A Survey of Knowledge Graph Applications in Entity Retrieval: Postprint

Authors: Ruan Guangce, Fanyuhang, Xia Lei

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] This paper systematically reviews the research trajectory and focal points of entity retrieval based on knowledge graphs, and explores future development directions in this field. [Methods/Process] It outlines the formal definition, implementation pathways, and primary data sources of knowledge graph-based entity retrieval; according to retrieval tasks, it categorizes entity retrieval into three implementation scenarios—matching retrieval, expansion retrieval, and recommendation retrieval—and provides a comprehensive survey of their implementation methods. [Results/Conclusion] As applications continue to deepen, research on knowledge graph-based entity retrieval has begun to focus on optimizing user retrieval experience and providing diverse retrieval results; future in-depth investigations will be conducted across multiple dimensions, including the interpretability of retrieval results and cross-domain knowledge graph retrieval.

Full Text

A Review of Knowledge Graph Applications in Entity Retrieval

Ruan Guangce[1]**, Fan Yuhang**[1]**, Xia Lei**[2] [1]Department of Information Management, Faculty of Economics and Management, East China Normal University, Shanghai 200241 [2]Shanghai Library, Shanghai 200031

Abstract: *[Purpose/Significance]* This paper systematically reviews the research trajectory and key focuses of entity retrieval based on knowledge graphs, and explores future development directions in this field. *[Method/Process]* We outline the formal definition, implementation pathways, and primary data sources for knowledge graph-based entity retrieval. Based on retrieval tasks, we categorize entity retrieval into three scenarios: matching retrieval, extended retrieval, and recommendation retrieval, and review their implementation methods. *[Result/Conclusion]* As applications deepen, research on knowledge graph-based entity retrieval has begun focusing on optimizing user retrieval experiences and providing diverse results. Future research will delve into areas such as retrieval result interpretability, cross-domain knowledge graph retrieval, and more.

Keywords: entity retrieval; knowledge graph; information retrieval **Classification Number:** G250 **DOI:** 10.13266/j.issn.0252-3116.2020.14.013

1 Introduction

Retrieval is a crucial pathway for people to obtain information and knowledge. With technological development, retrieval systems have evolved from simply returning information lists based on user queries to proactively providing information recommendation services [1]. However, users sometimes desire summarized answers and brief overviews when retrieving information—this process is known as entity retrieval. Characterized by directly providing answers or category lists, entity retrieval reduces information consumption costs for users while enhancing their experience. Research indicates that entity-based retrieval intentions account for 52% of all retrieval activities [2], making entity retrieval an indispensable function for modern retrieval systems.

The emergence of large-scale knowledge graphs (also known as knowledge bases) such as YAGO [3-5], DBpedia [6], Freebase [7], NELL [8], and Probase [9] has facilitated the extraction, organization, and management of knowledge from data resources for retrieval systems. Knowledge graphs utilize the structured characteristics of datasets, expressing knowledge through entities and their semantic relationships. This approach not only helps users retrieve and discover relevant entities of interest but also provides in-depth exploratory retrieval services. By understanding user intent at the semantic level and leveraging network information in knowledge graphs to implement complex associative queries, knowledge graphs have elevated the depth of knowledge understanding in retrieval processes.

In recent years, knowledge graph research has made significant progress, with numerous review articles published on topics including knowledge extraction, representation, and reasoning [10], as well as knowledge graph construction and fusion [11-12]. However, these studies lack systematic reviews specifically addressing the progress of knowledge graph applications in entity retrieval. To address this gap, we conducted a literature search in the Web of Science Core Collection using the query: topic=("knowledge graph" or "KG") and "retriev*" (search date: April 11, 2019). After disambiguation screening to exclude medical literature referring to "retrieval" as "recovery," we limited the scope to computer and information science fields. Manual reading removed papers not substantially related to knowledge graph-based entity retrieval, yielding 81 papers for this review.

1.1 Overview of Knowledge Graphs and Their Retrieval Applications

Knowledge graphs aim to describe entities and their relationships in the real or virtual world, essentially forming semantic networks. As early as the late 1960s, M. R. Quillian proposed simulating human language behavior by constructing complex element networks as a knowledge foundation, using labeled directed graphs to discover knowledge related to nodes through object attributes and semantic relationships [13]. Subsequently, R. F. Simmons [14] defined Quillian's model as a "semantic network" in natural language question-answering system research, considering it a preliminary semantic analysis method.

