Journal of Information Hiding and Multimedia Signal Processing ©2024 ISSN 2073-4212
Ubiquitous International Volume **15**, Number **3**, September **2024**

Role of Knowledge-based Data Models in Enhancing Information Retrieval Systems: A Review

Trinh-Dong Nguyen[a,b*], Trong T. Le[a,b], Trinh Huynh Ho Thi Mong[a,b]
[a]Software Engineering Department,
University of Information Technology, Vietnam
[b]Vietnam National University, Ho Chi Minh City 700000, Vietnam
{dongnt, tronglt}@uit.edu.vn, trinhhhtm@uit.edu.vn

Trong-The Nguyen[1,2], Thi-Kien Dao[1,2,*], Vinh-Tiep Nguyen[1,2]
[1]Multimedia Communications Lab.,
VNU-HCM, University of Information Technology, Vietnam
[2]Vietnam National University, Ho Chi Minh City 700000, Vietnam
{thent, kiendt, tiepnv}@uit.edu.vn

*Corresponding author: Thi-Kien Dao

Received June, 2024, revised July, 2024, accepted August, 2024.

Abstract. *The rapid expansion of digital data necessitates advanced information retrieval (IR) systems that can effectively navigate through vast amounts of information. While traditional keyword-based methods continue to be influential, their limitations in understanding context and semantics often lead to imprecise and irrelevant search results. Knowledge-based data models (KBDMs) offer a promising solution, using explicit representations of domain knowledge to enhance retrieval accuracy and user experience. This paper provides a comprehensive review of the current state of KBDMs for IR, exploring their various forms, functionalities, and impact on system performance. We examine how KBDMs address key challenges in IR, such as resolving ambiguity, expanding synonyms, and ranking based on context. Additionally, we discuss the integration of KBDMs with machine learning techniques to improve information extraction and relevance prediction. Finally, we identify open challenges and future research directions for KBDM-based IR systems.*
Keywords: Knowledge-based data models; Information retrieval; Machine learning techniques; Information extraction; Relevance prediction.

1. **Introduction.** The digital information landscape continues to expand, presenting challenges for efficient information retrieval [1][2]. Traditional keyword-based information retrieval systems often struggle to understand user intent, leading to inaccurate results [3]. Ambiguity, synonymy, and polysemy further complicate retrieval, frustrating users and impeding access to relevant information [4][5]. Fig. 1 shows an example of the rapid expansion of the digital information landscape. The everexpanding amount of digital information that people and organizations create, store, and access is known as the expanding digital universe, which includes everything from websites and social media posts to emails, documents, and multimedia content [6]. With the proliferation of digital devices and platforms, the amount of data generated and shared is increasing exponentially [7][8].

The expanding digital world poses various obstacles to managing and retrieving information [9]. The sheer volume of data frequently overwhelms traditional search engines

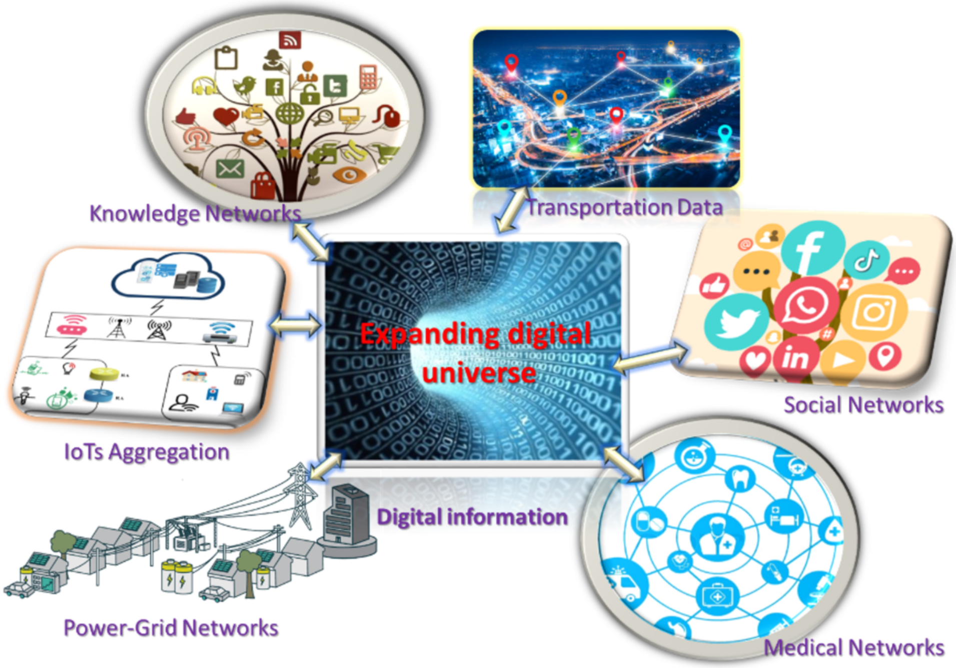

Figure 1. An example of the recent expansion of the digital information landscape

and databases, resulting in slower and less reliable results [10]. Additionally, the complexity of human language and the nuances of user intent make it difficult for these systems to understand and interpret search queries effectively. This limitation has prompted the development of knowledge-based digital models (KBDMs), which aim to understand user intent and context more accurately [11]. For example, KBDMs can analyze the keywords used in a search query and the context in which they are used, leading to more precise and relevant search results [12]. The transformative potential of KBDMs in revolutionizing the future of information retrieval is immense [13], as they can significantly enhance the user experience and improve the accessibility of relevant information in the expanding digital universe [14]. KBDMs offer a promising alternative by integrating explicit domain knowledge into retrieval [11][15]. KBDMs provide a new approach to information retrieval, moving beyond keyword matching to incorporate detailed representations of domain knowledge [16]. These models act as intricate maps, capturing the complex relationships between concepts, entities, and properties within specific domains. Unlike traditional methods, KBDMs enable systems to understand user queries in context, disambiguate terms, identify synonyms, and infer implicit connections between entities [17]. This semantic understanding enhances the IR experience, transforming it from a mundane task to an engaging journey of discovery. Imagine a user seeking information on the "treatment of chronic bronchitis." A traditional keyword-based system might flood them with articles on coughs, lungs, and antibiotics, while overlooking the crucial nuances of their query [18]. Conversely, a KBDM-powered system, armed with knowledge of lung diseases and their various management strategies, would not only retrieve relevant articles on bronchitis treatment but also highlight the importance of early diagnosis, preventive measures, and potential complications [19]. This level of insight and relevance is precisely what KBDMs bring to the table, elevating IR from a mere keyword chase to a meaningful information exchange [20]. By providing the muchneeded bridge between the data and the user's intent, KBDMs promise to usher in a new era of information retrieval [21]. In the following sections, we will delve deeper into the intricate workings of these models, explore their diverse forms and functionalities, and witness how they synergize with machine learning techniques to unlock further potential. With each step, we paint a vivid picture of a future where information retrieval is no longer a frustrating battle but an enlightening and efficient journey powered by the transformative power of knowledge-based data models.

