A Review of Scholar Profile Research: Postprint

Authors: Wang Shiqi, Liu Zhifeng, Wang Jimin, Wang Jimin

Date: 2022-11-26T00:00:00+00:00

Abstract

This study systematically reviews scholar profiling research to provide references for related investigations. Through literature investigation and analysis, it distinguishes and examines the concepts of scholar profiling and related notions, summarizes the construction process, key technologies, and primary applications of scholar profiling, and analyzes the current challenges facing this research. The construction process of scholar profiles comprises data collection, data preprocessing, scholar tag construction, and visual analysis; the main practical applications include expert recommendation, academic resource recommendation, and research capability evaluation. Current research faces challenges such as the difficulty of acquiring and fusing multi-source data, the complexity of dynamically updating scholar profiles, and the lack of effective evaluation mechanisms.

Full Text

A Review of Scholar Profiling Research

Wang Shiqi[1], **Liu Zhifeng**[1], **Wang Jimin**[1,2] [1]Department of Information Management, Peking University, Beijing 100871 [2]National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing 100871

Abstract: [Purpose/Significance] This paper systematically reviews research on scholar profiling to provide a reference for related studies. [Method/Process] Through literature research and analysis, this paper distinguishes scholar profiling from related concepts, summarizes the construction process, key technologies, and main applications of scholar profiling, and analyzes current research challenges. [Result/Conclusion] The scholar profiling construction process includes data collection, data preprocessing, scholar label construction, and visual analysis. Primary applications include expert recommendation, academic resource recommendation, and research capability evaluation. Current research faces challenges such as difficulty in multi-source data acquisition and fusion, challenges in dynamic updating of scholar profiles, and lack of effective evaluation mechanisms.

Keywords: user profile; scholar profile; academic data; expert recommendation
Classification Number: G250 **DOI:** 10.13266/j.issn.0252-3116.2022.20.008

In the era of big data, user profiling—enriched by abundant data resources and broad application prospects—has gradually become a research hotspot in library and information science. User profiling aims to be user-centered, extracting multi-dimensional user information from multi-source data to generate a series of labels that characterize user features. As a data service tool for depicting user characteristics and mining user needs, user profiling has been widely applied in libraries, e-commerce, social media, and smart healthcare [1-4]. With the rapid growth and accumulation of academic big data, managing and analyzing multi-source heterogeneous academic data has become a new challenge. Scholar profiling is a method that organizes and manages academic big data with researchers as the basic unit. Unlike general user profiling, scholar profiling targets researchers and aims to outline a comprehensive picture of scholars from massive academic data, offering broad application prospects in academic behavior analysis, scholar recommendation, and scholar evaluation.

The focus of scholar profiling is to characterize the attribute features of different scholars, which differs from general user profiling in research objects, data sources, technical methods, and application scenarios. Current research on scholar profiling primarily concentrates on its concept [5], construction process, and application studies. Research on construction methods includes data source identification, scholar profile dimension design [6], multi-source data fusion techniques [7-9], automatic extraction of scholar labels [8], and visualization of scholar profiles. Application studies based on scholar profiling include research collaborator recommendation [6] and discovery of scholars' research interests [10]. These studies demonstrate that scholar profiling research has achieved certain progress.

Some domestic scholars have summarized relevant research on user profiling concepts, methods, technologies, models, and applications from macro and micro perspectives [11-14]. However, regarding scholar profiling reviews, only Yuan Sha et al. [15] in 2018 provided a technical overview, identified problems, and discussed future directions from a computer science perspective, lacking a summary of model construction and main applications. In recent years, with the development of scholar profiling research, scholars have further explored the field from perspectives of data sources, model construction, technical methods, and application domains, proposing new scholar profiling models and application directions. Meanwhile, some controversies remain regarding the conceptual origins of scholar profiling. Therefore, this study first sorts out the concepts and construction process of scholar profiling, summarizes its main application directions and challenges, and prospects future research trends to provide references for related research.

1 Conceptual Analysis of Scholar Profiling and User Profiling

Some studies suggest that the concept of user profiling was first proposed by interaction design pioneer Alan Cooper in his 1999 book *The Inmates Are Running the Asylum* [7,12]. However, Cooper did not directly mention "user profiling" in the book. The relevant original text states that "A persona is a fictitious, specific and concrete representation of target users" [16]. The term "persona" mentioned here refers to using a fictional yet specific user to represent target users, thus it is usually translated as "typical user" or "user role." Persona is not the same concept as the precise user profiles constructed using data in many studies, so it is inaccurate to regard Cooper's work as the origin of user profiling.

