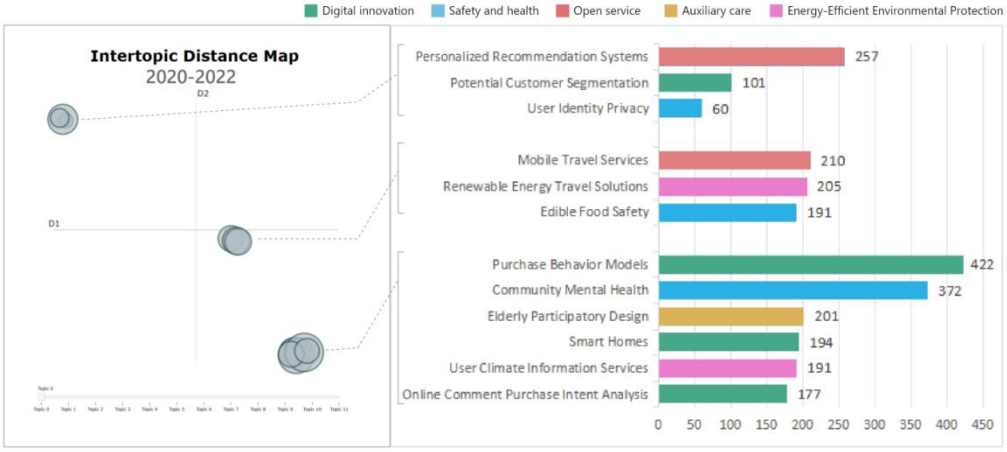
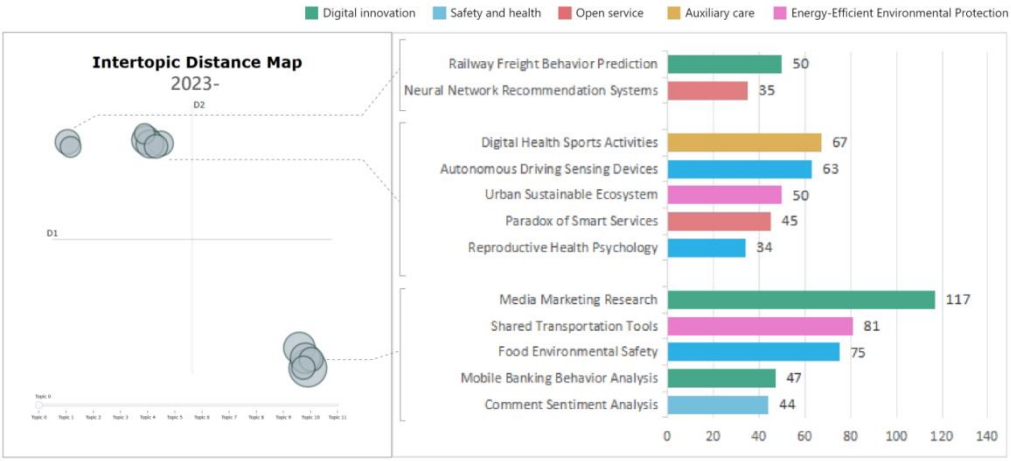


