



Published in final edited form as:

*Proc ACM Int Conf Inf Knowl Manag.* 2023 October ; 2023: 5021–5025. doi:10.1145/3583780.3614739.

## DiscoverPath: A Knowledge Refinement and Retrieval System for Interdisciplinarity on Biomedical Research

**Yu-Neng Chuang,**

Rice University

**Guanchu Wang,**

Rice University

**Chia-Yuan Chang,**

Texas A&M University

**Kwei-Herng Lai,**

Rice University

**Daochen Zha,**

Rice University

**Ruixiang Tang,**

Rice University

**Fan Yang,**

Rice University

**Alfredo Costilla Reyes,**

Rice University

**Kaixiong Zhou,**

Rice University

**Xiaoqian Jiang,**

UTHealth at Houston

**Xia Hu**

Rice University

### Abstract

The exponential growth in scholarly publications necessitates advanced tools for efficient article retrieval, especially in interdisciplinary fields where diverse terminologies are used to describe similar research. Traditional keyword-based search engines often fall short in assisting users who may not be familiar with specific terminologies. To address this, we present a knowledge graph

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

[ynchuang@rice.edu](mailto:ynchuang@rice.edu) .

based paper search engine for biomedical research to enhance the user experience in discovering relevant queries and articles. The system, dubbed *DiscoverPath*, employs Named Entity Recognition (NER) and part-of-speech (POS) tagging to extract terminologies and relationships from article abstracts to create a KG. To reduce information overload, *DiscoverPath* presents users with a focused subgraph containing the queried entity and its neighboring nodes and incorporates a query recommendation system enabling users to iteratively refine their queries. The system is equipped with an accessible Graphical User Interface that provides an intuitive visualization of the KG, query recommendations, and detailed article information, enabling efficient article retrieval, thus fostering interdisciplinary knowledge exploration. *DiscoverPath* is open-sourced at <https://github.com/ynchuang/DiscoverPath> with a demo video at Youtube.

## Keywords

Healthcare; Biomedical; Knowledge Graph; Recommender System; Information Retrieval System

## 1 INTRODUCTION

With an increasing number of papers being published over the years, search engines like Google Scholar and PubMed search engine<sup>1</sup> have become common tools for researchers and practitioners to identify papers of interest. For example, if a genetics professional wants to explore recent advances in Alzheimer's disease research outcomes, she can use relevant phrases such as "Alzheimer's disease" as search queries. The search engine will then match the queries with the content of the articles and generate a ranked list of relevant articles, as illustrated in the left-hand side of Figure 1.

However, in many cases, users may encounter difficulty in effectively identifying suitable search queries, especially in interdisciplinary fields where researchers from different backgrounds tend to employ diverse terminologies to describe similar research. Take, for instance, a genetics expert exploring Alzheimer's disease research within the proteomics domain. If she is not familiar with proteomics-related terminologies, she may struggle to grasp important queries like Apolipoprotein E4 (ApoE4)<sup>2</sup>, making it difficult to retrieve the relevant articles. This knowledge gap often leads to users spending significant amounts of time filtering out irrelevant information, thereby impeding the advancement of interdisciplinary research.

Unlike traditional search engines, a knowledge graph (KG) [13] offers a more concise and effective way to grasp the interconnectedness of key terminologies. A KG consists of nodes, which represent entities, and edges, which comprehensively represent relationships between those entities from real-world assertions [9, 12]. By utilizing a KG to represent the intricate relationships between terminologies, users can easily discover novel queries or relevant information when retrieving articles of interest [4, 11]. For example, a user can readily identify the ApoE4 protein as a relevant query if it is connected to Alzheimer's disease in

<sup>1</sup> <https://pubmed.ncbi.nlm.nih.gov/>

<sup>2</sup> ApoE4 is a protein that plays a significant role in mammalian fat metabolism and is closely associated with Alzheimer's disease research within the proteomics domain.

the KG, as shown in the screenshot of our system on the right of Figure 1. Several recent advancements [20, 24] have shown the effectiveness of KG-based search engines, but none of them focus on interdisciplinary exploration. Motivated by the concepts above, we ask: ***Can we build a KG-based paper search engine with a user-friendly interface for exploring and discovering relevant queries and articles, especially in interdisciplinary fields?***

However, it is non-trivial to build such a system due to several challenges. (i) *How to construct a KG encompassing various terminologies in research articles?* Detecting the specific terminologies and their corresponding relationships can be challenging, presenting a significant obstacle in constructing a KG. (ii) *How to enable users to identify and retrieve relevant articles efficiently?* The constructed KG may include an overwhelming number of entities and relations, making it challenging for users to identify the relevant information efficiently. (iii) *How to design the Graphical User Interface (GUI)?* The system’s intended users may lack a background in KG or computer science, making it challenging to design a GUI that is accessible and user-friendly for individuals in other disciplines.