2. **Background and Related Work.** This section reviews the approach to information retrieval that leverages advanced technologies known as knowledge-based digital models (KBDMs), their forms, and functionalities. It also presents their specific contributions to information retrieval (IR) systems. It is broken into the following subsections.

2.1. **Forms and Functionalities.** The knowledge-based digital model is one of the ways to use models with a new approach to information retrieval that leverages advanced technologies such as natural language processing (NLP) [22] and machine learning (ML) to understand user intent and context more accurately. Unlike traditional keyword-based systems, KBDMs go beyond simply matching search queries with keywords and instead focus on understanding the meaning of words and their relationships [23]. By analyzing the semantics of the language, KBDMs can provide more relevant and contextually accurate results, leading to a more precise and personalized user experience. As the digital universe continues to expand, KBDMs are poised to play a pivotal role in transforming the future of information retrieval, offering the potential to significantly enhance the accessibility of relevant information in this vast and evergrowing digital landscape. Fig. 2 illustrates an approach to KBDMs for information retrieval that leverages advanced technologies such as NLP and ML to better understand user intent and context. A cut-

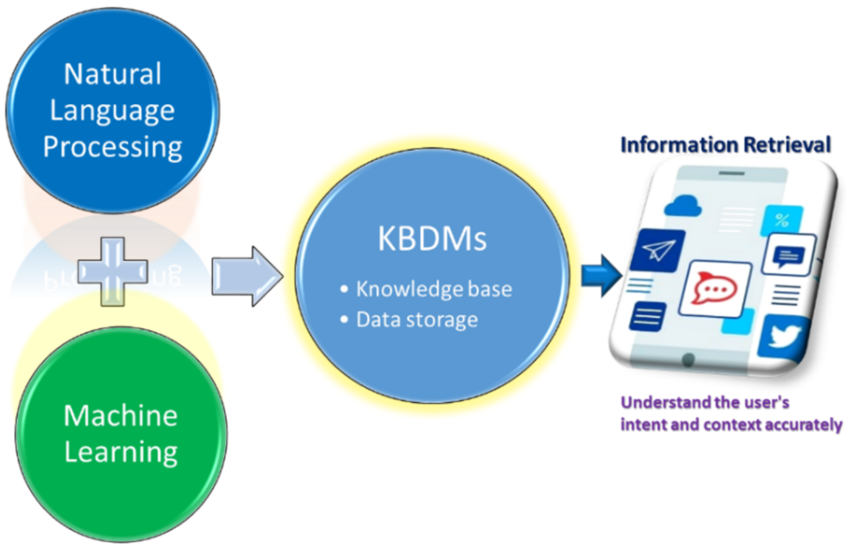

FIGURE 2. An approach to KBDMs for information retrieval that leverages advanced technologies such as NLP and ML to understand user intent and context more accurately

tingedge approach to information retrieval, KBDMs harness advanced technologies like NLP and machine learning to more accurately understand user intent and context [24]. Positioned as pivotal in the transformation of information retrieval, KBDMs hold the potential to significantly improve the accessibility of relevant information within the expansive and evolving digital landscape [25]. KBDMs exist in various forms, each offering unique capabilities for enhancing IR. Notable models commonly used include:

- Ontologies: Formal models capturing concepts, relationships, and properties within a specific domain. Ontologies enable disambiguation of terms, identify synonyms, and infer implicit relationships between entities.
- Semantic networks: Graphbased structures representing entities and their connections. Semantic networks facilitate reasoning over knowledge and inferring implicit information relevant to the user's query.
- Rule-based systems: Sets of logical rules guiding the retrieval process based on domain knowledge. Rule-based systems can leverage ontologies and semantic networks to refine search results and prioritize relevant information.

These KBDMs enhance IR systems in several ways:

- Context understanding: KBDMs allow systems to interpret queries within their domain context, disambiguating terms and identifying relevant concepts beyond literal keywords.
- Synonym expansion: KBDMs can identify synonyms and related terms, expanding the search space and retrieving information that might be missed using keywords alone.
- Relevance ranking: KBDMs enable ranking retrieved information based on its contextspecific relevance to the user's query, rather than solely relying on keyword frequency.
- Personalization: KBDMs can personalize search results by tailoring them to the user's interests, background knowledge, and previous interactions with the system.

Table 1 provides a concise overview of the functionalities and benefits of each KBDM form. Depending on the specific KBDM implementation and domain, additional functionalities and benefits might exist. KBDMs are not a monolithic entity, and they come in diverse flavors, each equipped with unique capabilities to enhance information retrieval (IR) systems. Choosing the right KBDM for a specific task depends on the nature of the domain, the desired functionalities, and the available resources.

2.2. **Knowledge-Based Data Models for IR.** The use of KBDMs promises a journey of discovery, not only of information but also of the interconnectedness of knowledge itself [15]. It encourages embracing challenges, harnessing potential, and embarking on an exciting journey towards a future where information retrieval is an exploration of the boundless sea of knowledge [21].The most common forms of KBDMs are those that make specific contributions to information retrieval, such as semantic networks, conceptual graphs, ontologies, knowledge graphs, rule-based systems, etc [3][47]. As mentioned, KBDMs are not a single type of monolithic entity; instead, they are available in various forms, each with unique qualities that improve IR systems [48]. Depending on the domain, the functionality required, and the resources available, the best KBDM for a given task will vary. They can be mixed and hybridized to capitalize on their advantages and produce even more potent IR systems. Rule-based systems can refine the results received from a combination of ontologies and semantic networks, and an ontology can be integrated into a semantic network to provide richer context comprehension. Fig. 3 illustrates an example of several valuable contributions of KBDMs to the IR systems. As observed in Fig. 3, knowledge-based data models play a crucial role in IR by structuring and organizing information to enhance search processes. Ontologies, semantic networks, and rulebased systems are among the most common forms of KBDMs, each offering specific contributions

Table 1. An overview of KBDM Forms and Functionalities.