Figure 16

2023-Topic Relationships and Descriptions



4.3.4 Periodic Topic Evolutionary Trends

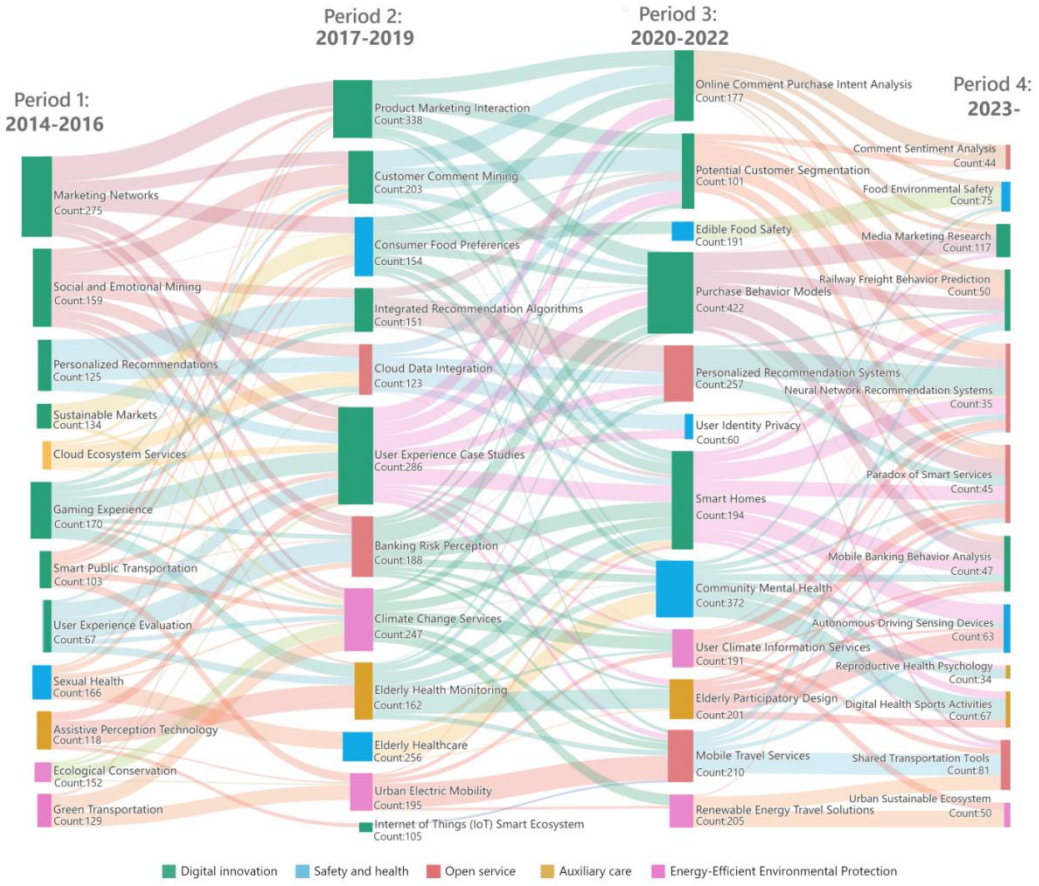
The topic words under each topic are converted into vector representations using a word2vec model, then aggregated for averaging, and the cosine similarity between the topics of adjacent phases is calculated [39].This is done to plot the Sangi diagram using Python's pyecharts library. Based on this, it is possible to see which themes gradually shrink over time, which themes continue to be of interest, and which themes emerge at a given time. The research themes for the four periods from 2014-2023 are shown in Figure 17. Where blocks represent themes, the flow graphs between the blocks represent the evolutionary path of these themes over time. The thickness of the blocks indicates the number of documents between themes; and

the height of the blocks indicates the strength of the flow of themes. The analysis revealed the following characteristics of the evolutionary trend of research themes across the datasets:

(1) Digital innovation has always been a hotspot for potential user research concerns, showing a certain expansion trend and overwhelming presence of flow intensity. Through observation, the extension from Marketing Networks to Product Marketing Interaction[40], Consumer Food Preferences, Customer Comment Mining, development to Online Comment Purchase Intent Analysis, Purchase Behaviour Models, evolving into Media Marketing Research, Comment Sentiment Analysis in 2023, and continuing to the present day towards fine-grained knowledge of potential users. Gaming Experience has evolved into User Experience Case Studies and started to expand and transform into Parados of Smart Services through Smart Homes, highlighting the development of cross-domain experience services.

Figure 17

Thematic Evolution



(2) Safety and health, in Period 1 Sexual Health concern shifted to Period 2 Elderly Healthcare, Consumer Food Preferences, and then developed into Community Mental Health,

indicating the increasing importance of mental health in society. The emergence of Consumer Food Preferences in Period 2, which evolved into Period 3 Edible Food Safety and Period 4 Food Environmental Safety, shows a deepening and broadening of the understanding of food safety issues, from individual-level consumer preferences to a more comprehensive and systemic approach. Period 3 User Identity Privacy theme cooled significantly, and Period 4 Autonomous Driving Sensing Devices appeared, suggesting the importance of autonomous driving technology in potential user research, as well as the concern for future mobility and technology trends. Concerns. Influenced by a variety of factors such as social structure, demographics, and medical advances, this reflects the evolution of society's perception of health and safety travel issues.