To address the challenges and promote interdisciplinary knowledge exploration, we demonstrate DiscoverPath, a KG-based retrieval system designed for biomedical research. This system aims to assist biomedical researchers in dynamically refining their queries and effectively retrieving articles. DiscoverPath is featured for:

- *A KG specifically for biomedical research.* We extract essential terminologies and relationships from the article abstracts to convert textual data into a KG. Additionally, each article is treated as an entity to aid in identifying the relevant articles.
- *A query recommendation system.* To mitigate information overload, we present users with a focused subgraph that only includes the queried entity and its neighboring nodes rather than displaying the entire KG. Moreover, we develop a query recommendation system, which allows users to iteratively modify their queries and observe the resulting subgraphs, facilitating a progressive “path” toward discovering the most relevant articles.
- *A user-friendly GUI.* We offer visualization of knowledge graphs, query recommendations, and detailed information display for each discovered article. An online demo is readily available.

## 2 DISCOVERPATH SYSTEM

Figure 2 presents an overview of DiscoverPath. The system consists of three modules: (i) a KG construction module responsible for extracting terminologies and relations from articles and constructing the KG, (ii) a query recommendation system module that offers relevant query suggestions to refine user queries, and (iii) an interdisciplinarity visualization module that provides a GUI.

## 2.1 Knowledge Graph Construction

We aim to construct a KG with “entities” and “relations” to illustrate the knowledge extracted from the abstract of the articles structurally. Typically, a KG comprises several triplets extracted from the sentences in the format of head entity, relation, and tail entities. Motivated by [23, 29], we construct the KG in two steps: (1) recognizing biomedical-related terminologies as entities and (2) extracting verbs and prepositions as relations between each biomedical-related term. Specifically, we adopt name entity recognition (NER) methods from BERN [15] to extract the potential biomedical-related entities in the abstracts, where BERN is a neural biomedical entity recognition tool trained on articles from PubMed. Then, we extract the relation from the raw abstract utilizing the pre-trained part-of-speech tagging model from NeuralCoref [22].

We have presented the constructed KG to biomedical scientists to validate the correctness of the generated knowledge graphs. The feedback from the scientists indicates that the constructed KG is consistent with the relationships of the articles.

## 2.2 Query Recommendation System

The query recommendation system aims to suggest queries for refinement, where the recommended queries can be either an article or a terminology. We train a recommendation model to learn implicit relationships among the article-to-terminology networks to generate the queries effectively.

The model is trained based on a collection of articles  $(\mathcal{A}, w, w') \in \mathcal{D}_{\mathcal{A}}$ , where  $w$  and  $w'$  denote the relevant and irrelevant terminology of article  $\mathcal{A}$ , respectively. Instantiated by [26, 27, 35], the objective is to learn the relationship between an article  $\mathcal{A}$  and a positive article  $\mathcal{A}^+$  that shares the same terminology as  $\mathcal{A}$ , and a negative article  $\mathcal{A}^-$  that contains the irrelevant terminology  $w'$  of  $\mathcal{A}$ . In the article-to-terminology network, the relations of each article  $\mathcal{A}$  are generally composed of its relevant terminologies and its paper relationship. The optimization criterion is to find an embedding matrix  $\Theta$  that maximizes the joint likelihood as follows:

$$\begin{aligned} \text{Objective} &:= \ln p(\Theta | \mathcal{D}_{\mathcal{A}}) \propto \ln \prod_{(\mathcal{A}, w, w') \in \mathcal{D}_{\mathcal{A}}} p(w > \mathcal{A} | \Theta) p(w' < \mathcal{A} | \Theta) \\ &= \sum_{(\mathcal{A}, w, w') \in \mathcal{D}_{\mathcal{A}}} \ln \sigma(\langle \Theta_{w\mathcal{A}}, (\Theta_{w\mathcal{A}^+} - \Theta_{w'\mathcal{A}^-}) \rangle) - \lambda \|\Theta\|^2, \end{aligned} \quad (1)$$