Form	Key Functionalities	Benefits for IR	Example Use Case
Ontologies [26]	Disambiguation of terms	Improved relevance by identifying correct meaning [8] [12]	Medical diagnosis retrieval based on symptoms and possible diseases [27]
	Synonym expansion	Broader search space, capturing related concepts [28]	Research paper search using keywords and synonyms for key terms [29]
	Inference of implicit relationships	Retrieving relevant information not explicitly mentioned [30]	Finding scientific articles related to a specific research topic but using broader keywords [31]
Semantic Networks [32]	Context understanding through network traversal	Identifying relevant entities and concepts beyond keywords [33]	Product recommendation based on user's previous purchases and connected items [34][35]
	Information extraction based on network structure	Enhancing document processing accuracy and relevance [36]	Extracting financial data from news articles based on interconnected entities are [37]
	Personalization through user profile integration	Tailored search results based on individual needs [38]	Music recommendation based on user's listening history and genre connections [39]
Rule-based Systems [40]	Prioritization of retrieved information based on relevance rules	Efficiently surfacing the most pertinent results [41]	Email filtering based on rules for spam, sender, and keywords [42]
	Filtering of irrelevant or misleading information	Refining search results for accuracy and quality [43]	News feed curation based on rules for factual content and source credibility [44][23]
	Customization to specific domains and tasks	Adaptability to diverse IR needs [45]	Legal document search based on case-specific rules and terminology [46]

to IR.Ontologies serve as formal models that encapsulate essential concepts, relationships, and properties within a specific domain [26][28]. They function as intricate dictionaries enriched with semantic information, excelling at disambiguation, synonym expansion, and inference [49]. By resolving term ambiguity, identifying synonyms, and allowing for reasoning over knowledge, ontologies ensure that relevant information is retrieved based on the user's intent, even if expressed using different words [48]. Semantic networks, on the other hand, resemble interconnected webs of nodes representing entities and their relationships [32][50]. They excel in context understanding, information extraction, and personalization. By traversing the links within the network, systems can grasp the context surrounding a query, extract relevant information from textual data, and tailor search results to individual needs and interests, resulting in a more personalized IR experience [34]. Rule-based systems operate as sets of if then statements guiding the retrieval process

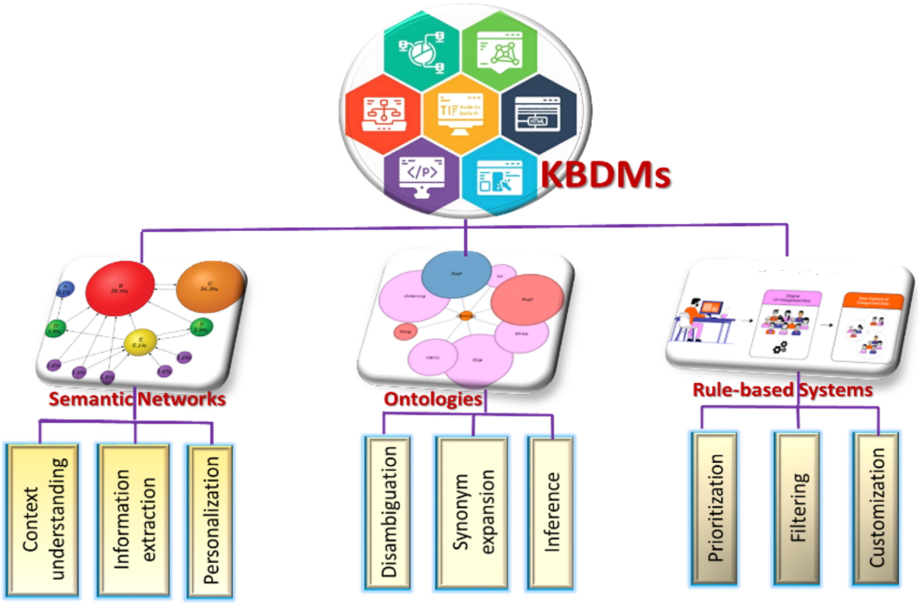

FIGURE 3. Several valuable contributions of KBDMs to the IR systems

based on stored knowledge [40]. They are particularly effective in prioritization, filtering, and customization [23]. By prioritizing retrieved information, filtering out irrelevant data, and being easily customized to specific domains and tasks, rule-based systems ensure that the most pertinent results are surfaced first, enhancing the overall accuracy and efficiency of the IR system [51]. As a whole, these KBDMs—which include ontologies, semantic networks, and rulebased systems—improve the process of finding information by giving us ways to clear up confusion, understand context, and make things our own. This makes search results more accurate and relevant to our needs. It's important to note that these KBDM forms are not mutually exclusive [52]. They can be combined and hybridized to leverage the strengths of each and create even more powerful IR systems. For instance, an ontology can be incorporated into a semantic network to provide richer context understanding, while rule-based systems can be used to refine the results retrieved from a combination of ontologies and semantic networks [48]. By understanding the diverse forms and functionalities of KBDMs, we can effectively unlock their potential for enhancing information retrieval systems. They provide the crucial semantic glue that binds keywords to context, transforming IR from a mechanical keyword chase into a meaningful journey of discovery within the realm of knowledge [53].

3. **KBDMs and Machine Learning.** The integration of KBDMs with machine learning techniques further enhances IR systems [54]. KBDMs provide valuable features and domain knowledge for training machine learning models, such as for[50]:

- Information extraction: KBDMs can guide the extraction of relevant information from textual data, improving the accuracy and efficiency of document processing.
- Relevance prediction: KBDMs can be used to train machine learning models to predict the relevance of retrieved information to the user's query, leading to more accurate and personalized search results.
- Adaptive IR: KBDMs can facilitate the development of adaptive IR systems that learn from user interactions and refine their search strategies over time.

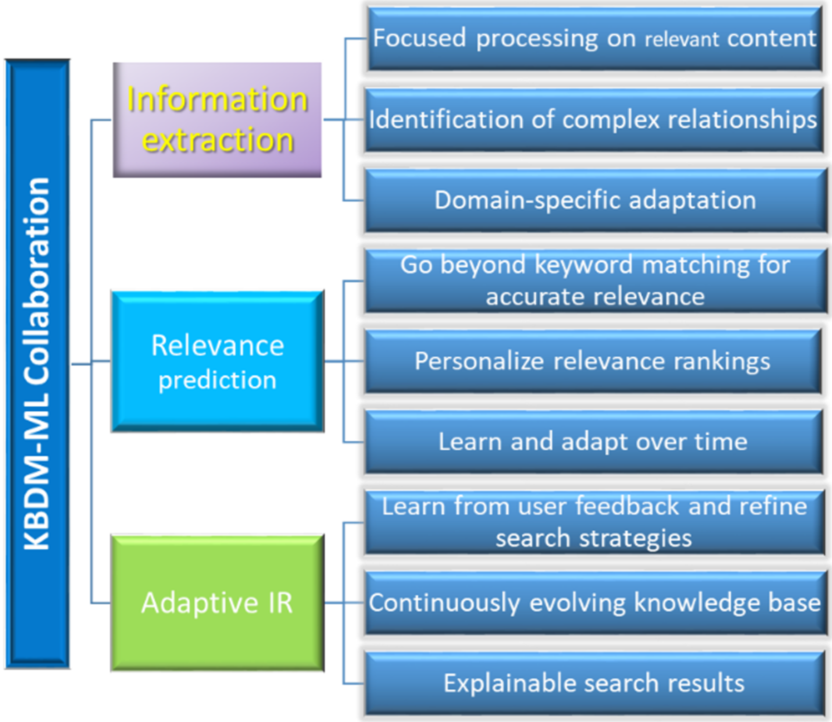

FIGURE 4. An example of the KBDMs and ML collaborating to enhance various aspects of IR