Energy-Efficient Environmental Protection, in the evolution from Green Transportation to Urban Electric Mobility and Climate Change Services, the number of literatures grows significantly. Since Period 2, the intensity of Energy-Efficient Environmental Protection flow has significantly increased, and although there is still a large gap compared to Digital innovation, it has gained steady growth on its own. The 2020-2022 epidemic has exposed the close relationship between human health and the environment. The demand for environmental protection and energy efficiency has become more pronounced among some potential users.

(4) Auxiliary care in Period 1-3 reflects a lower level of concern, whether it is Assistive Perception Technology in Period 1, Elderly Health Monitoring in Period 2, and Elderly Participatory Design in Period 3. Elderly Participatory Design, the increase in flow intensity also reflects the expectations and needs of society and individuals for auxiliary care. Auxiliary care requires a combination of technology and humane care to assist users in their daily activities, provide companionship, and even provide emergency services for auxiliary care applications, cost-effective and easy-to-implement solutions. Cost-effective and easy-to-implement solutions that require a high level of attention cannot be attracted to a wide range of highly valued.

Open service in Period2 research rise, with the development of big data, artificial intelligence attention has been significantly increased. It was found that Personalised Recommendation Systems evolved into Neural Network Recommendation Systems, Mobile Travel Services evolved into Shared Transportation Tools. collated to see that the The advancement of Open service technology promoted the evolution of the service model, and the research focused on the promotion of information technology in various industrial fields, after which the focus was placed more on the application of emerging technologies such as big data applications, data management, blockchain, artificial intelligence[41] and other emerging

technologies in various fields, which are closely related to the development of information technology and the changes in social needs.

The 10-year development of research themes of potential users is roughly as follows: in 2014-2016, the research focuses on user-driven innovation, open user experience, social mining, intelligent ecology, potential market development and other industries to open up the market; in 2017-2019, the research on Energy-Efficient In 2017-2019, research on Energy-Efficient, Environmental Protection, Open service, etc. is gradually increasing; in 2020-2022, "Purchase Intention", "Personalised Recommendation Service", etc. focusing on precise service will appear. In 2020-2022, topics focusing on precision services such as "purchase intention" and "personalised recommendation services" will emerge; in recent years, "connected services for potential customers", "intelligent service platforms", "digital sharing" and "open experiences" have been increasing. In recent years, there has been an emergence of research on "connected services for potential customers", "intelligent service platforms", "digital sharing" and "open experience". Overall, the scope of research continues to broaden and evolve from the initial multi-dimensional applications to focus on open services, and more towards the public service sector. This has led to advances in sensory experiences such as natural language processing, computer vision, speech recognition, etc., but it has also brought new challenges to computing resources and user data privacy.

5 CONCLUSION

The conclusion of an article should summarize the main findings of the study succinctly, highlighting the significant contributions to the research field. It should reiterate the objectives of the study and summarize the most important findings, emphasizing their relevance and practical or theoretical implications.The number of annual publications on prospective user research has generally levelled off over the past decade but can be expected to have declined in recent years. The United States is the most prolific country, and University Washington has produced important research. computer science information systems are the most influential discipline, and LI Y has published the most articles, which has contributed to the further development of potential user research. Given the complexity of potential user research, there is a close transnational relationship between countries, institutions and authors, and there is an endless hierarchy of hotspot emergence, which promotes innovation and challenges. In addition to this, this study utilises the BERTopic model to mine the overall prevalent themes of potential users and explore the historical theme evolution to inform user research conducted by

companies and phase researchers. This study found that the overall prevalent themes of potential user research cover 42 topics, and after dimensionality reduction based on UMAP themes, it is found that most of the themes focus on "Digital innovation", "Safety and health", "Auxiliary care" and "Open source", which are the most popular themes in user research. Auxiliary care" "Open service" "Energy-Efficient Environmental Protection". "Energy-Efficient Environmental Protection" and "Digital innovation", "Safety and health", "Auxiliary care", "Open service" and "Energy-Efficient Environmental Protection". Historical evolutionary trends are obtained through word2vec, and it is found that the trend of purchasing behaviour sensory data mining, public service domain and centralized open service development, while bringing new challenges of computing resources and security risks of user data privacy. Our results bring new ideas for potential user-related research and help further research on applications and technologies .