Another concept related to the origin of user profiling is "user profile." Cooper explicitly defined persona and user profile as two completely different design tools in his 2007 book *About Face 3: The Essentials of Interaction Design* [17]. Baxter et al. defined user profile as a detailed description of user attributes, such as job title, experience, education level, key tasks, age range, etc., representing a non-singular, exhaustive feature set of users [18]. The user profile concept is commonly used in computer science fields such as information filtering, tag recommendation systems [19], and personalized document retrieval [20]. Overall, both in definition and application scenarios, user profile better aligns with the definition of "user profiling" used in most domestic research: extracting user information from massive data and constructing a set of user labels [21].

Scholar profiling is an important branch of user profiling research. The research object of scholar profiling is researchers. As early as 2007, L. Yao et al. first mentioned that "unlike traditional user profiles established through manual input, they focus on how to automatically construct profiles for researchers using a unified approach by integrating dependency relationships between different types of information" [22]. Later, Fan Xiaoyu et al. defined researcher profiles as constructing labeled user models based on researchers' social attributes, research habits, and behaviors [23]. Yuan Sha et al. believe that scholar profiling is a process of automatically obtaining multi-dimensional information for constructing researcher user models from the open Internet through computer technology, thereby conducting data mining and application analysis to support specific applications such as expert selection and recommendation [15]. Unlike general user profiling, Qin Chenglei et al. emphasize that the core of scholar profiling is based on characteristics such as high data authority, large data scale, and easy data acquisition. Compared with general user profiling, scholar profiling is more accurate and involves deeper analysis.

2 Scholar Profiling Construction Process and Key Technologies

Scholar profiling construction methods and technologies are the research focus of this field. Currently, many studies have elaborated on the user profiling construction process from different perspectives and levels. Such research [25-26] often summarizes the user profiling construction process into steps such as collecting user feature data, extracting user feature information, and building user profile models. Scholar profiling differs from general user profiling in data sources, label system design, and applications. Wang Ruijie summarized the scholar profiling construction process into a data information layer, data processing layer, and profile analysis layer, where information in the profile analysis layer is divided into basic and advanced dimensions [27]. Chi Xuehua' s research divided profile construction into three parts: personal information description, research interest label discovery, and academic influence prediction, and summarized the key technologies of information extraction and text classification involved [28]. Fan Xiaoyu et al. conducted scholar profiling visualization research after completing model representation and label extraction [23]. Based on the above scholar profiling studies, the research framework for big data-based scholar profiling construction mainly includes four components: multi-source scholar data collection, data preprocessing, scholar label construction, and scholar profiling visualization [Figure 1: see original paper].

2.1 Data Sources and Collection Methods

In the era of academic big data, large academic databases, academic search engines, and academic social media platforms contain massive amounts of scholar-related achievement and behavioral data, providing a solid data foundation for scholar profiling. Scholar profiling data includes different types such as scholar personal data, research achievement data, and research social data. The data sources for these three categories are shown in Table 1 .

Scholar personal data refers to scholars' demographic characteristics, mainly including name, education, affiliation, title, and other data. Such data primarily comes from online websites and academic search engines such as scholars' personal homepages, Baidu Baike, Wikipedia, and the Aminer platform [29], where Aminer is a technology intelligence analysis and mining platform based on three types of data: researchers, scientific literature, and academic activities, currently containing 84 million researcher data. Personal data sources are extensive with inconsistent data structures, posing significant challenges to data collection. Current mainstream methods for obtaining scholar personal data mainly include two categories: one uses rule-based methods to extract required information such as name, affiliation, and address from search engine results or scholar homepages [28]; the other adopts machine learning algorithms such as CRF and LSTM sequence labeling models for entity recognition from collected homepage content [30]. However, existing machine learning-based automatic scholar homepage recognition technology cannot achieve high recognition rates, and a considerable proportion of experts and scholars still lack homepages or cannot be found through search engines [22].

Research achievement data is an important data source for constructing scholar profiles, mainly including academic papers, monographs, research projects, and patents. Research achievement data primarily exists in various academic information databases. Chinese data comes from the three major domestic electronic literature databases—CNKI, Wanfang, and Weipu [31], while foreign data comes from Web of Science and other academic databases [32]. These databases generally provide data export services in common formats, though web scraping can also be used for more comprehensive and flexible data capture. Currently, many digital libraries and academic search engines on the Internet, such as DBLP (Digital Bibliography & Library Project), CiteSeerX, ACM, and Google Scholar, can also serve as supplements to research achievement data. Additionally, there are databases storing project information for the National Natural Science Foundation and National Social Science Foundation, as well as the China Patent Database for patent data.