where  $w > \mathcal{A} w'$  denotes that article  $\mathcal{A}$  related to terminology  $w$  over terminology  $w'$  and  $\lambda$  is a regularization parameter. For every article  $\mathcal{A}$ , in Equation 1,  $\Theta_{w\mathcal{A}}$  represents the combination of an article  $\mathcal{A}$  and relevant terminology  $w$ ; and  $\Theta_{w\mathcal{A}} = \Theta_{\mathcal{A}} + \Theta_w$ .

## 2.3 Interdisciplinarity Visualization

The GUI of DiscoverPath, as depicted in Figure 3, encompasses three interfaces:

(i) a graph visualization section that presents a visualization of the KG, (ii) a query

recommendation field equipped with a search bar for suggesting queries, and (iii) a detailed information board for displaying information of selected entities.

**2.3.1 Graph Visualization Section & detailed information board.**—The KG constructed for biomedical research is extensive and complex, making it challenging for users to easily explore entity information at first glance when presented as a static image [1, 3]. To tackle this obstacle, we present the KG through interactive visualization. Users can effortlessly zoom in on a particular KG element and drag entities to access further insights via pop-up windows. Additionally, double-clicking an entity unveils relevant details on the right-hand side information board. By incorporating these interactive features, *DiscoverPath* offers a more user-friendly and intuitive way to explore interdisciplinary knowledge.

**2.3.2 Query Recommendation Field.**—During interdisciplinary exploration, initial queries provided by users may not always be precise, resulting in retrieval results that do not fully meet their expectations. We use query recommendations in *DiscoverPath* to address this issue. The query recommendations are displayed in two folders to interact with users, helping them refine their queries and obtain personalized and informative KG in the graph visualization section. The query recommendations can enhance the efficiency and effectiveness of interdisciplinary exploration.

## 2.4 Interface Implementation Details

*DiscoverPath* is implemented based on a client-server architecture consisting of a frontend interaction, a backend platform, and a graph database. More details are shown as follows:

- **Frontend (client).** The frontend interface is built upon React [18] framework, handling the communication with the backend via HTTP API endpoints.
- **Backend (API server).** The backend framework Flask [21] offers a flexible structure for API development and database connection configuration. The backend processes retrieve data and provide query recommendation algorithms for better search results.
- **Graph database.** Neo4j [19] graph database is used to store and manage the data. The graph database can efficiently manage complex and highly correlated data like KG.

Note that our system is not restricted to the biomedical domain and can be easily implemented in other research areas. We can also use other LLM-based KG construction methods [31, 36] and recommendation algorithms [6, 16, 25, 25, 28] in the implementation.

## 3 USER PIPELINE AND CASE STUDY

In this section, we present the user's perspective workflow of *DiscoverPath* with a case study focused on Alzheimer's disease.

### 3.1 Interdisciplinarity Discovery Pipeline

The user pipeline of *DiscoverPath* for interdisciplinarity discovery is illustrated as follows:

- **Step 1: Input an initial query.** Users are required to provide their queries in the search bar. After clicking the search bottom, *DiscoverPath* can retrieve the related subgraph for personalized exploration based on their initial queries.
- **Step 2: Observe retrieved KG.** After receiving the knowledge graphs provided by *DiscoverPath*, users can deeply explore the relevant information for interdisciplinarity discovery with a GUI and visualization. Recommended relevant papers are also listed in *DiscoverPath* for wider exploration.
- **Step 3: Explore additional knowledge:** If users need more information, *DiscoverPath* provides query recommendations to help the users amend their initial queries to retrieve better match-specific requirements. Users can go back to Step 1 with refined queries and move to Step 2 for advanced exploration. Step 3 is highly encouraged for whom are not familiar with the appropriate search queries in specific disciplines.

