

Editorial

New Horizons in Web Search, Web Data Mining, and Web-Based Applications

Jing Zhang ^{1,2,*}, Jipeng Qiang ³ and Cangqi Zhou ⁴¹ School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China² Engineering Research Center of Blockchain Application, Supervision and Management, Southeast University, Ministry of Education, Nanjing 211189, China³ School of Information and Engineering, Yangzhou University, Yangzhou 225127, China; jpqiang@yzu.edu.cn⁴ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; cqzhou@njjust.edu.cn

* Correspondence: jingz@seu.edu.cn

† Current address: Jiulonghu Campus, Southeast University, No. 2 SEU Road, Nanjing 211189, China.

1. Introduction

In today's era of rapid digitization and information technology advancement, web search and web data mining stand at the core of the technological progress of numerous web-based applications [1–3]. Web search is accompanied by the emergence of the Internet, and it continues to develop as Internet applications become increasingly diversified. It has evolved from the early days of navigating people to web pages of interest and providing people with rich content to automatically searching for relevant resources based on the user's characteristics, integrating related functions, and pushing personalized services. The root cause of achieving these exciting web application experiences is that we have a set of web data mining algorithms that continuously analyze massive amounts of web data and user-generated content [4]. They analyze large volumes of data in an automated or semi-automated manner to find hidden functional patterns like outliers, clusters, and association rules, classify targets into different categories, or link two different types of items (i.e., recommender systems).

In general, web search and web data mining are the main ways to extract valuable information from massive network data, and their models, algorithms, and techniques are constantly evolving. As a result, web applications tend to be autonomous, proactive, content-exploring, self-learning, socially collaborative, and location-aware. For example, through user click and eye-tracking modeling, search results can be optimized more accurately based on user characteristics [5]. Advanced autoencoder deep learning models make extracting information from heterogeneous contexts more efficient [6]. In web image search, semi-supervised pseudo-labeling and variational contrastive learning can be used to overcome the influence of noise and obtain better retrieval performance [7]. Embracing location-based social networks into web applications enables the users to register whenever they visit a specific point-of-interest (POI) through the so-called check-ins, or to establish social links with other users in the system [8]. Relying on multiple rounds of natural language, the interaction technology image search engine can obtain more semantically accurate retrieval results [9]. Crowdsourcing technology makes large-scale scientific research collaboration based on the web possible [10,11]. In summary, search engines improve the relevance and accuracy of search results by employing more complex algorithms, such as models based on machine learning and machine intelligence. Web data mining helps enterprises and organizations deeply understand customer needs and optimize products and services by applying complex statistical methods, machine learning, and deep learning technologies. Together with the development of cloud and mobile computing, web-based applications have become more powerful and diverse. These applications support the



Citation: Zhang, J.; Qiang, J.; Zhou, C. New Horizons in Web Search, Web Data Mining, and Web-Based Applications. *Appl. Sci.* **2024**, *14*, 530. <https://doi.org/10.3390/app14020530>

Received: 28 December 2023

Accepted: 3 January 2024

Published: 8 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

operation of multiple industries such as e-commerce [12], online education [13], and remote healthcare [14]. Innovations such as blockchain technology and the application of the Internet of Things have further expanded the possibilities of web applications [15,16], providing users with safer and more personalized services.

The articles published in this Special Issue have shown that web search, web data mining, and web-based applications are in a stage of rapid development. Different research and practices from various fields indicate that with the continuous emergence and application of new technologies, these fields will continue to drive social and technological progress.

2. An Overview of Published Articles

“Predicting Task Planning Ability for Learners Engaged in Searching as Learning Based on Tree-Structured Long Short-Term Memory Networks” by Pengfei Li, Shaoyu Dong, Yin Zhang, and Bin Zhang was published in November 2023, and it proposed a new method by which to predict the task planning ability of learners using network-based search engines in the context of searching as learning (SAL). This method not only improves the accuracy of predicting the task-planning ability of learners but also provides valuable insights for web-based search engines, recommendation systems, and instructional designers. The innovative contribution of this study lies in its ability to help create personalized and efficient search interfaces and support educators in designing more effective learning experiences based on the needs of individual learners.

“WSREB Mechanism: Web Search Results Exploration Mechanism for Blind Users” by Snober Naseer, Umer Rashid, Maha Saddal, Abdur Rehman Khan, Qaisar Abbas, and Yassine Daadaa was published in October 2023, and it introduced an innovative framework for improving the accessibility of network search for blind users and addressing the challenges they face due to information exchange and cognitive pressure. This study proposes a novel WSREB mechanism, which emphasizes accessibility and navigation of web documents while reducing the cognitive load in a non-linear and integrated way. It significantly improves the availability and accessibility of network content for business units. This study helps to redefine the paradigm of online search to promote inclusivity and optimize user experience for blind users, reflecting that technological development in web search increases the well-being of minority groups.