Fig. 4 illustrates an example of the KBDMs and ML collaborating to enhance various aspects of IR, ultimately leading to more effective and efficient retrieval of relevant information for users. In this collaboration, KBDMs such as ontologies, semantic networks, and rule-based systems work in conjunction with ML techniques to improve various aspects of IR. For example, in information extraction, ontologies provide a formal representation of domain knowledge, which ML models can leverage to extract meaningful features from unstructured text data. This collaboration enhances the ability to extract relevant features for tasks such as document classification and query understanding [55]. In relevance prediction, a rule-based system encodes domainspecific rules and logic, combining with ML techniques to predict and enhance information retrieval. The rule-based system can filter and preprocess data before feeding it into ML models, improving the quality of training data and ultimately enhancing the performance of the IR system. Furthermore, in adaptive IR, the semantic network is enriched using ML approaches to capture more nuanced relationships and meanings within the network. ML can identify and infer implicit relationships between concepts in the semantic network, leading to more accurate and contextually rich information retrieval. The collaboration illustrates how KBDMs and ML can work together to push the boundaries of IR, delivering unprecedented accuracy, efficiency, and personalization. The combination of explicit domain knowledge and datadriven insights creates a new frontier in information retrieval, showcasing the true potential of this synergistic alliance. Let's delve into the exciting possibilities that emerge from this collaboration:

a).Information Extraction Supercharged: Extracting relevant information from textual data is fundamental to effective IR. KBDMs provide invaluable guidance for ML models, acting as roadmaps that point to key entities, relationships, and concepts within the text [37]. This enables ML models to:

- Focus on the right content: KBDMs guide the extraction process, directing ML models to prioritize text sections that are most likely to contain relevant information based on the domain knowledge. This reduces processing time and improves extraction accuracy.
- Identify complex relationships: By leveraging the intricate networks of entities and relationships encoded in KBDMs, ML models can extract not only factual information but also nuanced connections between entities. This unlocks the potential for retrieving information relevant to the user's intent even if it is not explicitly stated in the query.
- Adapt to diverse domains: With KBDMs tailored to specific domains, ML models can readily adapt their extraction strategies to different contexts [56]. This ensures optimal performance across a wide range of IR applications.

b). Relevance Prediction Redefined: Traditional IR systems heavily rely on keyword matching to gauge relevance. However, this approach often produces inaccurate results due to ambiguity and context mismatch. KBDMs offer a crucial solution by enabling ML models to:

- Go beyond keywords: By incorporating the semantic understanding provided by KBDMs, ML models can move beyond simple keyword matching and consider the broader context of the query. This allows them to identify relevant information even if it uses different terminologies or conveys the same meaning through alternative phrasing [57].
- Personalize relevance rankings: KBDMs can inform ML models about user preferences and past interactions, enabling them to personalize relevance rankings [58]. This ensures that users receive search results that are not only objectively relevant to their query but also cater to their individual needs and interests.
- Learn and adapt over time: As KBDMs evolve and incorporate new knowledge, ML models trained on them can continuously learn and adapt their relevance prediction strategies [59]. This leads to ever improving IR experiences as the system refines its understanding of user intent and domain specific information.

c) Adaptive IR: A Dynamic Duo: KBDMs and ML form a dynamic duo, creating adaptive IR systems that learn from user interactions and refine their search strategies over time [60]. Imagine a system that remembers your previous searches, understands your evolving interests, and proactively surfaces information that anticipates your needs. This is the magic of adaptive IR:

- User feedback as fuel: KBDMs provide a structured framework for capturing user feedback and incorporating it into the system [61]. This feedback can be used to refine relevance models, prioritize specific entities, and personalize search results over time.
- Continuously evolving knowledge: KBDMs can be dynamically updated with new information and trends, ensuring that the system's knowledge base remains fresh and relevant [62]. This continuous learning process fuels the adaptive capabilities of the IR system.
- Explainable search results: KBDMs provide a foundation for explaining why certain results are deemed relevant [63]. This transparency builds trust with users and allows them to understand the reasoning behind the system's choices.

The confluence of KBDMs and ML marks a watershed moment in IR. By harnessing the power of explicit domain knowledge and data-driven insights, this alliance redefines how we retrieve and interact with information. As research in this area continues to advance, we can expect even more sophisticated IR systems that deliver personalized, adaptive,

Table 2. KBDMs and Machine Learning Synergies for Enhanced IR

KBDM-ML Collaboration	Benefit for IR	Example Application
Information Extraction	Focused processing on relevant content [64] .	Mining financial reports for specific companies and financial instruments [65].
	Identification of complex relationships [66].	Extracting connections between genes and diseases from research papers [67][68].
	Domain-specific adaptation [37].	Extracting legal entities and concepts from court documents [69].
Relevance Prediction	Go beyond keyword matching for accurate relevance [70].	Personalized news recommendations based on user interests and reading history [58].
	Personalize relevance rankings [71].	Prioritizing research articles based on user's past citations and research focus [59].
	Learn and adapt over time [72].	Continuously improve recommendation accuracy based on user feedback and engagement [57].
Adaptive IR	Learn from user feedback and refine search strategies [61].	Suggesting related research topics based on user's current search and past research efforts [60].
	Continuously evolving knowledge base [73].	Update product recommendations based on new market trends and user preferences [62].
	Explainable search results [74].	Provide transparent reasons for ranking specific results [63].

and context-aware search experiences, empowering users to navigate the ever-expanding information landscape with greater efficiency and success.

Table 2 provides a concise overview of the benefits and applications of KBDM-ML collaborations in IR. Depending on the specific KBDM and ML models used, additional benefits and applications might exist. The integration, which uses data-driven insights and explicit domain knowledge to transform information retrieval and interaction, marks a turning point in information retrieval. As research progresses, increasingly complex IR systems should provide context-aware, adaptive, and personalized search experiences, enabling users to successfully and efficiently traverse the constantly changing information landscape.

4. **Challenges and Future Directions.** The promise of KBDMs in revolutionizing IR is undeniable, which means that while the promise of KBDMs in transforming IR is apparent, their road is not without hurdles [50]. The KBDMs' remaining obstacles are depicted in Fig. 5. Navigating these obstacles and mapping out a future will take ongoing research and innovation in several crucial areas, such as knowledge acquisition and maintenance, integration with current systems, user engagement, and explainability.