While both semantic networks and knowledge graphs use graph-based knowledge representation, they differ significantly. In semantic networks, nodes represent objects and concepts, and edges represent relationships, with node and edge values definable by users, creating difficulties for multi-source data fusion. Knowledge graphs, however, represent entities or attribute values as nodes and relationships or attributes as edges, explicitly expressing object attributes and semantic relationships through triples. This makes knowledge graph construction more standardized, structurally concise, intuitively flexible, and ensures data quality.

With the development of Linked Open Data (LOD) projects [15], massive RDF (Resource Description Framework) data has been published, transforming the internet from a document web to a data web containing rich descriptions of entities and relationships. In May 2012, Google launched its Knowledge Graph to improve search engine capabilities and user experience. Initially built on public RDF data sources, modern knowledge graphs now automatically discover new entities and relationships from unstructured web data through machine learning and data mining [16]. Today, large-scale knowledge graphs have attracted substantial attention in both academia and industry, finding widespread applications in intelligent search, question answering, personalized recommendation, and content distribution.

Domain ontologies form a crucial step and the backbone of knowledge graphs. As term collections describing specific domains, ontologies define classes, relationships, functions, axioms, and instances that constrain and manage knowledge graphs at the schema level. Leveraging mature ontology libraries, knowledge graph-based entity retrieval has seen particularly deep research and practice in biomedicine, integrating heterogeneous data from single [17] or multiple sources [18-20] into knowledge graphs for retrieval.

1.2 Knowledge Graph-Based Entity Retrieval

An entity refers to any object or concept with specific semantics in the real or virtual world, representing the most basic element in knowledge graphs. Entities are interconnected through relationships, primarily expressed as (entity1-relationship-entity2) and (entity-attribute-attribute value) triples.

If we represent an entity as e, knowledge graph-based entity retrieval can be formalized as $\{D, q, R(q, e)\}$. Here, D represents the knowledge graph (knowledge base) composed of numerous entities and relationships; q represents the user's query; and $R(q, e)$ measures the relevance or similarity between the query and entity e in the dataset, providing a ranked list of entities based on this scoring.

The general implementation pathway for knowledge graph-based entity retrieval consists of two main processes: knowledge graph construction and retrieval implementation [Figure 1: see original paper].

Knowledge graph construction is an iterative process comprising three stages per iteration: (1) **Knowledge extraction** identifies entities, attributes, and relationships from multi-source heterogeneous data, forming ontological knowledge representations; (2) **Knowledge fusion** integrates new knowledge through entity disambiguation (resolving ambiguity from 同名 entities) and co-reference resolution (merging items referring to the same entity); (3) **Knowledge processing** structures the knowledge, typically through ontology construction that defines standards for classes, instances, attributes, relationships, and rules. Knowledge reasoning and quality assessment are crucial processes in this stage—reasoning establishes new associations between entities to enrich the graph, while assessment preserves high-confidence knowledge by discarding low-confidence entries.

During user retrieval, the core challenge is interacting user queries with the knowledge graph. The system must first convert natural language queries, entity pairs, or SPARQL statements into query subgraphs for matching against the entire knowledge graph. It then identifies semantic entities in the query and performs query expansion and reasoning on structural relationships. Finally, the system ranks results by relevance to provide the most suitable answers. Implementation requires appropriate storage and indexing strategies due to varying matching strategies and efficiency considerations.

1.3 Primary Datasets

Large-scale knowledge graphs form the foundation for entity retrieval. Mature international products include DBpedia from Leipzig University, Freebase from MetaWeb (acquired by Google in 2010), Google's Knowledge Graph and Knowledge Vault, Wikimedia Foundation's Wikidata, Max Planck Institute's YAGO, Microsoft's Probase and Bing Satori, Wolfram's WolframAlpha, Facebook's Social Graph, and Stanford's ImageNet. These are applied in products like Google Search Engine, Wikipedia, Apple Siri, and computer vision systems.

Domestically, commercial entities like Baidu and Sogou have developed knowledge graphs for their search engines. Research institutions have also created knowledge graphs including Fudan University's CN-DBpedia, Peking University's PKU-PIE, Tsinghua University's XLore, and the Institute of Automation's Belief-Engine.

2 Applications of Entity Retrieval

Knowledge graph-based entity retrieval demands more detailed data organization, considering not just external resource characteristics but also internal semantic features. Unlike traditional retrieval focusing on formal features, this approach emphasizes the "knowledge" concept itself. After receiving a query, the system must semantically understand user needs, form query subgraphs, and perform matching calculations with the knowledge graph. Results are returned as relevant entities, entity groups, or relationship triples, providing richer semantic features and structured representations than simple text-matching results.