Academic research social data is generated when researchers use academic social websites, including follow, interaction, and comment data. Academic social networking sites are platforms that help researchers conduct academic research-related communication and collaboration via the Internet. Unlike general social networking sites, academic social networking sites aim at academic exchange and cooperation. Currently, well-known foreign academic social platforms include ResearchGate [33] and Academia [34], while domestic platforms include ScholarNet [35] and XiaoMuChong. Additionally, co-authorship relationships in research achievements can serve as effective supplements to scholar social data.

Finally, data from different sources needs to be integrated and uniformly stored in databases for management. Different sources use varying description standards for experts and scholars, so it is first necessary to construct a unified metadata field table to map data from different sources to unified metadata fields, achieving multi-source data fusion. The biggest challenge in data fusion is the scholar entity disambiguation problem. Some research has explored this from perspectives such as semantic similarity [40], paper feature model similarity [41], and combined rule and feature model similarity [42]. However, existing scholar entity disambiguation methods still have room for improvement in accuracy, and future deep learning methods could further enhance disambiguation effects [43].

2.2 Data Preprocessing and Key Technologies

Scholar profiling data is multi-source and heterogeneous, requiring raw data to be cleaned and effectively fused to transform it into data suitable for constructing expert scholar profile models. Scholar profiling data preprocessing generally includes three steps: data cleaning, information extraction, and data fusion. In recent years, scholar profiling research has seen new explorations in data preprocessing. This paper summarizes the methods and technologies in preprocessing steps and some latest tools from a practical application perspective.

First, raw data needs to be cleaned to address data missing and redundancy issues. For missing data, supplementary approaches such as search engine retrieval or adding other data sources can be used. For example, L. Yao et al. extracted personal information from Wikipedia for researchers without scholar homepages [22]; M. Bravo et al. chose to extract publication data from DBLP to enrich original university researcher data [36]. For data redundancy, duplicate data needs to be deleted to ensure data uniqueness.

Second, required fields need to be extracted from text data. In scholar personal homepages, basic personal information is usually unstructured text data, requiring information extraction from unstructured text. The key technology involved is Named Entity Recognition (NER). NER technology is an important foundation for information extraction, referring to extracting specified types of entities from natural language text. In the scholar profiling field, when processing scholar personal data, the main entities extracted include name, position, institution, email, etc. NER methods mainly include rule-based learning methods, statistical machine learning methods, and deep learning methods. Early NER work mostly used manually compiled dictionaries and rules, which had high accuracy but low recall, were time-consuming, language-dependent, and had poor scalability. Statistical machine learning methods treat NER as a classification or sequence labeling problem [37], using manually annotated corpus to train classical machine learning classifiers such as HMM, ME, CRF, and SVM. The difficulty of such methods lies in requiring large-scale training corpus and how to construct feature engineering. Currently, deep learning methods are commonly used for entity recognition from unstructured text such as scholar homepages. This method uses word vectors to represent words and character vectors to represent characters, utilizing deep neural networks to solve the feature engineering construction problem of statistical machine learning methods and achieving good results. Commonly used NER tools include the Chinese Academy of Sciences' NLPIR system [38], Stanford' s Stanza, and Harbin Institute of Technology' s LTP system [39].

Finally, data from different sources needs to be integrated and uniformly managed. The biggest challenge in data fusion is the scholar entity disambiguation problem, which has been explored from perspectives such as semantic similarity [40], paper feature model similarity [41], and combined rule and feature model similarity [42]. However, existing methods still have room for improvement in accuracy, and future deep learning methods could further enhance entity disambiguation effects [43].

2.3 Scholar Label Construction

Scholar label construction is the process of mining scholar data, extracting scholar features, and identifying these features with standardized phrases.

Scholar labels have characteristics such as standardization, short text, semanticization, and specificity [44]. Scholar profiling labels can be divided into four categories: personal attribute labels, research interest attribute labels, academic capability labels, and scholar social labels.

Scholar personal attribute labels describe basic personal features, such as name, age, title, and educational background. Generally, personal attribute labels are relatively stable and do not change significantly in the short term. The source of personal attribute labels is mainly information extraction from scholar personal data. Additionally, some personal attribute labels can be artificially constructed according to specific tasks, such as academic age and institutional geographic distance [6].

Research interest attribute labels are the core of the scholar label system and are commonly used in expert recommendation scenarios. The number and content of research interest labels have no unified standard and are often constructed based on different application scenarios. Research interest labels mainly come from scholars' research achievement data. The most direct method is to use keywords from scholars' published papers as interest labels. However, the number of keywords is large and their quality varies, and some keywords cannot illustrate scholars' interest topics. Therefore, some scholars use algorithms such as LSI [45], LDA, and Doc2Vec [46] to mine scholars' research interest labels from relevant research achievements. Some scholars also extract terms representing scholars' research interests from different online data sources and integrate relevant terms representing research interests through Wikipedia to represent scholars' research interests [47]. Additionally, Shi Xiang et al. [48] found through reviewing literature on scholar interest identification that current research on scholar interest identification at the vocabulary-topic level is relatively mature, and future research directions mainly focus on network-level interest identification.