### 3.2 Case Study of *DiscoverPath*

Given Alzheimer's prominence as the fifth leading cause of death in adults over 65, we employ the query "*Alzheimer*" to showcase our system. In the visualized KG, yellow nodes denote articles' PubMed IDs, blue nodes signify query-related terms, and edges indicate node relationships. Paths link articles and terminology, while terminology-article-terminology paths uncover potential in-Readers are encouraged to search with other AD-related queries, including *MCI*, *amyloid*, *cognitive*, and *brain*. In Figure 4, the following insights can be discerned:

- **Terminologies Related with *Alzheimer*.** The neighbor terminology nodes of Alzheimer are: *rTMS*<sup>3</sup>, *Apo-E*<sup>4</sup>, *A $\beta$  clearance*<sup>5</sup>, etc. The edges connected to these nodes indicate the relationships of the terminologies with Alzheimer's disease. For instance, ApoE4 is the factor *for* Alzheimer's disease (AD) and Alzheimer's disease *is* a neurodegenerative disorder. The visualized relationships between these terminologies and Alzheimer's disease align with findings and discussions in previous works [2, 17].
- **Articles Related with *Alzheimer*.** The neighbor article nodes of Alzheimer's disease are the papers with PMID 28474569 [17], 33737172 [30], and 33115936 [7]. The edges linking these paper nodes reveal that Alzheimer's disease is a prevalent terminology within these papers.
- **Query Recommendation.** The recommended queries are shown beneath the KG visualization section, including "tau biomarkers", "group of lysosome enhancing

<sup>3</sup> *Transcranial magnetic stimulation*: A noninvasive form of brain stimulation.

<sup>4</sup> *Apolipoprotein-E*: A protein involved in the metabolism of fats of mammals.

<sup>5</sup> A complex process mediated by various systems and cell types.

compounds”, and “amyloid beta”, etc. These recommended queries provide more specific information than *Alzheimer* for deeper and more accurate exploration.

## 4 LIVE AND INTERACTIVE PART

In the demo session, we will present a live demo and video to explore a AD topic using our system, focusing on the queries of “*Alzheimer*”. Comprehensive instructions on how to utilize our system will be provided, along with a selection of suggested queries to explore, such as “*CSF biomarkers*” and “*MCI*”.

## 5 CONCLUSION AND FUTURE WORK

In this work, we build *DiscoverPath*, a KG-based retrieval system of biomedical data and articles, aiming to assist scientists in exploring interdisciplinary knowledge in biomedical research. The goal of *DiscoverPath* is to promote interdisciplinary exploration and improve the understanding of diseases, ultimately benefiting patients. *DiscoverPath* is a full-stack system with effective KG-based retrieval, query recommendations, and friendly graphic user interaction. In the future, we will keep updating this system, including extending the biomedical databases, exploring more effective ways of knowledge graph construction, developing more efficient approaches for query recommendation [8], explainable and fairness demonstrations [5, 9, 10, 14, 32], and data-centric AI methods [33, 34].

## ACKNOWLEDGEMENTS

XJ is CPRIT Scholar in Cancer Research (RR180012), and he was supported in part by Christopher Sarofim Family Professorship, UT Stars award, UTHealth startup, the National Institute of Health (NIH) under award number R01AG066749, R01AG066749-03S1, R01LM013712, U24LM013755, U01TR002062, U01CA274576, U54HG012510 and the National Science Foundation (NSF) #2124789. The authors would like to thank Chen-Chuan Chang and Dr. Chih-Ming Chen for their generous feedback and assistance.

## REFERENCES

- [1]. Archambault Daniel, Purchase Helen C, and Pinaud Bruno. 2011. Difference map readability for dynamic graphs. In *Graph Drawing: 18th International Symposium, GD 2010, Konstanz, Germany, September 21–24, 2010. Revised Selected Papers 18*. Springer, 50–61.
- [2]. Gloria S Benson Chris Bauer, Hausner Lucrezia, Couturier Samuel, Lewczuk Piotr, Peters Oliver, Michael Hüll Holger Jahn, Jessen Frank, Pantel Johannes, et al. 2022. Don’t forget about tau: the effects of ApoE4 genotype on Alzheimer’s disease cerebrospinal fluid biomarkers in subjects with mild cognitive impairment—data from the Dementia Competence Network. *Journal of Neural Transmission* 129, 5–6 (2022), 477–486. [PubMed: 35061102]
- [3]. Burch Michael and Weiskopf Daniel. 2014. A flip-book of edge-splatted small multiples for visualizing dynamic graphs. In *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*. 29–38.
- [4]. Chang Chia-Yuan, Chen Ning, Chiang Wei-Ting, Lee Chih-Hen, Tseng Yu-Hsuan, Wang Chuan-Ju, Chen Hsien-Hao, and Tsai Ming-Feng. 2020. Query Expansion with Semantic-Based Ellipsis Reduction for Conversational IR.. In *TREC*.
- [5]. Chang Chia-Yuan, Chuang Yu-Neng, Lai Kwei-Herng, Han Xiaotian, Hu Xia, and Zou Na. 2023. Towards Assumption-free Bias Mitigation. *arXiv preprint arXiv:2307.04105* (2023).