While similar to ontology-based retrieval, knowledge graph retrieval differs by emphasizing data layer construction, enabling additional graph algorithm methods like structural similarity and path distance.

Zhang et al. [35] categorized entity retrieval into relevant entity retrieval and similar entity retrieval. We argue that the core research focus lies in calculating semantic matching between unstructured user needs and knowledge graph networks. As applications mature, research increasingly emphasizes optimizing user experience and providing diverse results. Building on [35], we classify knowledge graph-based entity retrieval into three task scenarios: matching retrieval, extended retrieval, and recommendation retrieval.

2.1 Matching Retrieval

Knowledge graph-based matching retrieval incorporates both semantic and structural similarity, enriching semantic features while presenting new challenges for matching calculations and candidate entity ranking. The general approach creates "documents" for each entity by treating related triples as its content, then calculates relevance between these documents and user queries to rank candidates [36-37].

Two primary matching models exist: **structural entity models** and **hierarchical entity models** [Figure 2: see original paper]. Structural models represent entities using predicate structures without leveraging inter-predicate semantic information. Hierarchical models employ a two-level structure using entity predicate types, providing richer semantic features for matching calculations. Experiments show hierarchical models outperform both unstructured and structural models.

Efficiency is another key research concern. Primary strategies include index construction and data partitioning. Lashkari et al. [38] combined inverted indexes, treaps, and wavelet trees to build semantic hybrid index structures for entities, types, and keywords. Katib et al. [39] designed the RIQ tool using filtered indexes and separate indexes for similar RDF graphs. For partitioning, Hao et al. [40] proposed an association-oriented method (Assc) that significantly reduces inter-partition interactions during queries. Zheng et al. [41] utilized skyline entities for data partitioning in skyline queries to avoid unnecessary candidate partitions. Other optimizations include candidate set pruning through knowledge graph summarization [42] and pre-stored path views for rewriting path queries [43].

With widespread application of large-scale knowledge graphs, researchers identified significant issues of missing relationships and attributes (e.g., in DBpedia) and uncertainty in node relationships. Improving matching accuracy has become a research priority. Common solutions include graph cutting and entity prediction. Yuan et al. [44] iteratively cut uncertain graphs to prune unqualified matches, though this cannot resolve reachability issues in uncertain graphs. Lin et al. [45] addressed this through probabilistic calculations that adjust uncertain edge probabilities toward 0 or 1.

To handle knowledge graph incompleteness, research focuses on prediction methods. Probabilistic language models predict missing triples and entities to fill gaps [46-47], though most assume conditional independence between candidate entities and query keywords, ignoring semantic associations. Some studies calculate inclusion degrees between entity categories and query keywords [48] or introduce proximity concepts [49] to obtain similar candidate entity sets.

2.2 Extended Retrieval

Leveraging knowledge graph networks, extended retrieval offers greater flexibility by inferring user needs from entity pairs or natural language rather than structured statements. This assists users unfamiliar with knowledge graphs or query languages. The goal has shifted from meeting basic needs to addressing diverse, multi-level retrieval requirements.

Entity expansion requires finding entities with shared semantic features. Calculating inter-entity similarity to identify sets similar to user queries is fundamental. Sun et al. [50] computed similarity based on different semantic perspectives according to query emphasis. Mottin et al.'s EQ algorithm [51] used subgraph isomorphism and strong simulation to build equivalence relations for returning similar answer entities. Jayaram et al.'s GQBE [52] constructed maximal query graphs and lattices, outperforming EQ.

For natural language queries, extended retrieval primarily employs rule- and template-based methods. Gupta and Bendersky [53] decomposed long queries into sub-queries, assigned entity weights, and performed query expansion. Zheng et al. [54] used templates to generate structured SPARQL from complex natural language questions. Shi et al. [55] generated query graphs covering different question types by using RDF predicate constraints. Yahya et al. [56] extracted new triples from unstructured sources to expand knowledge graphs.

Identifying latent user needs is another challenge. Chen et al. [57] used path-based semantic features to describe entity commonalities, clarifying desired result set characteristics. Recent perspectives infer latent intentions by analyzing structural features of input entities [58] or calculating information gain from entity types [59]. Zheng et al. [60] employed instance-driven methods to match semantically equivalent but structurally different subgraphs.