Fig. 5 shows several promising avenues for future research directions for KBDMs in IR. The future of research in KBDMs in information retrieval (IR) may involve several promising directions. One such avenue is the exploration of automated KBDM construction, which entails leveraging natural language processing and information extraction techniques to streamline the generation and upkeep of KBDMs. By automating these processes, re-searchers aim to enhance the efficiency and scalability of knowledge base

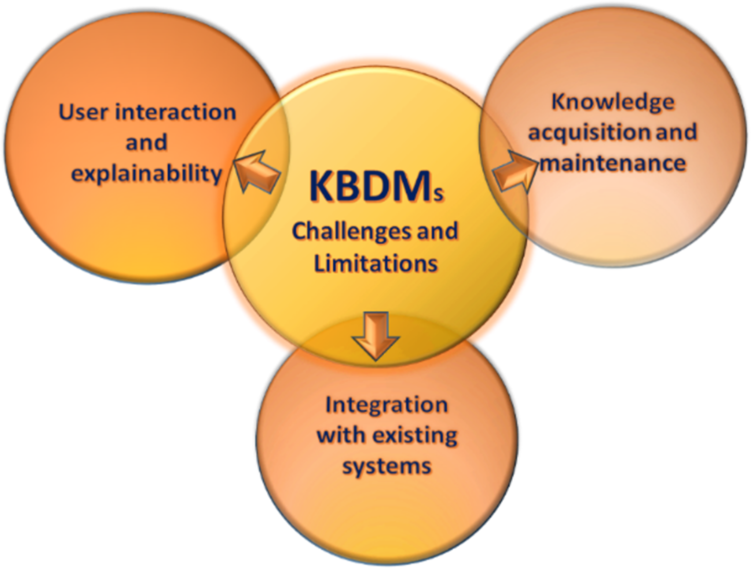

Figure 5. The KBDMs' challenges and limitations remain for enhancing information retrieval systems

development. Another compelling area for future research is the development of hybrid KBDM-ML models. This entails devising innovative approaches that integrate KBDMs with deep learning and other machine-learning techniques. By combining these methodologies, researchers seek to create IR systems that are more resilient and adaptable, capable of handling complex and dynamic information retrieval tasks [68]. Furthermore, future research may also delve into user-centric KBDMs. This involves the exploration of interactive KBDMs that can adjust to individual user preferences and knowledge levels, ultimately providing personalized and explainable search experiences [69]. By tailoring the search experience to the user and offering transparent explanations for search results, these user-centric KBDMs have the potential to enhance the overall usability and effectiveness of IR systems. Table 3 summarizes the key challenges and potential solutions for KBDM-powered IR, along with the exciting future opportunities they present. As research and development progress, these challenges can be addressed, paving the way for a future where KBDMs unlock the full potential of IR systems. While KBDMs offer significant potential for improving IR, challenges remain in several key areas. Firstly, knowledge acquisition and maintenance pose significant hurdles, as building and maintaining comprehensive KBDMs can be costly and time-consuming [85]. Additionally, integrating KBDMs with existing IR infrastructure requires careful consideration of data formats and computational demands to ensure seamless integration. Lastly, user interaction and explain-ability are ongoing challenges, as making KBDM-based systems transparent and allowing users to understand the reasoning be-hind retrieval decisions remains a complex task [68]. Looking ahead, future research directions for KBDMs in IR encompass several promising avenues. One such direction involves automated KBDM construction, which entails leveraging natural language processing and information extraction techniques to auto-mate KBDM generation and maintenance. Another area of interest is the development of hybrid KBDM-ML models, which involves devising novel approaches that combine KBDMs with deep learning and other machine learning techniques to create more robust and adaptive IR systems. Furthermore, exploring user-centric KBDMs, which adapt to individual user preferences and knowledge levels, offers personalized and

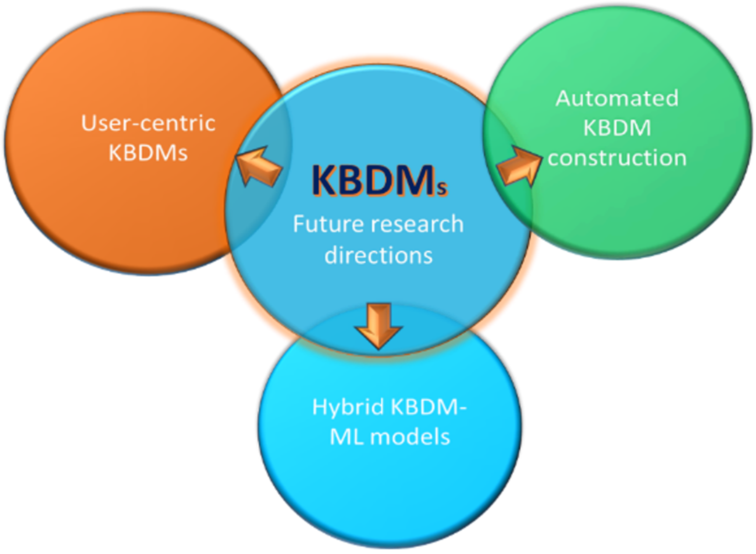

FIGURE 6. Several promising avenues for future research directions for KB-DMs in IR

TABLE 3. Serveral challenges and potential solutions for KBDM-powered IR

Challenge	Potential Solution	Future Opportunity
Knowledge Acquisition and Maintenance [20]	Automated KBDM construction using NLP and information extraction	Democratized KBDM creation and maintenance [75]
	Domain-specific knowledge repositories	Deeper understanding and nuanced search in specialized fields [76]
Integration with Existing Systems [48]	Standardized data formats and APIs	Wider adoption and seamless integration across diverse IR applications [77]
	Lightweight KBDMs for resource-constrained systems	Scalability and accessibility for various platforms [78]
User Interaction and Explainability [15]	Interactive KBDMs that adapt to user preferences	Personalized and intuitive search experiences [79]
	Explainable AI techniques for transparent reasoning and ranking	Building trust and user engagement with KBDM-powered systems [80]
Ethical Considerations [81]	Bias detection and mitigation strategies	Fair and impartial information retrieval across diverse user groups [82]
	Transparent reporting of potential biases in KBDMs	Responsible development and deployment of KBDM-powered IR systems [81]
Domain-Specific Adaptation [60]	Automatic KBDM tailoring for diverse fields	Customizable and adaptable IR solutions for various domains [83]
	Transfer learning techniques for leveraging knowledge across domains	Improved efficiency and accuracy in domain-specific information retrieval [84]

explainable search experiences, addressing the ongoing challenges in user interaction and transparency [83].

a Knowledge Acquisition and Maintenance: Building and maintaining comprehensive KBDMs is a resource-intensive endeavor. Automating KBDM construction through nat-ural language processing and information extraction techniques holds immense poten-tial, but challenges remain in handling ambiguity, evolving terminology, and domain-specific nuances.

b Integration with Existing Systems: Seamless integration of KBDMs with existing IR infrastructure requires careful consideration of data formats, computational demands, and legacy system compatibility. Overcoming these hurdles will ensure wider adoption and unlock the full potential of KBDMs across diverse IR applications.

c User Interaction and Explainability: Making KBDM-based systems transparent and allowing users to understand the reasoning behind retrieved results is crucial for build-ing trust and promoting user engagement. Research into interactive KBDMs that adapt to user preferences and knowledge levels, while providing insightful explanations for search results, will be instrumental in achieving this goal.

d Ethical Considerations: KBDMs can inherit biases present in the data they are built upon. Mitigating these biases and ensuring fair and impartial information retrieval across diverse user groups necessitates careful attention to data provenance, ethical con-siderations in knowledge model construction, and transparent reporting of potential bi-ases.

e Domain-Specific Adaptation: While generic KBDMs offer a starting point, domain-specific models tailored to specific fields can provide deeper and more nuanced understanding. Further research into efficient and automated methods for building and adapt-ing KBDMs to diverse domains is crucial for ensuring their widespread applicability.