Scholar academic capability labels are also an important component of the scholar profiling label system. Scholars' academic capabilities can be reflected from two aspects: the quality of academic achievements and academic influence. The quality of academic achievements can be represented by indicators such as the level of published journals, the level of hosted funds, and the authority of affiliated institutions. Scholars' academic influence is commonly represented by derived indices such as the h-index and g-index [49-50].

Scholar relationship labels mainly come from research social data and research achievement data. Research social data is generated when researchers use academic social platforms. Common academic social websites include foreign platforms such as ResearchGate, Academia.edu, and domestic platforms such as Research Friend and ScholarNet. Such websites provide social data including followed scholars, questions, and answers. Additionally, scholars' cooperative relationship networks can be obtained from co-authorship relationships in research achievements such as papers.

2.4 Scholar Profiling Visualization Analysis

Scholar profiling visualization refers to presenting the previously constructed large number of different types and weighted scholar profiling labels in graphical forms, which can intuitively and clearly display and analyze various attributes of experts and scholars. Among them, building tag clouds for user information is a commonly used visualization method. This method can build word clouds with different-sized tags according to different label weights, visually presenting user information. Many mature tools are now available for user tag visualization, such as WordArt, tagCloud, Tagul [51], and Tagxedo [52]. Additionally, visualization methods can be used to analyze scholars' rich research achievement data. Some scholars have summarized dozens of existing data visualization tools, techniques, and systems for analyzing academic data. For example, social network visualization tools such as Pajek and Gephi can be used to visualize and analyze scholars' cooperative, citation, or follow relationships and community division. Common statistical charts such as bar charts and radar charts can also be used to display scholars' research achievement publication timelines and research interests [53].

3 Application Research on Scholar Profiling

3.1 Expert Recommendation

Research interest is one of the most important attribute features of scholars. Through research interest labels in scholar profiles and similarity calculation algorithms, expert and research collaborator recommendations in related fields can be achieved. In expert recommendation, R. Thiagarajan et al. solved the expert discovery problem by using ontology-based spreading activation networks to calculate similarity between user profiles [54]. Hu Chengfang et al. [55] proposed a design concept for an expert database system in the Lancang-Mekong water resources cooperation field based on profiling technology, achieving intelligent talent recommendation based on Lancang-Mekong cooperation needs by constructing a talent profiling model based on spatiotemporal attributes. L. M. De Campos et al. [56] built summary user profiles and extracted different hidden topics of interest to experts through clustering expert text information, based on which they implemented expert recommendation applications.

Additionally, research collaborator recommendation based on academic preferences is another important application of scholar profiling. In recent years, many studies on research collaborator recommendation based on scholar profiling have emerged. These studies mainly collect scholars' academic interests, academic capabilities, and cooperative networks [57] from literature, patents, and social media to build scholar profiles, and then recommend research collaborators by fusing similarity across multiple dimensions of scholars.

3.2 Academic Resource Recommendation

Academic resource information recommendation is an important research issue in the library and information field. Recommending academic resource information based on scholar profiles is currently one of the important means. Such recommendations have many applications in the library field, where books are recommended by matching scholars'interest and behavior features from their profiles with book content. Some scholars have built individual and group profiles of readers and achieved personalized book recommendations by comprehensively reflecting readers' borrowing behavior characteristics [58]. Subsequent studies have attempted to integrate richer Internet data. For example, Liu Haiou et al. [59] proposed personalized learning resource recommendation services based on scholar profiling in the big data era, providing dynamic personalized resource recommendations for researchers by combining their basic information, research interests, and social interaction data. With the continuous rise of academic social media and the increasing richness of academic social data, research on building academic new media user profiles based on academic social data and precise recommendation models for academic new media information based on user profiles has also made certain progress [60].

3.3 Scholar Research Capability Evaluation

Scholar profiling technology has also been applied to evaluate scholars' past academic capabilities and predict their future potential capabilities. Scholars' academic capabilities are primary indicators considered by higher education institutions, research centers, and enterprises when making talent recruitment and funding decisions, making objective and accurate evaluation crucial. For example, Han Xu et al. proposed a method for calculating scholar capability indices and ranking scholars based on user profiling technology, using the computer field as an example [61]. Xiong Huixiang et al. divided research capabilities into academic achievement quality and academic influence, using a weighted topic model to represent scholars' research capabilities [62]. M. Lee et al. [63] proposed researcher performance index models that can measure qualitative and quantitative performance, researcher influence, and growth potential, evaluating researchers from multiple perspectives to help them improve their research capabilities. As an important dimension of scholar profiling, academic capabilities can be evaluated through academic capability labels in scholar profiles.