- [6]. Chen Chih-Ming, Tsai Ming-Feng, Lin Yu-Ching, and Yang Yi-Hsuan. 2016. Query-based music recommendations via preference embedding. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 79–82.
- [7]. Chu Che-Sheng, Li Cheng-Ta, Andre R Brunoni Fu-Chi Yang, Tseng Ping-Tao, Tu Yu-Kang, Stubbs Brendon, André F Carvalho Trevor Thompson, Yeh Ta-Chuan, et al. 2021. Cognitive effects and acceptability of non-invasive brain stimulation on Alzheimer's disease and mild cognitive impairment: a component network meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry* 92, 2 (2021), 195–203. [PubMed: 33115936]
- [8]. Chuang Yu-Neng, Chen Chih-Ming, Wang Chuan-Ju, Tsai Ming-Feng, Fang Yuan, and Lim Ee-Peng. 2020. TPR: Text-aware preference ranking for recommender systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 215–224.
- [9]. Chuang Yu-Neng, Lai Kwei-Herng, Tang Ruixiang, Du Mengnan, Chang Chia-Yuan, Zou Na, and Hu Xia. 2022. Mitigating relational bias on knowledge graphs. *arXiv preprint arXiv:2211.14489* (2022).
- [10]. Chuang Yu-Neng, Wang Guanchu, Yang Fan, Liu Zirui, Cai Xuanning, Du Mengnan, and Hu Xia. 2023. Efficient xai techniques: A taxonomic survey. *arXiv preprint arXiv:2302.03225* (2023).
- [11]. Dong Junnan, Zhang Qinggang, Huang Xiao, Duan Keyu, Tan Qiaoyu, and Jiang Zhimeng. 2023. Hierarchy-Aware Multi-Hop Question Answering over Knowledge Graphs. In *Proceedings of the ACM Web Conference 2023*. 2519–2527.
- [12]. Dong Junnan, Zhang Qinggang, Huang Xiao, Tan Qiaoyu, Zha Daochen, and Zihao Zhao. 2023. Active ensemble learning for knowledge graph error detection. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 877–885.
- [13]. Hogan Aidan, Blomqvist Eva, Cochez Michael, d'Amato Claudia, de Melo Gerard, Gutierrez Claudio, Kirrane Sabrina, Emilio Labra Gayo José, Navigli Roberto, Neumaier Sebastian, et al. 2021. Knowledge graphs. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–37.
- [14]. Jiang Zhimeng, Han Xiaotian, Fan Chao, Liu Zirui, Zou Na, Mostafavi Ali, and Hu Xia. 2022. Fmp: Toward fair graph message passing against topology bias. *arXiv preprint arXiv:2202.04187* (2022).
- [15]. Lee Jinhyuk, Yoon Wonjin, Kim Sungdong, Kim Donghyeon, Kim Sunkyu, Chan Ho So, and Kang Jaewoo. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240. [PubMed: 31501885]
- [16]. Liu Shih-Yang, Hsien Hao Chen Chih-Ming Chen, Tsai Ming-Feng, and Wang Chuan-Ju. 2022. IPR: Interaction-level Preference Ranking for Explicit Feedback. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1912–1916.
- [17]. Marcelli Serena, Corbo Massimo, Iannuzzi Filomena, Negri Lucia, Blandini Fabio, Nistico Robert, and Feligioni Marco. 2018. The involvement of post-translational modifications in Alzheimer's disease. *Current Alzheimer Research* 15, 4 (2018), 313–335. [PubMed: 28474569]
- [18]. Meta. 2013. React. Retrieved Feb 15th, 2023 from <http://reactjs.org>
- [19]. Neo4j. 2007. Neo4j. Retrieved Feb 15th, 2023 from <https://neo4j.com>
- [20]. Nicholson David N and Greene Casey S. 2020. Constructing knowledge graphs and their biomedical applications. *Computational and structural biotechnology journal* 18 (2020), 1414–1428. [PubMed: 32637040]
- [21]. Ronacher Armin. 2010. Flask. Retrieved Feb 15th, 2023 from <https://flask.palletsprojects.com/en/2.2.x/>
- [22]. Srinivasa-Desikan Bhargav. 2018. *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd.
- [23]. Stewart Michael, Enkhsaikhan Majigsuren, and Liu Wei. 2019. Icdm 2019 knowledge graph contest: Team uwa. In *2019 IEEE international conference on data mining (ICDM)*. IEEE, 1546–1551.
- [24]. Tan Qiaoyu, Liu Ninghao, and Hu Xia. 2019. Deep representation learning for social network analysis. *Frontiers in big Data* 2 (2019), 2. [PubMed: 33693325]