As the future of IR is poised to undergo a transformative shift by embracing the synergistic power of KBDMs and ML. By addressing existing challenges and pursuing promising future directions, we can expect remarkable advancements. One such development is the emergence of intelligent personal assistants, powered by KBDM-enhanced IR[85]. These assistants will anticipate information needs, proactively surface relevant data, and adapt to evolving user preferences, fundamentally reshaping how individuals inter-act with information. Additionally, KBDMs can facilitate the creation of decentralized and collaborative knowledge networks, democratizing access to information and foster-ing collaborative knowledge creation across diverse domains. Furthermore, as KBDMs become more sophisticated and interconnected, IR systems will be capable of semantic reasoning and inference, enabling them to retrieve information aligned with user intent and offering deeper understanding of the domain [87]. By embracing the challenges and opportunities presented by KBDMs, we can chart a course towards a future where information retrieval becomes a seamless and insightful journey of discovery. The future of IR lies in harnessing the power of knowledge to em-power users and unlock the vast potential of the information landscape. This shift will not only address existing challenges but also pave the way for a new era where information retrieval is no longer a tedious chore but an enriching experience, offering personalized and contextually rich insights. As KBDMs and ML converge, the future of IR holds the promise of transforming how individuals access and interact with information, ultimately shaping a more intuitive and empowering information landscape.

5. **Conclusion.** This study highlighted the potential of knowledge-based data models (KBDMs) to revo-lutionize information retrieval. KBDMs go beyond traditional keyword-based approach-es by enabling semantic understanding and context-aware search, leading to more accu-rate and personalized information retrieval experiences. KBDMs have shown the ability to enhance query understanding, improve relevance ranking and recommendation sys-tems, increase accuracy in search results, facilitate knowledge discovery, and

support complex queries. Despite challenges such as knowledge modeling complexity and scalability, advance-ments in machine learning (ML), artificial intelligence (AI), and knowledge graph adop-tion offer promising solutions for overcoming these hurdles. The future of information retrieval lies in integrating KBDMs with ML, AI, and standardized knowledge represen-tations, paving the way for the next generation of intelligent informa-tion retrieval sys-tems. This review study can also be a valuable resource for researchers, practitioners, and organizations looking to leverage KBDMs for improved information retrieval. Em-bracing emerging trends and addressing critical challenges will be crucial to fully real-izing the transformative potential of KBDMs and unlocking a new era of information access and utilization.

Acknowledgment. This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2024-26-08.

REFERENCES

[1] D. Bawden, "Information and digital literacies: a review of concepts," *Journal of documentation,* vol. 57, no. 2, pp. 218-259, 2001.

[2] M. Kobayashi and K. Takeda, "Information retrieval on the web," *ACM computing surveys (CSUR),* vol. 32, no. 2, pp. 144-173, 2000.

[3] V. Gupta, D. K. Sharma, and A. Dixit, "Review of information retrieval: Models, performance evaluation techniques and applications," *International Journal of Sensors Wireless Communications and Control,* vol. 11, no. 9, pp. 896-909, 2021.

[4] P. Mahalakshmi and N. S. Fathima, "An Art of Review on Conceptual based Information Retrieval.," *Webology*, vol. 18, no. 1, 2021.

[5] T. Gupta and G. Vidyapeeth, "Keyword extraction: a review," *International Journal of Engineering Applied Sciences and Technology*, vol. 2, no. 4, pp. 215-220, 2017.

[6] Y. Zhang et al., "Neural information retrieval: A literature review," *arXiv preprint arXiv:1611.06792*, vol. 1611, no. 06792, 2016.

[7] J. Best, *The limits of transparency: Ambiguity and the history of international finance.* Cornell University Press, 2018.

[8] J. Feliu, J. Vivaldi, and M. T. Cabré, "Ontologies: a review," *https://repositori.upf.edu/handle/10230/1295*, 2002.

[9] F. Fischer, J. Blakesley, P. Wojcik, and R. Jaschke, "Preface: World Literature in an Expanding Digital Space," *Journal of Cultural Analytics,* vol. 8, no. 2, pp. 1-14, 2023.

[10] J. E. Murphy, C. J. Lewis, C. A. McKillop, and M. Stoeckle, "Expanding digital academic library and archive services at the University of Calgary in response to the COVID-19 pandemic," *Ifla Journal,* vol. 48, no. 1, pp. 83-98, 2022.

[11] Y. Ge and Q. J. Wu, "Knowledge-based planning for intensity-modulated radiation therapy: a review of data-driven approaches," *Medical physics,* vol. 46, no. 6, pp. 2760-2775, 2019.

[12] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 4, pp. 2226-2238, 2013.

[13] T. T. Nguyen, J. S. Pan, and T. K. Dao, "An Improved Flower Pollination Algorithm for Optimizing Layouts of Nodes in Wireless Sensor Network," *IEEE Access,* vol. 7, pp. 75985-75998, 2019, doi: 10.1109/ACCESS.2019.2921721.

[14] M. B. Bashir, M. S. Abd Latiff, A. A. Ahmed, A. Yousif, and M. E. Eltayeeb, "Content-based information retrieval techniques based on grid computing: A review," *IETE Technical Review,* vol. 30, no. 3, pp. 223-232, 2013.

[15] A. Ladj, Z. Wang, O. Meski, F. Belkadi, M. Ritou, and C. Da Cunha, "A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective," *Journal of Manufacturing Systems,* vol. 58, pp. 168-179, 2021.

[16] H. Wang, H. Li, X. Wen, and G. Luo, "Unified modeling for digital twin of a knowledge-based system design," *Robotics and Computer-Integrated Manufacturing,* vol. 68, p. 102074, 2021.

[17] J. Wang, G. Su, C. Wan, X. Huang, and L. Sun, "A Keyword-Based Literature Review Data Generating Algorithm—Analyzing a Field from Scientific Publications," *Symmetry,* vol. 12, no. 6, p. 903, 2020.

[18] M. Schuhmacher and S. P. Ponzetto, "Knowledge-based graph document modeling," *in Proceedings of the 7th ACM international conference on Web search and data mining*, 2014, pp. 543-552.

[19] J.-S. Kang, K. Chung, and E. J. Hong, "Multimedia knowledge-based bridge health monitoring using digital twin," *Multimedia Tools and Applications,* vol. 80, no. 26-27, pp. 34609-34624, 2021.

[20] D. A. Alrahbi, M. Khan, S. Gupta, S. Modgil, and C. J. Chiappetta Jabbour, "Challenges for developing health-care knowledge in the digital age," *Journal of Knowledge Management,* vol. 26, no. 4, pp. 824-853, 2022.

