Registration Form for Representative Achievements

Applicant: Yanghua Xiao **College/School: College of Computer Science and Artificial Intelligence**

Research Category: [ ] Basic Innovation [√] Applied Innovation

Achievement Title:

Research on Large-Scale Knowledge Graph Construction, Reasoning, and Application

List of Supporting Materials for the Achievement (itemize the high-level papers, monographs, authorized patents, policy advisory reports, etc. that constitute this achievement):

I. High-Level Papers

1. Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, Yanghua Xiao (corresponding author). CN-DBpedia: A never-ending Chinese knowledge extraction system. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE 2018): 428-438 (cited more than 430 times on Google Scholar)

2. Lu Wang, Yanghua Xiao (corresponding author), Bin Shao, Haixun Wang. How to Partition a Billion-Node Graph. 2014 IEEE 30th International Conference on Data Engineering (ICDE 2014): 568-579 (ICDE 2024 Ten-Year Influential Paper)

3. Siyu Yuan, Jiangjie Chen, Ziquan Fu, Xuyang Ge, Soham Shah, C. R. Jankowski, Deqing Yang, Yanghua Xiao. Distilling Script Knowledge from Large Language Models for Constrained Language Planning. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023): 4303-4325 (ACL 2023 Outstanding Paper Award)

II. Monographs

Yanghua Xiao et al. (eds.), Knowledge Graph: Concepts and Technologies, Publishing House of Electronics Industry, 2020

III. Authorized Patents

1. System and method for entity recognition and linking based on CN-DBpedia, Patent No.: 201810173270.0

2. Error-correction apparatus for a Chinese general concept graph, Patent No.: 202010303271.X

3. A concept extraction system based on knowledge prompting, Patent No.: 202210419758.3

Brief Description of the Achievement (200-300 characters):

Knowledge graphs are an important infrastructure for artificial intelligence, yet their deployment faces three core challenges: high construction cost, limited reasoning capability, and difficulty of adaptation. This project studies the construction, reasoning, and application of large-scale knowledge graphs: it establishes a mechanism for continuous knowledge acquisition in open environments, proposes a knowledge acquisition method based on language model distillation, and develops a methodology for large-scale graph partitioning, search, and question answering. The work has yielded 114 CCF Class A papers, 3 authorized invention patents, and 2 monographs as editor-in-chief. CN-DBpedia has been downloaded more than 50,000 times with over 1.8 billion API calls; the achievements have won the ICDE 2024 Ten-Year Influential Paper Award and the ACL 2023 Outstanding Paper

Award, as well as the Second Prize of the Natural Science Award of the Ministry of Education and the First Prize of the Science and Technology Award of the Shanghai Computer Society (Natural Science Award). They have been deployed at scale in enterprises such as Huawei, Meituan, and Alibaba, and have received a number of collaborative awards including the Huawei Spark Award and the Huawei Outstanding Partner Award.

Notes:

I. Scientific Contributions:

The essence of intelligence is the discovery and application of knowledge. Endowing machines with human knowledge is one of the most important ways to raise the level of machine intelligence and an indispensable path toward cognitive intelligence. As a large-scale semantic network, the knowledge graph is an important form for representing human knowledge and delivering knowledge services, and it has gradually become a key infrastructure driving the development of the artificial intelligence industry. However, the industrial deployment of knowledge graphs still faces three core challenges: at the construction level, the enormous scale entails high cost; at the reasoning level, sparse structure and implicit semantics constrain cognitive depth; and at the application level, diverse scenarios demand efficient response and flexible adaptation. In response, this project takes the application requirements of knowledge graphs as its starting point and conducts in-depth research on the construction, reasoning, and application of large-scale knowledge graphs, yielding the following three scientific findings:

1. A mechanism for continuous knowledge acquisition oriented to open environments was established. To overcome the bottlenecks of traditional knowledge acquisition methods in cost, scale, and quality, a new route for knowledge graph construction was proposed, centered on "automatic extraction – on-demand updating – continuous refinement – crowdsourced verification."

2. A knowledge acquisition method based on a language model distillation mechanism was proposed. To meet the demand for acquiring knowledge with complex structures, the project proposed a knowledge acquisition method that exploits the distillation mechanism of pre-trained language models, significantly reducing the cost of knowledge acquisition while improving the scale and quality of the acquired knowledge.

3. A methodology of partitioning, search, and question answering was proposed for the efficient use of large-scale graphs. To address the storage, retrieval, and interaction bottlenecks that

ultra-large-scale knowledge graphs face in real deployments, the project achieved the first efficient partitioning of a billion-node knowledge graph in the world and received the first ICDE 2024 Ten-Year Influential Paper Award in the Chinese mainland.

The above scientific findings have yielded 114 papers published in Class A journals or conferences recognized by the China Computer Federation (CCF), 3 authorized invention patents, and 2 textbooks and monographs as editor-in-chief, which have been adopted as teaching materials by a number of universities and licensed for publication in Taiwan, China. The applicant has repeatedly been invited to give keynote speeches or academic tutorials at top academic conferences; the results have been well applied in construction tasks such as large-scale encyclopedic and domain-specific graphs, deployed in multiple business lines of Huawei, and honored with the Huawei Spark Award and the Huawei Outstanding Partner Award; they have been deployed at scale on world-class service platforms such as Alibaba and Meituan, winning the first Alibaba Research Award and the Meituan Scientific Research Cooperation Innovation Award, and have also supported the establishment and operation of several university-enterprise joint laboratories. Related research has additionally received the Second Prize of the Natural Science Award of the Ministry of Education and the First Prize of the Science and Technology Award of the Shanghai Computer Society (Natural Science Award) (see Figures 1 and 2).