- [25]. Tan Qiaoyu, Liu Ninghao, Zhao Xing, Yang Hongxia, Zhou Jingren, and Hu Xia. 2020. Learning to hash with graph neural networks for recommender systems. In Proceedings of The Web Conference 2020. 1988–1998.
- [26]. Tan Qiaoyu, Zhang Xin, Huang Xiao, Chen Hao, Li Jundong, and Hu Xia. 2023. Collaborative Graph Neural Networks for Attributed Network Embedding. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [27]. Tan Qiaoyu, Zhang Xin, Liu Ninghao, Zha Daochen, Li Li, Chen Rui, Choi Soo-Hyun, and Hu Xia. 2023. Bring your own view: Graph neural networks for link prediction with personalized subgraph selection. In Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. 625–633.
- [28]. Wang Xiang, He Xiangnan, Cao Yixin, Liu Meng, and Chua Tat-Seng. 2019. Kgat: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 950–958.
- [29]. Wu Xuehong, Duan Junwen, Pan Yi, and Li Min. 2023. Medical knowledge graph: Data sources, construction, reasoning, and applications. *Big Data Mining and Analytics* 6, 2 (2023), 201–217.
- [30]. Xu Lu, Pan Cai-Long, Wu Xiang-Hui, Song Jing-Jing, Meng Ping, Li Lei, Wang Li, Zhang Zhiren, and Zhang Zhi-Yuan. 2021. Inhibition of Smad3 in macrophages promotes A $\beta$  efflux from the brain and thereby ameliorates Alzheimer's pathology. *Brain, Behavior, and Immunity* 95 (2021), 154–167. [PubMed: 33737172]
- [31]. Ye Hongbin, Zhang Ningyu, Chen Hui, and Chen Huajun. 2022. Generative knowledge graph construction: A review. *arXiv preprint arXiv:2210.12714* (2022).
- [32]. Yuan Jiayi, Tang Ruixiang, Jiang Xiaoqian, and Hu Xia. 2023. LLM for Patient-Trial Matching: Privacy-Aware Data Augmentation Towards Better Performance and Generalizability. *arXiv preprint arXiv:2303.16756* (2023).
- [33]. Zha Daochen, Zaid Pervaiz Bhat Kwei-Herng Lai, Yang Fan, and Hu Xia. 2023. Data-centric AI: Perspectives and Challenges. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM).
- [34]. Zha Daochen, Zaid Pervaiz Bhat Kwei-Herng Lai, Yang Fan, Jiang Zhimeng, Zhong Shaochen, and Hu Xia. 2023. Data-centric Artificial Intelligence: A Survey. *arXiv preprint arXiv:2303.10158* (2023).
- [35]. Zhou Huachi, Chen Hao, Dong Junnan, Zha Daochen, Zhou Chuang, and Huang Xiao. 2023. Adaptive Popularity Debiasing Aggregator for Graph Collaborative Filtering. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval.
- [36]. Zhu Xiangru, Li Zhixu, Wang Xiaodan, Jiang Xueyao, Sun Penglei, Wang Xuwu, Xiao Yanghua, and Yuan Nicholas Jing. 2022. Multi-modal knowledge graph construction and application: A survey. *IEEE Transactions on Knowledge and Data Engineering* (2022).