“A Neural-Network-Based Landscape Search Engine: LSE Wisconsin” by Matthew Haffner, Matthew DeWitte, Papia F. Rozario, and Gustavo A. Ovando-Montejo was published in August 2023, and it introduced a search engine, namely, LSE Wisconsin, which extends the perspectives of remote sensing research by implementing image retrieval based on terrain and vegetation features. The new method proposed in this study indicates that the VGG16 and ResNet-50 networks typically produce more favorable results, marking an important step towards developing more comprehensive and high-resolution landscape search engines. This study helps to create powerful and user-friendly digital resources for the research community and users, improving the accessibility and practicality of remote sensing data in various applications.

“Web Page Content Block Identification with Extended Block Properties” by Kiril Griazev and Simona Ramanauskaitė was published in May 2023 and proposed an innovative method for web content block recognition, which is of great significance for automatically integrating web content into other systems. The main technological advancement lies in the ability to describe, in detail, the scope and variants of each content block through text similarity and document object model (DOM) tree analysis. Compared to manual tagging and other existing methods, it can recognize more content blocks, reducing at least 70% of manual tagging work. This work led to a full understanding of the web page structure, making automated integration and transformation of web content possible.

“EFCMF: A Multimodal Robustness Enhancement Framework for Fine-Grained Recognition” by Rongping Zou, Bin Zhu, Yi Chen, Bo Xie, and Bin Shao was published in January 2023, and it proposed an innovative method for fine-grained recognition in multi-mode data. It enhances the learning ability of multimodal data complementarity by randomly

deactivating modal features in the constructed multimodal fine-grained recognition model, solving challenges such as pattern loss and resistance attacks. EFCMF improves the processing of missing modal scenes without additional training. It is worth noting that compared to traditional models under adversarial conditions, it achieves significantly higher accuracy and shows a 27.13% performance improvement.

“Link Prediction with Hypergraphs via Network Embedding” by Zijuan Zhao, Kai Yang, and Jinli Guo was published in December 2022 and introduced a new link prediction method using hypergraphs and network embedding (HNE), demonstrating technological progress in the field of network analysis and providing a new perspective for studying complex relationships. Hypergraphs provide a natural way to represent complex higher-order relationships. The findings of this paper have broad implications, proposing potential applications in different fields such as online social network recommendations and bioinformatics by integrating hypergraphs and network embedding methods.

“Unsupervised Domain Adaptation via Stacked Convolutional Autoencoder” by Yi Zhu, Xinke Zhou, and Xindong Wu was published in December 2022, and it proposed a new unsupervised domain adaptation method that significantly improves domain adaptation technology by using the Stacked Convolutional Sparse Autoencoder (SCSA). It obtains higher-level representations for unsupervised domain adaptation by performing layer projection from the original data. SCSA effectively addresses the challenges of performance degradation caused by ineffective optimization and data redundancy in deep neural networks. Compared with existing methods, it shows superior classification accuracy of up to 89.3%. This research effectively improves the efficiency of using unsupervised methods to transfer knowledge in different domains.

“Development of a Web Application for the Detection of Coronary Artery Calcium from Computed Tomography” by Juan Aguilera-Alvarez, Juan Martínez-Nolasco, Sergio Olmos-Temois, José Padilla-Medina, Víctor Sámano-Ortega, and Micael Bravo-Sanchez was published in November 2022, and it introduced a novel web application that uses Agaston technology for semiautomatic quantification of coronary artery calcium (CAC). This study makes an important advancement in cardiovascular disease analysis. The innovative approach in the system provides accessibility to any device through internet connectivity, which significantly simplifies the processes of healthcare professionals and improves the practicality and efficiency of cardiovascular risk assessment. This system not only simplifies the workflow of cardiologists but may also help with the early detection and management of cardiovascular diseases.

“Fuzzy MLKNN in Credit User Portrait” by Zhuangyi Zhang, Lu Han, and Muzi Chen was published in November 2022, and it proposed an improved fuzzy MLKNN multi-label learning algorithm. The new algorithm solves the subjectivity problem caused by the discretization of credit data and provides more dimensional portraits for credit users. It weakens the subjectivity of credit data after discretization by introducing intuitionistic fuzzy numbers and better realizes the multi-label portrait of credit users by using the corresponding fuzzy Euclidean distance. Compared with traditional MLKNN algorithms, it significantly improves performance, especially in reducing one error. The method creatively combines fuzzy set theory with multi-label learning, paving the way for more sophisticated credit data analysis and potentially aiding in more accurate credit risk assessments.

“Prompt Tuning for Multi-Label Text Classification: How to Link Exercises to Knowledge Concepts?” by Liting Wei, Yan Li, Yi Zhu, Bin Li, and Lejun Zhang was published in October 2022, and it proposed a novel multi-label text classification prompt adjustment method (PTMLTC). The proposed method automatically links exercises with knowledge concepts in educational environments. Specifically, the relevance scores of exercise content and knowledge concepts are learned by a prompt tuning model with a unified template, and then the multiple associated knowledge concepts are selected with a threshold. It solves the cost and time challenges of requiring a large amount of training data in traditional multi-label text classification methods and performs significantly better than existing methods in terms of efficiency and accuracy on the self-constructed Exercises–Concepts dataset

of the Data Structure course. This innovative method not only simplifies the process of connecting educational content but also has the potential for wider application in intelligent education systems.