Interactive approaches help users complete extended retrieval. Zheng et al. [61] asked clarifying questions to resolve ambiguity while minimizing interaction costs. Arenas et al. [62] used faceted search allowing users to explore by clicking entity property values. Zhang and Li's Maverick system [63] discovered special instance frameworks to help explore unique entities. Aeberli et al.'s JEDI system [64] provided information about actual connections between knowledge graph nodes, including paths and intermediate nodes.

As retrieval targets non-expert users, result diversity and multi-level presentation gain importance. Lissandrin et al.'s X2Q [65] provided query suggestions based on user phrases, returning formal entity names and hierarchical structures. Arnout and Elbassuoni [66] used maximum marginal relevance to avoid homogeneous results. Tonon et al. [67] returned appropriate entity types based on contextual information including frequency and type hierarchy.

2.3 Recommendation Retrieval

Entity recommendation provides relevant entity suggestions based on user queries. Compared to traditional recommendation systems, it must address: (1) entity disambiguation; (2) cross-domain recommendations using knowledge graphs; (3) providing recommendation justifications; and (4) making results more intuitive and readable.

Common methods analyze directly connected entities, extracting those with known relationships as relevant sets [1]. For indirect relationships, similarity calculations determine relevance [68]. Some approaches compute similarity using entity attributes and description texts, such as comparing encyclopedia articles [1] or using Latent Dirichlet Allocation (LDA) to model description documents [69].

Recommendation justification explains why recommended entities are relevant, acting as a bridge between "domain experts" and users. Justifications typically use existing or predicted relationships [70-71], but suffer from insufficient information or lack of semantic depth. Manual annotation can generate explanatory sentences [72], but doesn't scale. Voskarides et al. [73] automatically generated templates from knowledge graph patterns (e.g., "person-author-book" → "[Book] is a [type] by [author]"), though coverage remains limited by existing relationships.

3 Future Directions and Challenges

3.1 Multi-Relational Entity Reasoning

Knowledge graphs are highly structured data containing machine-"understandable" semantic information. While unary relations correspond to entities and binary relations to entity pairs in triples, higher-order relations involve greater semantic complexity. Current entity reasoning focuses primarily on binary relations, though over one-third of entities in Freebase have multi-relational structures [74].

Representing multi-relational expressions requires more complex logic, reducing flexibility and reasoning capability. Multi-relational reasoning demands data organized by relation types, decreasing knowledge base flexibility. Future research must balance logical expression complexity with reasoning capability, minimizing information loss while ensuring flexible reasoning processes.

3.2 Retrieval Result Interpretability

Interpretability—providing semantic explanations for results—is crucial for entity expansion and recommendation, especially in domains requiring expert knowledge. Current methods mostly explain based on shared semantics without addressing knowledge gaps or indirect semantics, limiting explanation quality.

Knowledge graph incompleteness challenges interpretability. Solutions must consider: (1) temporal dynamics of entity relationships [72]; (2) explaining entities without direct relationships; and (3) bridging knowledge gaps. Time-aware reasoning and path-based explanations for indirect relationships are promising directions.

3.3 Integration with Unstructured Data

Entity retrieval requires rich data sources. Unstructured data (text, video, audio) far exceeds structured data in volume. Integrating structured knowledge graphs with unstructured data can supplement emerging entities and relationships.

Unstructured data contains richer entity relationships. Mining these can extend knowledge graphs and improve retrieval effectiveness [75]. Neural networks now enable unified representation of different data types in shared spaces, making it possible to construct entity relationships from unstructured data—a worthwhile exploration direction.

3.4 Data Quality Issues

Data quality problems—knowledge gaps and errors—directly impact retrieval results. Most knowledge graphs suffer from incompleteness; for instance, 93.8% of people lack birthplace information and 78.5% lack nationality information in Freebase [77]. Systems ignoring these issues produce less accurate results.

Solutions include: (1) **Knowledge completion** using rule-based methods [78] or representation learning [79] to fill gaps, though these cannot add entirely new entities/relationships; (2) **Knowledge reasoning** to infer new facts from existing knowledge [80]. Errors arise from extraction mistakes, semantic ambiguity, or outdated data. While some supervised methods detect specific errors (e.g., numerical data [81]), unified error detection models are lacking. Human supervision is costly, motivating automated reasoning-based solutions.

3.5 User Experience

Current evaluation still relies on traditional metrics like precision, recall [42, 47], query time [38], and scalability [46, 50]. Some studies collect user feedback through questionnaires [24, 30], but user experience research remains fragmented.