4 Current Research Challenges

A body of research has explored scholar profiling from perspectives of concepts, model construction, key technologies, and applications, providing a theoretical and practical foundation for subsequent research. With the rise of the open science movement, multi-source heterogeneous scientific big data provides a rich data foundation for scholar profiling while also bringing greater challenges. The challenges in scholar profiling research mainly include difficulty in multi-source data acquisition and fusion, challenges in dynamic updating research, and lack of evaluation studies.

4.1 Difficulty in Multi-Source Data Acquisition and Fusion

The development of the Internet has generated massive amounts of data, making it possible to build precise and rich scholar profiles based on multi-source heterogeneous data. However, how to filter out required scholar webpages from massive Internet pages and dynamically capture and store them in real-time is one of the challenges facing scholar profiling. The current common method is to first search using scholar names and institutions in search engines, then filter results according to rules. However, this method has low recall and precision rates [64], and identifying required scholar personal homepages remains a research challenge. Additionally, existing research shows that a considerable proportion of scholars have no homepage or cannot be found through crawlers or search engines [22]. For these scholars, personal information needs to be obtained from other sources, such as academic publications or social platforms.

Data fusion is another challenge in scholar profiling. In the process of multi-source heterogeneous data fusion, issues such as scholar name ambiguity require entity disambiguation techniques. Current data fusion research mainly adopts methods based on machine learning, graphs, and heuristic rules, with machine learning algorithms being predominant [30]. Overall, data acquisition and fusion technologies in big data-based scholar profiling research have made certain progress, but there remains much exploration space in solving time complexity for large-scale data and incremental disambiguation [65].

4.2 Challenges in Dynamic Updating of Scholar Profiles

Scholars' relevant information is constantly changing. To ensure the accuracy and timeliness of scholar profiles, dynamically updating attribute features across different dimensions is crucial. However, most current scholar profiling research ignores the issue of dynamic updating. Possible solutions for dynamic updating of scholar profiles include feedback mechanism-based updating and data-driven updating. Feedback mechanism-based scholar profile updating generally uses manual methods to modify and improve scholars' personal attribute information. For example, the AMiner scholar profiling system uses a "claim-edit" approach to encourage scholars themselves to manually modify and supplement profile information [10]. This approach can ensure the accuracy of updated information but is too inefficient for large-scale scholar profiles.

Data-driven scholar profile updating generally uses automated methods to update scholars' interest attributes, academic capability attributes, and other information. This automated updating technology based on research achievement data needs to solve problems such as real-time data collection and building efficient trigger and update mechanisms [15]. Overall, research on dynamic updating of scholar profiles requires substantial human investment, has certain technical thresholds, and requires long research cycles. Therefore, current research on dynamic updating of scholar profiles is relatively scarce.

4.3 Lack of Scholar Profiling Evaluation Studies

Most current research on scholar profiling focuses on how to construct scholar profiles, innovating processes such as multi-source heterogeneous data fusion, information extraction, label organization, and weight calculation. However, few studies scientifically and objectively evaluate constructed scholar profile models from aspects such as authenticity, timeliness, and accuracy. Only a small number of studies provide theoretical reviews of constructed models [66] or illustrate scholar profiling applications in certain fields with subjective evaluations [6], lacking evaluation studies based on large-scale data or scientifically objective test samples. This affects the universality of scholar profiling models and hinders model improvement.

The main reason for the lack of scholar profiling evaluation studies is the absence of evaluable data. Scholar profiling evaluation data includes manually annotated data by researchers and indicator data such as usage evaluations, usage frequency, and demand matching degree generated during the application of scholar profiles. For large-scale user profiles, manual annotation is difficult. Only through in-depth and extensive applications can relevant indicator data be obtained to support evaluation studies of scholar profiling systems.

5 Summary and Outlook

This paper systematically reviews research on the construction and application of scholar profiling. First, it summarizes the concept of scholar profiling as the process of extracting multi-dimensional attribute features of scholars from massive multi-source heterogeneous academic data for application and analysis. Second, it sorts out the construction process, including data collection, data preprocessing, scholar label extraction, and scholar profiling visualization analysis, involving key technologies such as scholar-related entity extraction, data fusion, and visualization. Finally, it points out challenges in multi-source data acquisition and fusion, dynamic updating of scholar profiles, and evaluation mechanisms.