3. Conclusions

The objectives of this Special Issue on “New Horizons in Web Search, Web Data Mining, and Web-Based Applications” were successfully achieved through the incorporation of groundbreaking research in these domains. Each contribution significantly advanced the understanding and capabilities of web-based technologies, focusing on enhancing information retrieval, intelligent data analysis, and innovative application development. The collective impact of these studies is profound, aligning with the core purpose of science and research: to enhance human experiences and capabilities in the digital age. This issue stands as a testament to the potential of web technologies in shaping a more informed, efficient, and connected world.

Author Contributions: Conceptualization, J.Z.; Investigation, J.Z. and J.Q.; Writing—original draft preparation, J.Q. and C.Z.; Writing—review and editing, J.Z. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: I would like to extend my appreciation to the authors for their diligent research, the reviewers for providing insightful comments and constructive suggestions, as well as the editors and proofreading team for their meticulous attention to detail, ensuring high-quality publishing in terms of both research content and printing standards.

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

References

1. Hitzler, P. A review of the semantic web field. *Commun. ACM* **2021**, *64*, 76–83. [\[CrossRef\]](#)
2. Ristoski, P.; Paulheim, H. Semantic Web in data mining and knowledge discovery: A comprehensive survey. *J. Web Semant.* **2016**, *36*, 1–22. [\[CrossRef\]](#)
3. Al-asadi, T.A.; Obaid, A.J.; Hidayat, R.; Ramli, A.A. A survey on web mining techniques and applications. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2017**, *7*, 1178–1184. [\[CrossRef\]](#)
4. Gheisari, M.; Hamidpour, H.; Liu, Y.; Saedi, P.; Raza, A.; Jalili, A.; Rokhsati, H.; Amin, R. Data Mining Techniques for Web Mining: A Survey. *Artif. Intell. Appl.* **2023**, *1*, 3–10. [\[CrossRef\]](#)
5. Zhang, R.; Xie, X.; Mao, J.; Liu, Y.; Zhang, M.; Ma, S. Constructing a comparison-based click model for web search. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 270–283.
6. Ahmad, F.; Abbasi, A.; Kitchens, B.; Adjeroh, D.; Zeng, D. Deep learning for adverse event detection from web search. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 2681–2695. [\[CrossRef\]](#)
7. Yavuz, M.C.; Yanikoglu, B. VCL-PL: Semi-supervised learning from noisy web data with variational contrastive learning. In Proceedings of the 2022 26th International Conference on Pattern Recognition, Montreal, QC, Canada, 21–25 August 2022; pp. 740–747.
8. Sánchez, P.; Bellogín, A. Point-of-interest recommender systems based on location-based social networks: A survey from an experimental perspective. *Acm Comput. Surv.* **2022**, *54*, 1–37. [\[CrossRef\]](#)
9. Tan, F.; Cascante-Bonilla, P.; Guo, X.; Wu, H.; Feng, S.; Ordonez, V. Drill-down: Interactive retrieval of complex scenes using natural language queries. *arXiv* **2019**, arXiv:1911.03826.
10. Simpson, R.; Page, K.R.; De Roure, D. Zooniverse: Observing the world’s largest citizen science platform. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Republic of Korea, 7–11 April 2014; pp. 1049–1054.
11. Zhang, J. Knowledge learning with crowdsourcing: A brief review and systematic perspective. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 749–762. [\[CrossRef\]](#)
12. Tang, A.K. A systematic literature review and analysis on mobile apps in m-commerce: Implications for future research. *Electron. Commer. Res. Appl.* **2019**, *37*, 100885. [\[CrossRef\]](#)
13. Criollo-C; S; Guerrero-Arias, A.; Jaramillo-Alcázar, Á.; Luján-Mora, S. Mobile learning technologies for education: Benefits and pending issues. *Appl. Sci.* **2021**, *11*, 4111. [\[CrossRef\]](#)
14. Pires, I.M.; Marques, G.; Garcia, N.M.; Flórez-Revuelta, F.; Ponciano, V.; Oniani, S. A research on the classification and applicability of the mobile health applications. *J. Pers. Med.* **2020**, *10*, 11. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Yang, W.; Aghasian, E.; Garg, S.; Herbert, D.; Disiuta, L.; Kang, B. A survey on blockchain-based internet service architecture: Requirements, challenges, trends, and future. *IEEE Access* **2019**, *7*, 75845–75872. [[CrossRef](#)]
16. Da Xu, L.; Viriyasitavat, W. Application of blockchain in collaborative internet-of-things services. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 1295–1305.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.