Result diversity enhances user experience by avoiding redundancy. Arenas et al. [62] considered both relevance and diversity in evaluation metrics. Uyar and Aliyu [82] compared Google and Bing knowledge graphs on entity type coverage and query support. However, no universal user experience metrics exist yet—a gap requiring future work.

3.6 Cross-Domain Knowledge Graph Retrieval

While knowledge graphs often focus on specific domains, user needs may span multiple or even uncovered domains, reducing result quality. Fionda and Pirrò's RECAP [83] supports retrieval across multiple RDF-encoded knowledge graphs using generic SPARQL. Chi et al. [30] built cross-disciplinary scientific resource retrieval by extracting metadata from diverse sources.

Entity disambiguation and alignment remain critical. Yan et al. [84] used character embeddings to measure entity similarity for matching, reducing computational costs. Current cross-domain solutions mostly build new knowledge graphs with generic features, which are limited by domain themes and require substantial resources.

3.7 Personalized Recommendation

Personalization enhances user experience in two ways: (1) recommending interesting content where retrieval becomes recommendation (e.g., citation [31] or Twitter recommendations [85]); (2) tailoring retrieval processes and results to individual users.

User profiles can be inferred solely from queries [33] when personal data is unavailable, or enhanced with historical data for better accuracy. For ambiguous entities, past search themes can aid disambiguation. Similar users' queries can inspire query expansion. Knowledge graphs' semantic processing capabilities provide a foundation for intelligent, personalized information applications.

References

[1] HUANG J Z, DING S Q, WANG H F, et al. Learning to recommend related entities with serendipity for Web search users[J]. ACM transactions on Asian and low-resource language information processing, 2018, 17(3): 1-22. [2]