**CCS CONCEPTS**

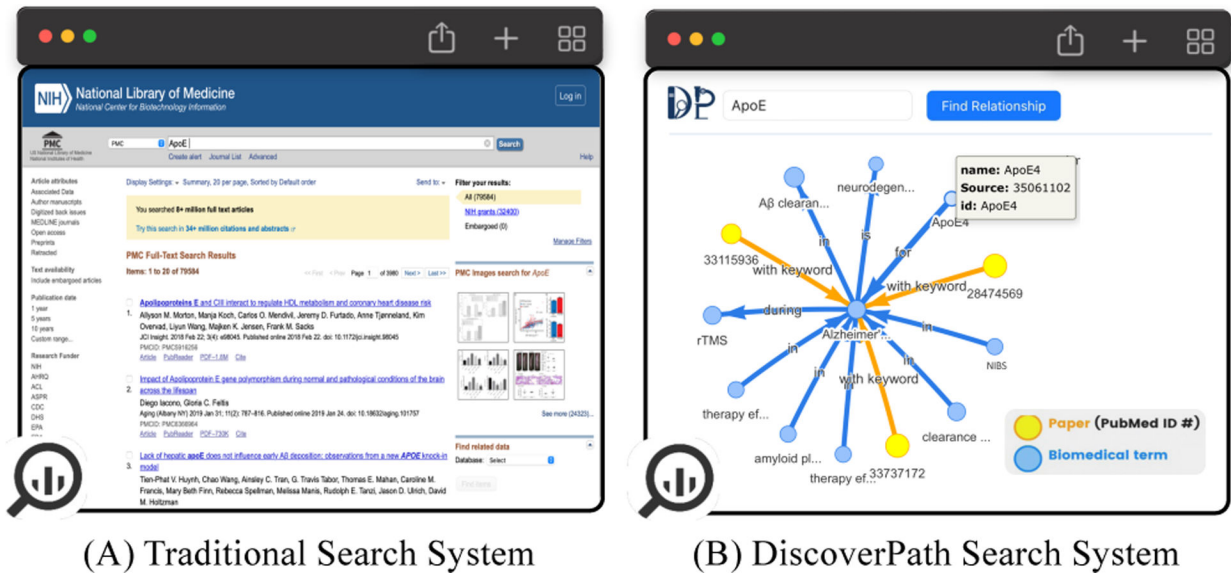
- Applied computing → Health care information systems; • Information systems → Information retrieval.

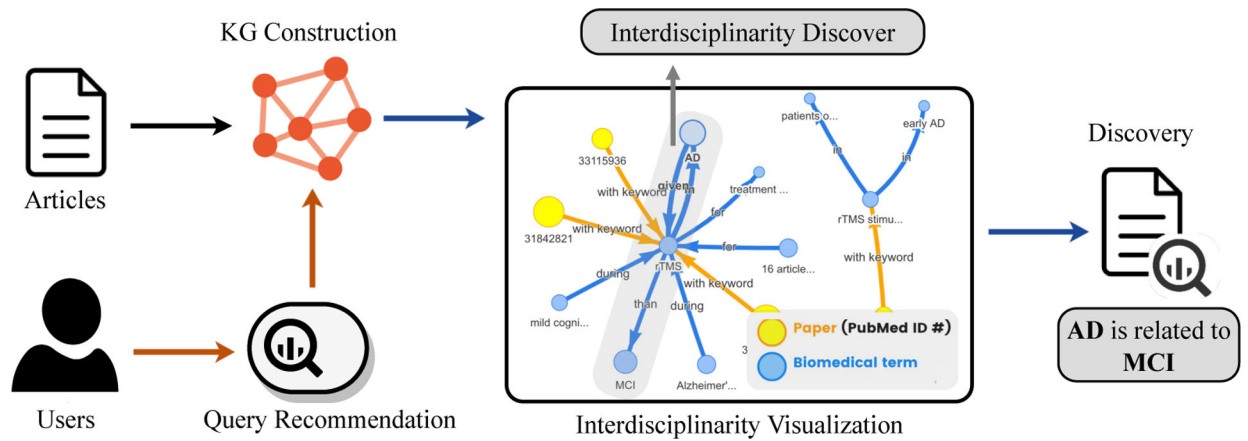
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 2:**

An overview of DiscoverPath for interdisciplinarity knowledge discovery. The KG construction module extracts key terminologies from research articles and converts them into a KG. Starting from an initial query, users progressively refine and query based on suggested related queries from the query recommendation while examining subgraphs generated by the interdisciplinarity visualization. Throughout this process, users can access detailed information on each article of interest.