The research recommendations are provided as follows:

(1) Development of social services in relation to national conditions

Under the guidance of the current national development strategy, the direction of research by potential users is in line with the national strategy. The research results should be closer to the needs of society, be able to solve practical problems and improve the overall level of social services in the country. For example, simplify the digital service process to make it more popular and convenient, to achieve the actual benefits of social services; by focusing on the digital transformation of small and medium-sized enterprises resource issues, to provide industry intelligent service solutions research.

(2) Strengthen the foundation to give full play to the advantages of the discipline

Potential user research can be developed in depth on the existing foundation. For example, the advantages of traditional disciplines such as economics, computer science, and psychology can be combined to promote in-depth exploration of potential user research in the areas of digital technology and the psychology of user behaviour, such as combining the data analysis methods of computer science to study the relationship between user demand and supply for specific products or services. At the same time, user experience research methods from psychology are applied to gain a deeper understanding of users' emotions, attitudes, and behaviours. This comprehensive research approach can provide a more comprehensive insight into the dynamics of potential users and can better support the sustainability of social services and research hotspots.

(3) Enhancing the life-like sensory experience

Sensory experience is an important consideration in various areas of design, product development and user experience research. Aspects such as visual perception, audio prompts and haptic feedback of automated driving systems can improve the user's sense of confidence and safety, thus enhancing the overall user experience. Potential user sensory experience research can better meet user needs, make technology more intelligent and closer to users' actual feelings, improve user satisfaction, and gain an edge in market competition.

(4) Focus on computing resources and data security

With the depth of potential user research, researchers need more computing resources to deal with large-scale data, user privacy and security protection inevitable choice. Decentralise computing resources, reduce dependence on a single cloud service provider, consider edge computing applications, focus on quantum computing, introduce encryption technology, access control policy applications, etc., to ensure the security of data involved in potential user research.

REFERENCES

Slater, S. F., & Narver, J. C. (1998). Customer-led and market-oriented: let's not confuse the two. Strategic management journal, 19(10), 1001-1006.

Howard, J. A., & Sheth, J. N. (1969). The theory of buyer behavior. New York, 63, 145.

Read, S., Dew, N., Sarasvathy, S. D., Song, M., & Wiltbank, R. (2009). Marketing under uncertainty: The logic of an effectual approach. Journal of marketing, 73(3), 1-18.

DemandGen Report， 2021， 2021 Lead Nurturing & Acceleration Benchmark Survey: Moving Into The Modern Era With Highly Targeted, Account-Based LeadNurturing,DemandGenReport ,https://www.demandgenreport.com/resources/2021-lead-nurturing-acceleration-benchmark-survey-moving-into-the-modern-era-with-highly-targeted-account-based-lead-nurturing/6963/

Brandon Purcell， 2023， Make Customer Lifetime Value Your Polaris For Long-TermGrowth,ForresterResearch ,https://www.forrester.com/report/make-customer-lifetime-value-your-polaris-for-long-term-growth/RES150315?ref_search=0_1712477309254

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. Scientometrics, 52, 291-314.

Prahalad, C. K., & Ramaswamy, V. (2004). The future of competition: Co-creating unique value with customers. Harvard Business Press.

Rogers, D. L. (2016). The digital transformation playbook: Rethink your business for the digital age. Columbia University Press.

Thelwall, M. (2008). Bibliometrics to webometrics. Journal of information science, 34(4), 605-621.

Thompson, D. F., & Walker, C. K. (2015). A descriptive and historical review of bibliometrics with applications to medical sciences. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 35(6), 551-559.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Bai, Y., Li, H., & Liu, Y. (2021). Visualizing research trends and research theme evolution in E-learning field: 1999–2018. Scientometrics, 126, 1389-1414.

Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ebrahim, N. A. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. arXiv preprint arXiv:1305.0377.

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. Publications, 9(1), 12.

Li, W., & Zhao, Y. (2015). Bibliometric analysis of global environmental assessment research in a 20-year period. Environmental Impact Assessment Review, 50, 158-166.

Park, H. W., Yoon, J., & Leydesdorff, L. (2016). The normalization of co-authorship networks in the bibliometric evaluation: the government stimulation programs of China and Korea. Scientometrics, 109, 1017-1036.

Meng, L., Wen, K. H., Brewin, R., & Wu, Q. (2020). Knowledge atlas on the relationship between urban street space and residents' health—a bibliometric analysis based on VOSviewer and CiteSpace. Sustainability, 12(6), 2384.