POUND J, MIKA P, ZARAGOZA H. Ad-Hoc object retrieval in the Web of data[C]//Proceedings of the 19th international conference on World Wide Web. New York: ACM Press, 2010: 771-780. [3] SUCHANEK F M, KASNECI G, WEIKUM G. YAGO: a core of semantic knowledge[C]//Proceedings of the 16th international conference on World Wide Web. New York: ACM Press, 2007: 697-706. [4] HOFFART J, SUCHANEK F M, BERBERICH K, et al. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia[J]. Artificial intelligence, 2013, 194(1): 28-61. [5] MAHDISOLATANI F, BIEGA J, SUCHANEK F. YAGO3: a knowledge base from multilingual Wikipedias[EB/OL]. [2020-05-11]. https://hal-imt.archives-ouvertes.fr/hal-01699874/document. [6] AUER S, BIZER C, KOBILAROV G, et al. DBpedia: a nucleus for a web of open data[C]//Proceedings of the 6th international semantic Web conference. Berlin: Springer-Verlag, 2007: 722-735. [7] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//Proceedings of the 2008 ACM SIGMOD international conference on management of data. New York: ACM Press, 2008: 1247-1250. [8] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an architecture for never-ending language learning[C]//Proceedings of the 24th AAAI conference on artificial intelligence. Menlo Park: AAAI, 2010: 1306-1313. [9] WU W, LI H, WANG H, et al. Probase: a probabilistic taxonomy for text understanding[C]//Proceedings of the 2012 ACM SIGMOD international conference on management of data. New York: ACM Press, 2012: 481-492. [10] XU Z L, SHENG Y P, HE L R, et al. A survey on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606. [11] LIU Q, LI Y, DUAN H, et al. Knowledge graph construction techniques[J]. Journal of computer research and development, 2016, 53(3): 582-600. [12] LIN H L, WANG Y Z, JIA Y T, et al. A survey of knowledge fusion methods for network big data[J]. Chinese journal of computers, 2017, 40(1): 1-27. [13] QUILLIAN M R. Word concepts: a theory and simulation of some basic semantic capabilities[J]. Behavioral science, 1967, 12(5): 410-430. [14] SIMMONS R F. Natural language question answering systems: 1969[J]. Communications of the ACM, 1970, 13(1): 15-30. [15] BIZER C, HEATH T, BERNERS-LEE T. Linked data-the story so far[J]. International journal on semantic Web and information systems, 2009, 5(3): 1-12. [16] DONG X L, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a Web-scale approach to probabilistic knowledge fusion[C]//Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM Press, 2014: 601-610. [17] LAMURIAS A, FERREIRA J D, CLARKE L A, et al. Generating an allergenic cell therapy knowledge graph from literature[J]. Frontiers in immunology, 2017, 8(11): 1-12. [18] MUNGALL C J, MCMURRY J A, KÖHLER S, et al. The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species[J]. Nucleic acids research, 2017, 45(1): 712-722. [19] YU T, LI J, YU Q, et al. Knowledge graph for TCM health preservation: design, construction, and applications[J]. Artificial intelligence in medicine, 2017, 77(3): 48-52. [20] SHI L, LI S, YANG X, et al. Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services[J]. BioMed research international, 2017, Article ID 2858423. [21] FATHALLA S. Detecting human diseases relatedness[J]. International journal on semantic Web and information systems, 2018, 14(3): 120-133. [22] BAKAL G, TALARICO P, KAKANI E V, et al. Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations[J]. Journal of biomedical informatics, 2018, 82(6): 189-199. [23] WANG X, WANG Y, GAO C, et al. Automatic diagnosis with efficient medical case searching based on evolving graphs[J]. IEEE access, 2018, 6: 53307-53318. [24] CHI Y, YU C, QI X, et al. Knowledge management in healthcare sustainability: a smart healthy diet assistant in traditional Chinese medicine culture[J]. Sustainability, 2018, 10(11): 4197-4217. [25] RUAN T, HUANG Y, LIU X, et al. QAnalysis: a question-answer driven analytic tool on knowledge graphs for leveraging electronic medical records for clinical research[J]. BMC medical informatics and decision making, 2019, 19(1): 82-94. [26] ADAMS B, JANOWICZ K. Thematic signatures for cleansing and enriching place-related linked data[J]. International journal of geographical information science, 2015, 29(4): 556-579. [27] FANG Y X, CHENG R, LI X D, et al. Effective community search over spatial graphs[J]. Proceedings of the VLDB endowment, 2016, 9(13): 1521-1524. [28] QIAO B, FANG K, CHEN Y, et al. Building thesaurus-based knowledge graph based on schema layer[J]. Filomat, 2018, 32(5): 1947-1952. [29] GU L, XIA Y, YUAN X, et al. Research on the model for tobacco disease prevention and control based on case-based reasoning and knowledge graph[J]. Sustainability, 2018, 10(4): 995-1015. [30] CHI Y, QIN Y, SONG R, et al. Knowledge graph in smart education: a case study of entrepreneurship scientific publication management[J]. Cluster computing, 2017, 20(1): 81-91. [31] AYALA-GÓMEZ F, DARÓCZY B, BENCZÚR A, et al. Global citation recommendation using knowledge graphs[J]. Journal of intelligent & fuzzy systems, 2018, 34(5): 3089-3100. [32] LIU D, LAI C. Mining group-based knowledge flows for sharing task knowledge[J]. Decision support systems, 2011, 50(2): 370-386. [33] TOMEO P, FERNÁNDEZ-TOBÍAS I, CANTADOR I, et al. Addressing the cold start with positive-only feedback through semantic-based recommendations[J]. International journal of uncertainty, fuzziness and knowledge-based systems, 2017, 25(S2): 57-78. [34] KARIM F, LYTRA I, MADER C, et al. DESERT: a continuous SPARQL query engine for on-demand query answering[J]. International journal of semantic computing, 2018, 12(3): 373-397. [35] ZHANG X L, CHEN Y G, MA D H, et al. A survey on entity search[J]. Journal of software, 2017, 28(6): 1584-1605. [36] NEUMAYER R, BALOG K, NORVÅG K. On the modeling of entities for ad-hoc entity search in the Web of data[C]//Proceedings of the advances in information retrieval, 33rd European conference on IR research. Berlin: Springer-Verlag, 2012: 133-145. [37] BALOG K, NØRVÅG K. On the use of semantic knowledge bases for temporally-aware entity retrieval[C]//Proceedings of the 5th workshop on exploiting semantic annotations in information retrieval. New York: ACM Press, 2012: 1-2. [38] LASHKARI F, ENSAN F, BAGHERI E, et al. Efficient indexing for semantic search[J]. Expert systems with applications, 2017, 73(5):