[21] J. H. Gennari et al., "The evolution of Protégé: an environment for knowledge-based systems development," *International Journal of Human-computer studies*, vol. 58, no. 1, pp. 89-123, 2003.

[22] K. Chowdhary and K. R. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence,* pp. 603-649, 2020.

[23] J. A. Bernard, "Use of a rule-based system for process control," *IEEE Control Systems Magazine,* vol. 8, no. 5, pp. 3-13, 1988.

[24] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association,* vol. 18, no. 5, pp. 544-551, 2011.

[25] K. S. Jones, "Natural language processing: a historical review," *Current issues in computational linguistics: in honour of Don Walker,* pp. 3-16, 1994.

[26] V. Ermolayev, S. Batsakis, N. Keberle, O. Tatarintseva, and G. Antoniou, "Ontologies of Time: Review and Trends.," *International Journal of Computer Science & Applications,* vol. 11, no. 3, 2014.

[27] N. Shoaip, A. Rezk, S. El-Sappagh, L. Alarabi, S. Barakat, and M. M. Elmogy, "A comprehensive fuzzy ontology-based decision support system for alzheimer's disease diagnosis," *IEEE Access,* vol. 9, pp. 31350-31372, 2020.

[28] D. Oberle, "How ontologies benefit enterprise applications," *Semantic Web,* vol. 5, no. 6, pp. 473-491, 2014.

[29] K. Munir and M. S. Anjum, "The use of ontologies for effective knowledge modelling and information retrieval," *Applied Computing and Informatics,* vol. 14, no. 2, pp. 116-126, 2018.

[30] T.-K. Dao, S.-C. Chu, T.-T. Nguyen, T.-D. Nguyen, and V.-T. Nguyen, "An Optimal WSN Node Coverage Based on Enhanced Archimedes Optimization Algorithm," *Entropy,* vol. 8, no. 24, 2022, doi: 10.3390/e24081018..

[31] D. Dermeval et al., "Applications of ontologies in requirements engineering: a systematic review of the literature," *Requirements engineering,* vol. 21, pp. 405-437, 2016.

[32] J. Borge-Holthoefer and A. Arenas, "Semantic networks: Structure and dynamics," *Entropy,* vol. 12, no. 5, pp. 1264-1302, 2010.

[33] L. Tamine and L. Goeuriot, "Semantic information retrieval on medical texts: Research challenges, survey, and open issues," *ACM Computing Surveys (CSUR),* vol. 54, no. 7, pp. 1-38, 2021.

[34] L. Berkani, S. Belkacem, M. Ouafi, and A. Guessoum, "Recommendation of users in social networks: A semantic and social based classification approach," *Expert Systems,* vol. 38, no. 2, p. e12634, 2021.

[35] H. B. de Barros Pereira et al., "Systematic review of the 'semantic network' definitions," *Expert Systems with Applications,* 118455, 2022.

[36] B. Mitra and N. Craswell, "Neural models for information retrieval," *arXiv preprint arXiv:1705.01509,* 2017.

[37] J. L. Martinez-Rodriguez, A. Hogan, and I. Lopez-Arevalo, "Information extraction meets the semantic web: a survey," *Semantic Web,* vol. 11, no. 2, pp. 255-335, 2020.

[38] A. Singh and A. Sharma, "Web semantics for personalized information retrieval," *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications,* pp. 795-810, 2018.

[39] M. Schedl, P. Knees, B. McFee, and D. Bogdanov, "Music recommendation systems: Techniques, use cases, and challenges," *in Recommender Systems Handbook,* Springer, 2021, pp. 927-971.

[40] N. Masri et al., "Survey of rule-based systems," *International Journal of Academic Information Systems Research (IJAISR),* vol. 3, no. 7, pp. 1-23, 2019.

[41] F. Aghaeipoor, M. Sabokrou, and A. Fernández, "Fuzzy Rule-Based Explainer Systems for Deep Neural Networks: From Local Explainability to Global Understanding," *IEEE Transactions on Fuzzy Systems,* 2023.

[42] A. Al-Ajeli, E. S. Al-Shamery, and R. Alubady, "An intelligent spam email filtering approach using a learning classifier system," *International Journal of Fuzzy Logic and Intelligent Systems,* vol. 22, no. 3, pp. 233-244, 2022.

[43] T. A. Rana and Y.-N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert systems with applications,* vol. 89, pp. 273-285, 2017.

[44] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment analysis for fake news detection," *Electronics,* vol. 10, no. 11, p. 1348, 2021.

[45] M. R. Besharati and M. Izadi, "KARB Solution: Compliance to Quality by Rule Based Benchmarking," *arXiv preprint arXiv:2007.05874,* 2020.

[46] P. Papadopoulos, M. Soflano, Y. Chaudy, W. Adejo, and T. M. Connolly, "A systematic review of technologies and standards used in the development of rule-based clinical decision support systems," *Health and Technology,* vol. 12, no. 4, pp. 713-727, 2022.

[47] T.-T. Nguyen, C.-S. Shieh, M.-F. Horng, T.-K. Dao, and T.-G. Ngo, "Parallelized Flower Pollination Algorithm with a Communication Strategy," *Proceedings - 2015 IEEE International Conference on Knowledge and Systems Engineering, KSE 2015* , doi: 10.1109/KSE.2015.22.

[48] G. J. Kowalski, *Information retrieval systems: theory and implementation,* vol. 1. springer, 2007.

[49] N. Guarino, "Formal ontology, conceptual analysis and knowledge representation," *International journal of human-computer studies,* vol. 43, no. 5-6, pp. 625-640, 1995.

[50] T.-T. Nguyen, J.-S. Pan, S.-C. Chu, J. F. Roddick, and T.-K. Dao, "Optimization Localization in Wireless Sensor Network Based on Multi-Objective Firefly Algorithm," *Journal of Network Intelligence,* vol. 1, no. 4, pp. 130-138, 2016.

[51] T. Xia, "A constant time complexity spam detection algorithm for boosting throughput on rule-based filtering systems," *IEEE Access,* vol. 8, pp. 82653-82661, 2020.

[52] G. J. Kowalski and M. T. Maybury, Information storage and retrieval systems: theory and implementation, *Springer Science & Business Media* , vol. 8., 2000.

[53] V. N. Gudivada, D. L. Rao, and A. R. Gudivada, "Information retrieval: concepts, models, and systems," *in Handbook of statistics,* vol. 38, Elsevier, 2018, pp. 331-401.

[54] P. Borlund, "The IIR evaluation model: a framework for evaluation of interactive information retrieval systems," *Information research,* vol. 8, no. 3, pp. 3-8, 2003.

[55] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering,* vol. 160, no. 1, pp. 3-24, 2007.