To address challenges in existing research, the following suggestions are proposed: For scholar profiling data acquisition, advanced text classification algorithms such as XGBoost and LightGBM can be used to identify scholar homepages from search engine results. For information extraction from scholar homepages, large-scale pre-trained models for text representation and deep neural network models can be used to achieve researcher information extraction, ensuring good universality across various domains without relying on manual features. Additionally, some open scientific big data platforms, such as the Chinese Academy of Sciences Knowledge Service Platform [67] and the Guangdong-Hong Kong-Macao Science and Technology Resource Big Data Service Platform

[68], integrate different types of data including scholar personal data, literature resource data, and social media data, providing data supplements for scholar profiling. For data fusion, existing academic knowledge graphs can be combined with deep learning methods to improve the accuracy of scholar-related data fusion. For the lack of scholar profiling evaluation data, training data for expert scholar profiles can be obtained through competitions related to scholar profiling, such as the 2017 Open Academic Precision Profiling Competition and CCKS2021: AMiner Scholar Profiling Competition [69]. Finally, for insufficient research on dynamic updating of scholar profiles, researchers should strengthen practical studies on scholar profiling, closely integrate research with industry, extend research cycles, and obtain usage and feedback data on scholar profiles during practical research to support dynamic updating and evaluation studies of scholar profiles.

Author Contributions: Wang Shiqi: Conducted literature research and drafting; Liu Zhifeng: Conducted literature research and revision; Wang Jimin: Proposed research questions and overall framework, revised the paper.

References: [1] Xu Pengcheng, Bi Qiang, Zhang Han, et al. Data-driven digital library user profiling model construction[J]. *Library and Information Service*, 2021, 65(12): 147-154. [2] Li Jiahui, Zhao Gang. Research on e-commerce user profiling based on big data[J]. *E-commerce*, 2019(1): 46-49. [3] Teng Chun' e, He Chunyu. Construction and application of online medical community user profiles[J]. *Library and Information Service*, 2021, 65(12): 147-154. [4] Xu Hailing, Zhang Haitao, Wei Mingzhu, et al. Research on social media user profiling construction and resource aggregation model[J]. *Library and Information Service*, 2019, 63(9): 109-115. [5] Wang Yajiao, Lu Jia, Ke Xiaojing. Research on academic profiling application in scientific journals[J]. *Chinese Editors Journal*, 2021(4): 45-49. [6] Dong Wenhui, Xiong Huixiang, Du Jin, et al. Research on research collaborator recommendation based on scholar profiling[J/OL]. *Data Analysis and Knowledge Discovery* [2022-05-23]. http://kns.cnki.net/kcms/detail/10.1478.G2.20220407.1054.002.html. [7] Fan Xiaoyu. Research on major project team member recommendation based on multi-source science and technology management data[D]. Xi' an: Xidian University, 2018. [8] Geng Q, Chuai Z, Jin J. Automatic construction of academic profile: a case of information science domain[J]. *Journal of Information Science*, 2021(4): 016555152199804. [9] Holanda O, Elias E, Costa E, et al. Towards an agent-based approach for automatic generation of researcher profiles using multiple data sources[C]//IEEE/WIC/ACM International Conference on Web Intelligence. New York: IEEE, 2013: 163-166. [10] Tang J, Zhang J, Yao L, et al. ArnetMiner: extraction and mining of academic social networks[C]//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas: ACM, 2008. [11] Chen Huixiang, Shao Bo. Research status and implications of user profiling in foreign library field[J]. *Research on Library Science*, 2017(20): 16-20. [12] Liu Haiou, Sun Jingjing, Su Yanjie, et al. Review of user profiling research at home and abroad[J]. *Information Studies: Theory & Application*, 2018, 41(11): 155-160.