92-114. [39] KATIB A, SLAVOV V, RAO P. RIQ: fast processing of SPARQL queries on RDF quadruples[J]. Journal of Web semantics, 2016, 37-38(S): 90-111. [40] HAO Y, LI G, YUAN P, et al. An association-oriented partition method for streaming graph query[J]. IEEE transactions on knowledge and data engineering, 2016, 28(7): 1805-1819. [41] ZHENG W, LIAN X, ZOU L, et al. Online subgraph skyline analysis over knowledge graphs[J]. IEEE transactions on knowledge and data engineering, 2016, 28(7): 1805-1819. [42] SONG Q, WU Y, LIN P, et al. Mining summaries for knowledge graph search[J]. IEEE transactions on knowledge and data engineering, 2018, 30(10): 1887-1900. [43] LI J, WU F D, YANG X, et al. APPQKW: an efficient algorithm for processing path queries on the knowledge graph under ecosystem[J]. Ekoloji, 2019, 28(107): 2031-2040. [44] YUAN Y, WANG G, CHEN L, et al. Efficient pattern matching on big uncertain graphs[J]. Information sciences, 2016, 339(4): 369-394. [45] LIN X, PENG Y, CHOI B, et al. Human-powered data cleaning for probabilistic reachability queries on uncertain graphs[J]. IEEE transactions on knowledge and data engineering, 2017, 29(7): 1452-1465. [46] HE L, LIU B, LI G, et al. Knowledge base completion by variational bayesian neural tensor decomposition[J]. Cognitive computation, 2018, 10(6): 1075-1084. [47] KLIEGR T, ZAMAZAL O. LHD2.0: a text mining approach to typing entities in knowledge graphs[J]. Journal of Web semantics, 2016, 39(8): 47-61. [48] CAO T H, HUYNH D T. Subsumption degrees between entity types for approximate knowledge retrieval[J]. International journal of uncertainty, fuzziness and knowledge-based systems, 2007, 15(1): 21-42. [49] JIN J, LUO J, KHEMMARAT S, et al. Querying Web-scale knowledge graphs through effective pruning of search space[J]. IEEE transactions on parallel and distributed systems, 2017, 28(8): 2342-2356. [50] SUN Y Z, HAN J W, YAN X F, et al. PathSim: metapath-based top-K similarity search in heterogeneous information networks[J]. PVLDB, 2011, 4(11): 992-1003. [51] MOTTIN D, LISSANDRINI M, VELEGRAKIS Y, et al. Exemplar queries: a new way of searching[J]. The VLDB journal, 2016, 25(6): 741-765. [52] JAYARAM N, KHAN A, LI C, et al. Querying knowledge graphs by example entity tuples[J]. IEEE transactions on knowledge and data engineering, 2015, 27(10): 2797-2811. [53] GUPTA M, BENDERSKY M. Information retrieval with verbose queries[J]. Foundations and trends in information retrieval, 2015, 9(3/4): 209-354. [54] ZHENG W, YU J X, ZOU L, et al. Question answering over knowledge graphs[J]. Proceedings of the VLDB endowment, 2018, 11(11): 1373-1386. [55] SHIN S, JIN X, JUNG J, et al. Predicate constraints based question answering over knowledge graph[J]. Information processing & management, 2019, 56(3): 445-462. [56] YAHYA M, BERBERICH K, RAMANATH M, et al. Exploratory querying of extended knowledge graphs[J]. Proceedings of the VLDB endowment, 2016, 9(13): 1521-1524. [57] CHEN J, CHEN Y, ZHANG X, et al. Entity set expansion with semantic features of knowledge graphs[J]. Journal of Web semantics, 2018, 52/53(10): 33-44. [58] METZGER S, SCHENKEL R, SYDOW M. QBEES: query-by-example entity search in semantic knowledge graphs based on maximal aspects, diversity-awareness and relaxation[J]. Journal of intelligent information systems, 2017, 49(3):