[56] T.-T. Nguyen, T.-K. Dao, T.-T.-T. Nguyen, and T.-D. Nguyen, "An Optimal Microgrid Operations Planning Using Improved Archimedes Optimization Algorithm," *IEEE Access,* vol. 10, pp. 67940-67957, 2022, doi: 10.1109/ACCESS.2022.3185737

[57] Y. Song, X. Shi, and X. Fu, "Evaluating and predicting user engagement change with degraded search relevance," *in Proceedings of the 22nd international conference on World Wide Web,* 2013, pp. 1213-1224.

[58] S. Raza and C. Ding, "News recommender system: a review of recent progress, challenges, and opportunities," *Artificial Intelligence Review,* pp. 1-52, 2022.

[59] X. Y. Ni, H. Huang, and W. P. Du, "Relevance analysis and short-term prediction of PM2. 5 concentrations in Beijing based on multi-source data," *Atmospheric environment,* vol. 150, pp. 146-161, 2017.

[60] J. Sanz-Cruzado, P. Castells, C. Macdonald, and I. Ounis, "Effective contact recommendation in social networks by adaptation of information retrieval models," *Information Processing & Management,* vol. 57, no. 5, p. 102285, 2020.

[61] C. Z. Bond et al., "Adaptive optics with an infrared pyramid wavefront sensor at Keck," *Journal of Astronomical Telescopes, Instruments, and Systems,* vol. 6, no. 3, p. 39003, 2020, doi: 10.1117/1.JATIS.6.3.039003.

[62] Z. Wen, M. X. Zhou, and V. Aggarwal, "Context-aware, adaptive information retrieval for investigative tasks," *in Proceedings of the 12th international conference on Intelligent user interfaces,* 2007, pp. 122-131.

[63] C. Behnert and D. Lewandowski, "Ranking search results in library information systems—Considering ranking approaches adapted from web search engines," *The Journal of Academic Librarianship,* vol. 41, no. 6, pp. 725-735, 2015.

[64] S. G. Small and L. Medsker, "Review of information extraction technologies and applications," *Neural computing and applications,* vol. 25, pp. 533-548, 2014.

[65] P. Ding, L. Zhuoqian, and D. Yuan, "Textual information extraction model of financial reports," *in Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City,* 2019, pp. 404-408.

[66] J. Piskorski and R. Yangarber, "Information extraction: Past, present and future," *Multi-source, multilingual information extraction and summarization,* pp. 23-49, 2013.

[67] À. Bravo, J. Piñero, N. Queralt-Rosinach, M. Rautschka, and L. I. Furlong, "Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research," *BMC bioinformatics,* vol. 16, pp. 1-17, 2015.

[68] M. Bundschus, A. Bauer-Mehren, V. Tresp, L. Furlong, and H.-P. Kriegel, "Digging for knowledge with information extraction: a case study on human gene-disease associations," *in Proceedings of the 19th ACM international conference on Information and knowledge management,* 2010, pp. 1845-1848.

[69] A. V Zadgaonkar and A. J. Agrawal, "An overview of information extraction techniques for legal document analysis and processing.," *International Journal of Electrical & Computer Engineering,* vol. 11, no. 6, pp. 5450-5457, 2021.

[70] Z. Gharibshah and X. Zhu, "User response prediction in online advertising," *aCM Computing Surveys (CSUR),* vol. 54, no. 3, pp. 1-43, 2021.

[71] J. Liu, C. Liu, and N. J. Belkin, "Personalization in text information retrieval: A survey," *Journal of the Association for Information Science and Technology*, vol. 71, no. 3, pp. 349-369, 2020.

[72] P. N. Bennett et al., "Modeling the impact of short-and long-term behavior on search personalization," *in Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval,* 2012, pp. 185-194.

[73] Y. He, E. D. Zamani, S. Lloyd, and C. Luo, "Agile incident response (AIR): Improving the incident response process in healthcare," *International Journal of Information Management,* vol. 62, p. 102435, 2022.

[74] Q. Ai and L. Narayanan. R, "Model-agnostic vs. model-intrinsic interpretability for explainable product search," *in Proceedings of the 30th ACM International Conference on Information & Knowledge Management,* 2021, pp. 5-15.

[75] O. Gudes, E. Kendall, T. Yigitcanlar, and V. Pathak, "Knowledge-based approach for planning healthy cities: the case of Logan-Beaudesert, Australia," *Proceedings of the 3rd Knowledge Cities World Summit-From Theory to Practice,* pp. 796-810, 2010.

[76] C.-H. Chuang, S. E. Jackson, and Y. Jiang, "Can knowledge-intensive teamwork be managed? Examining the roles of HRM systems, leadership, and tacit knowledge," *Journal of management,* vol. 42, no. 2, pp. 524-554, 2016.

[77] M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, "Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications," *Automation in Construction,* vol. 101, pp. 111-126, 2019.

[78] J. Govea, E. Ocampo Edye, S. Revelo-Tapia, and W. Villegas-Ch, "Optimization and Scalability of Educational Platforms: Integration of Artificial Intelligence and Cloud Computing," *Computers,* vol. 12, no. 11, p. 223, 2023.

[79] R. Zimmermann et al., "Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence," *Journal of Research in Interactive Marketing,* vol. 17, no. 2, pp. 273-298, 2023.

[80] D. Kelly, "Methods for evaluating interactive information retrieval systems with users, Foundations and Trends", *in Information Retrieval*, vol. 3, no. 1-2, pp. 1-224, 2009.

[81] M. Schedl, E. Gómez, and E. Lex, "Retrieval and Recommendation Systems at the Crossroads of Artificial Intelligence, Ethics, and Regulation," *in Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval,* 2022, pp. 3420-3424.

[82] M. Eriksen, "Evaluating Fairness in Information Retrieval Systems: A Study on the Performance of the FAIR Metric." *Master's thesis, UIS*, 2023.

[83] H. Abu-Rasheed, C. Weber, J. Zenkert, M. Dornhöfer, and M. Fathi, "Transferrable framework based on knowledge graphs for generating explainable results in domain-specific, intelligent information retrieval," *in Informatics,* 2022, vol. 9, no. 1, p. 6.

[84] M.-T. Nguyen, D. T. Le, and L. Le, "Transformers-based information extraction with limited data for domain-specific business documents," *Engineering Applications of Artificial Intelligence,* vol. 97, p. 104100, 2021.

[85] T.-K. Dao, T.-T. Nguyen, T.-X.-H. Nguyen, T.-D. Nguyen, "Recent Information Hiding Techniques in Digital Systems: A Review", *Journal of Information Hiding and Multimedia Signal Processing,* vol. 15, no. 1, pp. 10-20, March 2024.

[86] C. Cole, "A theory of information need for information retrieval that connects information to knowledge," *Journal of the American Society for Information Science and Technology,* vol. 62, no. 7, pp. 1216-1231, 2011.

[87] T.-T. Nguyen, T.-K. Dao, V.-T. Nguyen, T.-D. Nguyen, T. T. Le, Q.-K. Le, "Integrating Knowledge-Based Approaches for Predictive Socioeconomic Indicator Analysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 15, No. 2, pp. 63-75, June 2024.