[13] Song Meiqi, Chen Ye, Zhang Rui. Review of user profiling research[J]. *Information Science*, 2019, 37(4): 171-177. [14] Zhang Haitao, Xu Hailing, Zhang Xiaohui, et al. Current status and prospects of user profiling research in library and information field at home and abroad[J]. *Library and Information Service*, 2019, 63(7): 127-134. [15] Yuan Sha, Tang Jie, Gu Xiaotao. A survey of scholar profiling technology in open Internet[J]. *Journal of Computer Research and Development*, 2018, 55(9): 1903-1919. [16] Cooper A. *The inmates are running the asylum: why high tech products drive us crazy and how to restore the sanity*[M]. 2nd ed. New York: Pearson Higher Education, 2004. [17] Cooper A, Reimann R, Cronin D. *About face 3: the essentials of interaction design*[M]. Hoboken: John Wiley & Sons, 2007. [18] Baxter K, Courage C, Caine K. *Understanding your users: a practical guide to user research methods*[M]. Burlington: Morgan Kaufmann, 2015. [19] Xu Dihua, Wang Zhijian, Lin Qiaomin, et al. A preference-based personalized tag recommendation system[J]. *Application Research of Computers*, 2011, 28(7): 2573-2575, 2579. [20] Maleszka M, Mianowska B, Nguyen NT. A method for collaborative recommendation using knowledge integration tools and hierarchical structure of user profiles[J]. *Knowledge-based Systems*, 2013, 47: 1-13. [21] Zhang Haitao, Luan Yu, Zhou Honglei. User profiling: advancing towards knowledge[J]. *Document, Information & Knowledge*, 2020(5): 131-134. [22] Yao L, Tang J, Li J. A unified approach to researcher profiling[C]//IEEE/WIC/ACM International Conference on Web Intelligence. Fremont: IEEE, 2007: 359-366. [23] Fan Xiaoyu, Dou Yongxiang, Zhao Pengwei, et al. Research on researcher profiling method integrating multi-source data[J]. *Library and Information Service*, 2018, 62(15): 31-40. [24] Qin Chenglei, Zhang Chengzhi. Problems and countermeasures of peer review in big data environment[J]. *Information Studies: Theory & Application*, 2021, 44(4): 99-112. [25] Yao Yuan, Zhang Hui, Hao Qun, et al. Ontology-based user profiling method[C]//Proceedings of the 22nd Annual Conference on Network New Technology and Applications of China Computer Users Association Network Application Branch. 2018: 232-238. [26] Gao Guangshang. Review of user profiling construction methods[J]. *Data Analysis and Knowledge Discovery*, 2019, 3(3): 25-35. [27] Wang Ruijie. Research on researcher profiling and application based on multi-source information fusion[D]. Chengdu: University of Electronic Science and Technology of China, 2020. [28] Chi Xuehua. Research on automatic construction of precise scholar profiles[D]. Nanjing: Nanjing University of Science and Technology, 2019. [29] Zhang J, Tang J. Name disambiguation in AMiner[J]. *Science China (Information Sciences)*, 2021, 64(4): 214-216. [30] Niu Haibo, Luo Wei, Yin Zhongbo, et al. An open scholar profiling method based on Internet information: CN108090223B[P]. 2020-05-12. [31] Chang Ning, Dou Yongxiang, Xu Wei. Research on author name disambiguation in scientific literature based on multi-source data[J]. *Information Science*, 2021, 39(6): 108-116. [32] Liu G, Yang L. Popular research topics in recent journal publications of library and information science[J]. *The Journal of Academic Librarianship*, 2019, 45(3): 278-287. [33] ResearchGate[EB/OL]. [2022-05-10]. https://www.researchgate.net. [34] Academia.edu - Share