Mazzei, D., & Ramjattan, R. (2022). Machine Learning for Industry 4.0: A Systematic Review Using Deep Learning-Based Topic Modelling. Sensors, 22(22), 8641.

Smith, B. P., Hoots, B., DePadilla, L., Roehler, D. R., Holland, K. M., Bowen, D. A., & Sumner, S. A. (2023). Using Transformer-Based Topic Modeling to Examine Discussions of Delta-8 Tetrahydrocannabinol: Content Analysis. Journal of Medical Internet Research, 25(1), e49469.

Church, K. W. (2017). Word2Vec. Natural Language Engineering, 23(1), 155-162.

Di Gennaro, G., Buonanno, A., & Palmieri, F. A. (2021). Considerations about learning Word2Vec. The Journal of Supercomputing, 1-16.

Wang, H., Tong, L., Takeuchi, R., & George, G. (2016). Corporate social responsibility: An overview and new research directions: Thematic issue on corporate social responsibility. Academy of Management journal, 59(2), 534-544.

Yu, Q., Shao, H., He, P., & Duan, Z. (2013). World scientific collaboration in coronary heart disease research. International journal of cardiology, 167(3), 631-639.

Han, P., Shi, J., Li, X., Wang, D., Shen, S., & Su, X. (2014). International collaboration in LIS: global trends and networks at the country and institution level. Scientometrics, 98, 53-72.

Deng, X., Zhai, J., Lv, T., & Yin, L. (2017). Efficient vector influence clustering coefficient based directed community detection method. IEEE Access, 5, 17106-17116.

Sun, W., Huang, P., Song, H., & Feng, D. (2020). Bibliometric analysis of acute pancreatitis in Web of Science database based on CiteSpace software. Medicine, 99(49).

Elazhary, H. (2019). Internet of Things (IoT), mobile cloud, cloudlet, mobile IoT, IoT cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions. Journal of network and computer applications, 128, 105-140.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for information Science, 24(4), 265-269.

Osareh, F. (1996). Bibliometrics, citation analysis and co-citation analysis: A review of literature I.

Pouder, R., & St. John, C. H. (1996). Hot spots and blind spots: Geographical clusters of firms and innovation. Academy of management review, 21(4), 1192-1225.

Bornmann, L., Haunschild, R., & Hug, S. E. (2018). Visualizing the context of citations referencing papers published by Eugene Garfield: A new type of keyword co-occurrence analysis. Scientometrics, 114, 427-437.

Gottschalk, S., & Demidova, E. (2019). EventKG–the hub of event knowledge on the web–and biographical timeline generation. Semantic Web, 10(6), 1039-1070.

Selvin, E., Parrinello, C. M., Sacks, D. B., & Coresh, J. (2014). Trends in prevalence and control of diabetes in the United States, 1988–1994 and 1999–2010. Annals of internal medicine, 160(8), 517-525.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Frontiers in sociology, 7, 886498.

Weißer, T., Saßmannshausen, T., Ohrndorf, D., Burggräf, P., & Wagner, J. (2020). A clustering approach for topic filtering within systematic literature reviews. MethodsX, 7, 100831.

Matheron, G., Perrin, N., & Sigaud, O. (2020, September). PBCS: Efficient exploration and exploitation using a synergy between reinforcement learning and motion planning. In International Conference on Artificial Neural Networks (pp. 295-307). Cham: Springer International

Rong, X. (2014). word2vec parameter learning explained. arXiv preprint arXiv:1411.2738.

Jang, B., Kim, I., & Kim, J. W. (2019). Word2vec convolutional neural networks for classification of news articles and tweets. PloS one, 14(8), e0220976.

Shinde, A. V., Patil, D. D., & Tripathi, K. K. (2024). A Comprehensive Survey on Recommender Systems Techniques and Challenges in Big Data Analytics with IOT Applications. Revista De Gestão Social E Ambiental, 18(2), e05195. https://doi.org/10.24857/rgsa.v18n2-097

Almustafa, E., Assaf, A., & Allahham, M. (2023). Implementation of Artificial Intelligence for Financial Process Innovation of Commercial Banks. Revista De Gestão Social E Ambiental, 17(9), e04119.