333-366. [59] SHAN Y, LI M, CHEN Y. Constructing target-aware results for keyword search on knowledge graphs[J]. Data & knowledge engineering, 2017, 110(7): 1-23. [60] ZHENG W G, ZOU L, et al. Semantic SPARQL similarity search over RDF knowledge graphs[J]. Proceedings of the VLDB endowment, 2016, 9(11): 840-851. [61] ZHENG W, CHENG H, YU J X, et al. Interactive natural language question answering over knowledge graphs[J]. Information processing & management, 2019, 56(3): 445-462. [62] ARENAS M, CUENCA GRAU B, KHARLAMOV E, et al. Faceted search over RDF-based knowledge graphs[J]. Journal of Web semantics, 2016, 37-38(3): 55-74. [63] ZHANG G, LI C. Maverick[J]. Proceedings of the VLDB endowment, 2018, 11(12): 1934-1937. [64] AEBERLI C, MONTOYA G, SETTY V, et al. Discovering diversified paths in knowledge bases[J]. Proceedings of the VLDB endowment, 2018, 11(12): 2002-2005. [65] LISSANDRINI M, MOTTIN D, VELEGRAKIS Y, et al. X(2)Q: your personal example-based graph explorer[J]. Proceedings of the VLDB endowment, 2018, 11(12): 2026-2029. [66] ARNAOUT H, ELBASSUONI S. Effective searching of RDF knowledge graphs[J]. Journal of Web semantics, 2018, 48(1): 66-82. [67] TONON A, CATASTA M, PROKOFYEV R, et al. Contextualized ranking of entity types based on knowledge graphs[J]. Journal of Web semantics, 2016, 37/38(3): 170-183. [68] BINBI M A, HAO M A, HSU BO-JUNE (PAUL), et al. Learning to recommend related entities to search users[C]//Proceedings of the 8th ACM international conference on Web search and data mining. New York: Association for Computing Machinery, 2015: 139-148. [69] BLEI M D, NG Y A, JORDAN I M. Latent dirichlet allocation[J]. Journal of machine learning research, 2003, 3(4/5): 993-1022. [70] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[J]. Journal of Web semantics, 2016, 39(8): 47-61. [71] JI G L, LIU K, HE S Z, et al. Knowledge graph completion with adaptive sparse matrix[EB/OL]. [2020-05-11]. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11982/11693. [72] ALTHOFF T, DONG X L, MURPHY K, et al. TimeMachine: timeline generation for knowledge-base entities[C]//Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining. New York: Association for Computing Machinery, 2015: 19-28. [73] VOSKARIDES N, MEIJ E, DERIJKE M. Generating descriptions of entity relationships[C]//Proceedings of 39th European Conference on IR Research. Switzerland: Springer Nature, 2017: 317-329. [74] WEN J, LI J, MAO Y, et al. On the representation and embedding of knowledge bases beyond binary relations[EB/OL]. [2020-05-11]. https://arxiv.org/pdf/1604.08642.pdf. [75] LEE J Y, MIN J K, OH A, et al. Effective ranking and search techniques for Web resources considering semantic relationships[J]. Information processing and management, 2014, 50(1): 132-155. [76] SUCHANEK F M, GROSS-AMBLARD D, ABITEBOUL S. Watermarking for ontologies[C]//The Semantic Web-ISWC 2011. Berlin: Springer-Verlag, 2011: 697-713. [77] MIN B, GRISHMAN R, WAN L, et al. Distant supervision for relation extraction with an incomplete knowledge base[C]//Proceedings of the 2013 conference of the North American chapter of the ACL. Stroudsburg: ACL, 2013: 777-782.

[78] CHEN Y, GOLDBERG S L, WANG D Z, et al. Ontological pathfinding[C]//Proceedings of the 2016 ACM SIGMOD international conference on management of data. New York: ACM Press, 2016: 835-846. [79] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Proceedings of the 29th conference on artificial intelligence. Palo Alto: AAAI, 2015: 2181-2187. [80] GARDNER M, MITCHELL T M. Efficient and expressive knowledge base completion using subgraph feature extraction[C]//Proceedings of the conference on empirical methods in natural language processing. Stroudsburg: Association for Computational Linguistics, 2015: 1488-1498. [81] WIENAND D, PAULHEIM H. Detecting incorrect numerical data in DBpedia[C]//Proceedings of the semantic Web: trends and challenges, 11th international conference. Berlin: Springer-Verlag, 2014: 504-518. [82] UYAR A, ALIYU F M. Evaluating search features of Google knowledge graph and Bing satori[J]. Online information review, 2015, 39(2): 197-213. [83] FIONDA D, PIRRÒ G. Explaining and querying knowledge graphs by relatedness[J]. Proceedings of the VLDB endowment, 2017, 10(12): 1913-1916. [84] YAN J, XU C, LIN, et al. Optimizing model parameter for entity summarization across knowledge graphs[J]. Journal of combinatorial optimization, 2019, 37(1): 293-318. [85] PLAKARIDOU D, STAVRAKAS Y, VASSILIOU Y. Tweet and follower personalized recommendations based on knowledge graphs[J]. Journal of ambient intelligence and humanized computing, 2018, 9(6): 2035-2049.

Author Contributions: Ruan Guangce designed the overall framework and revised the paper; Fan Yuhang collected materials and wrote the initial draft; Xia Lei participated in paper revision.

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