research[EB/OL]. [2022-05-10]. https://www.academia.edu/. [35] ScholarNet - SCHOLAT[EB/OL]. [2022-05-10]. https://www.scholat.com/. [36] Bravo M, Reyes-Ortiz JA, Cruz I. Researcher profile ontology for academic environment[C]//Science and Information Conference. Cham: Springer, 2019: 799-817. [37] Sun J, Xu JG, Cen ZW. Chinese researcher profile annotation based on conditional random fields with semantic rules[C]//World Congress on Engineering. v. III.: International Association of Engineers. London: WCE, 2011: 1818-1822. [38] Zhang Huaping, Shang Jianyun. NLPIR-Parser: big data semantic intelligent analysis platform[J]. *Corpus Linguistics*, 2019, 6(1): 87-104. [39] Che W, Li Z, Liu T. LTP: a Chinese language technology platform[C]//23rd International Conference on Computational Linguistics. Beijing: Coling 2010, Demonstrations, 2010: 13-16. [40] Demartini G. Finding experts using Wikipedia[C]//International Conference on Finding Experts on the Web with Semantics. Karlsruhe: CEUR-WS, 2007. [41] Zeng Jianrong, Zhang Yangsen, Wang Siyuan, et al. Same-name expert disambiguation based on multi-feature fusion[J]. *Journal of the China Society for Scientific and Technical Information*, 2014, 33(6): 613-621. [42] Zhu Yunxia. Research on author name disambiguation in Chinese bibliographic data[J]. *Library and Information Service*, 2014, 58(23): 143-148, 142. [43] Wen Pingmei, Ye Zhiwei, Ding Wenjian, et al. Review of named entity disambiguation research progress[J]. *Information Network Security*, 2017(7): 18-24. [44] Zhao Gang, Yao Xingren. Abnormal behavior detection model based on user profiling[J]. *Information Network Security*, 2017(7): 18-24. [45] Xie Peng. Method for identifying scholar interest labels for academic literature[J]. *Technology Intelligence Engineering*, 2019, 5(3): 65-73. [46] Chi Xuehua, Liu Lifan, Zhang Chengzhi. Research on scholar research interest label discovery based on academic papers[J]. *Data Analysis and Knowledge Discovery*, 2020, 4(9): 15-25. [47] Amini B, Ibrahim R, Othman MS, et al. Capturing scholar' s knowledge from heterogeneous resources for profiling in recommender systems[J]. *Expert Systems with Applications*, 2014, 41(17): 7945-7957. [48] Shi Xiang, Liu Ping. Review of scholar research interest identification[J/OL]. *Data Analysis and Knowledge Discovery* [2022-05-05]. http://kns.cnki.net/kcms/detail/10.1478.G2.20211213.1739.008.html. [49] Hirsch JE. An index to quantify an individual' s scientific research output[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(46): 16569-16572. [50] Xiong Huixiang, Yang Xueping, Jiang Wuxuan, et al. Scholar recommendation research based on academic capability and cooperation relationship network[J]. *Information Science*, 2019, 37(5): 71-78. [51] WordArt.com - Word Cloud Art Creator[EB/OL]. [2022-05-05]. https://wordart.com/. [52] Tagxedo - Word Cloud with Styles[EB/OL]. [2022-05-05]. http://www.tagxedo.com/. [53] Liu J, Tang T, Wang W, et al. A survey of scholarly data visualization[J]. *IEEE Access*, 2018, 6: 19205-19221. [54] Thiagarajan R, Manjunath G, Stumptner M. Finding experts by semantic matching of user profiles[D]. Karlruhe: CEUR-WS, 2008. [55] Hu Chengfang, Li Ji, Wang Chunfang, et al. Research on Lancang-Mekong water resources cooperation expert database construction based on profiling technology[J]. *Changjiang Technical Economic*, 2021, 5(6): 100-106. [56] De Campos LM, Fernández-Luna JM, Huete JF, et al. Automatic construction of multi-faceted user profiles using text clustering and its application to expert recommendation and filtering problems[J]. *Knowledge-based Systems*, 2020, 190: 105337. [57] Ding Y, Yan E, Frazho A, et al. PageRank for ranking authors in co-citation networks[J]. *Journal of the American Society for Information Science and Technology*, 2009, 60(11): 2229-2243. [58] He Juan. Application research on personalized book recommendation based on combination of individual and group user profiles[J]. *Information Studies: Theory & Application*, 2019, 42(1): 129-133, 160. [59] Liu Haiou, Liu Xu, Yao Sumei, et al. Research on personalized learning precision services based on big data deep profiling[J]. *Research on Library Science*, 2019(15): 68-74. [60] Li Yujia, Wang Yicheng. Research on precise recommendation model for academic new media information based on user dynamic profiling[J]. *Information Science*, 2022, 40(1): 88-93, 101. [61] Han Xu, Li Han, Zhang Limin, et al. Research on scholar ranking technology and implementation based on academic behavior[J]. *Computer Knowledge and Technology*, 2019, 15(26): 1-3, 5. [62] Xiong Huixiang, Du Jin, Dai Quanquan, et al. Research on scholar academic influence evaluation based on topics and multi-dimensional measurement indicators[J]. *Information Studies: Theory & Application*, 2021, 44(8): 22-27, 21. [63] Lee M, Cho M, Jeong C, et al. Researcher profiling for researcher analysis service[C]//SWCIB2014 Workshop, Collocated with JIST2014 Conference. Chiang Mai: JIST (Workshops & Posters). 2014: 18-23. [64] Zhao JP, Liu TW, Shi JQ. Improving academic homepage identification from the Web using neural networks[C]//International Conference on Computational Science. London: Springer, 2019: 551-558. [65] Shen Yao, Wang Yi, Yao Yifan, et al. Review of author name disambiguation methods for academic literature[J]. *Data Analysis and Knowledge Discovery*, 2020, 4(8): 15-27. [66] Hu Yuan, Mao Ning. Construction of digital library knowledge community user model based on user profiling[J]. *Library Theory and Practice*, 2017(4): 82-85, 97. [67] Chinese Academy of Sciences Knowledge Service Platform (las.ac.cn)[EB/OL]. [2022-05-18]. https://www.las.ac.cn/. [68] Guangdong-Hong Kong-Macao Science and Technology Resource Big Data Service Platform[EB/OL]. [2022-05-18]. https://talent.dgut-gba.cn/. [69] CCKS2021: AMiner Scholar Profiling - Biendata[EB/OL]. [2022-05-18]. https://www.biendata.xyz/competition/ccks_{aminer}_{profiling}/.

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

Source: ChinaXiv —Machine translation. Verify with original.