The overall results of the project have attracted considerable attention from international peers. The applicant was invited by international conferences in this field such as ADMA, CIKM, and DASFAA to present the research results systematically to international peers in the form of academic tutorials; commissioned by the CCF, the applicant also organized CCF frontier tutorial programs on knowledge graph construction, which generated a strong response, and was named a CCF Outstanding Academic Director and selected as a CCF Distinguished Speaker for two consecutive years. The applicant has also been invited to give talks at well-known universities at home and abroad and at Fortune Global 500 companies, and has delivered more than 300 keynote or invited talks in total at major international conferences such as WAIC and the Bund Conference.

II. Evaluation of Representative Publications

The Chinese encyclopedic knowledge graph CN-DBpedia constructed by the applicant (Paper 1) is regarded as a high-quality knowledge source and knowledge acquisition method and has been widely used in tasks such as domain graph construction. A number of well-known scholars, including Jinlong Wang, Academician of the Chinese Academy of Sciences, and Yike Guo, Foreign Member of the Chinese Academy of Engineering, have spoken highly of it and cited it many times,

and it has been cited more than 430 times on Google Scholar. Academician Jinlong Wang commented that “CN-DBpedia is one of the earliest open encyclopedic Chinese knowledge graphs in China and also one of the largest at present.” In research published at IJCAI 2019, a top conference on artificial intelligence, IEEE/ACM Fellow Philip S. Yu used CN-DBpedia as the core knowledge graph source for entities to construct his heterogeneous social event information network. In a paper in IEEE Access 2019, IEEE Fellow Mohsen Guizani regarded the encyclopedic graph and the concept graph constructed by the project as typical representatives of large-scale knowledge graphs (“there are several large-scale knowledge bases … such as CN-Probase and CN-DBpedia”).

The multi-level label propagation (MLP) graph partitioning method proposed by the applicant (Paper 2) has been recognized by international peers as a classic graph partitioning method, and many well-known scholars, including several IEEE/ACM/AAAS Fellows, have cited or compared it alongside the METIS method proposed by Vipin Kumar, an AAAI/IEEE/ACM/AAAS/SIAM Fellow. Professor M. Tamer Özsu, Fellow of the Royal Society of Canada and IEEE/ACM/AAAS Fellow, considers the MLP method to be one of the two major partitioning methods alongside METIS. MLP was recognized by ICDE (a CCF Class A conference), one of the three authoritative international conferences on databases, and received the ICDE 2024 Ten-Year Influential Paper Award, the first time this honor has been granted in the Chinese mainland; the award committee held that this work greatly reduces the response latency of graph systems and the cost of graph data access, thereby enabling large model systems that need to access world knowledge to be deployed effectively in the future. The related work on community search has been recognized and followed up by numerous scholars and applied by scholars from different industries to social network analysis, bioinformatics research, e-commerce recommendation systems, anomalous behavior analysis, and many other scenarios, providing high-quality retrieval solutions with accurate results and rapid response.

The knowledge acquisition method based on a language model distillation mechanism proposed by the applicant (Paper 3) stood out among the 4,864 submissions to ACL 2023, a top conference in natural language processing, and won the ACL 2023 Outstanding Paper Award. This work was the first to define the constrained language planning task and proposed an over-generate-then-filter knowledge distillation method that distills script knowledge from large language models; it built the CoScript dataset containing 55,000 scripts, improved the accuracy of constrained language planning by 26% over previous methods, and effectively endowed small language models with constrained language planning capability, and it has been cited 67 times on

Google Scholar. The organizing committee commented that the work "teaches smaller, more efficient language models how to perform constrained language planning by distilling knowledge from larger models"; it has also been introduced as a representative method by well-known scholars such as IEEE/AAAI/AAAS/ACM Fellow Amit Sheth.

The applicant led the writing of the textbook Knowledge Graph: Concepts and Technologies. Since its first edition in 2020, it has been distributed in more than 40,000 copies, adopted as a textbook by 17 universities, repeatedly ranked among the top three computer book bestsellers on major platforms such as JD.com, described by readers as a "must-read" for learning about knowledge graphs, and licensed for publication in Taiwan, China (officially published in 2021) (see Figure 3). The applicant also built the Knowledge Graph digital-intelligence textbook platform and received the Special Prize for Textbook Development of Fudan University. The book has been listed by the Publishing House of Electronics Industry as a key title for promotion, and the applicant was named an Outstanding Author of the Publishing House of Electronics Industry for 2019. Drawing on the knowledge graph research results, the applicant served as editor-in-chief of Artificial Intelligence (Senior High School) and as a co-editor of Big Data Technology and several other textbooks and teaching references, built the Knowledge Graph online textbook platform to actively promote digital-intelligence textbook technology, and was appointed a member of the technical expert group of the first batch of the Ministry of Education Department of Higher Education's project on education and teaching resources in key fields and new-type textbook development. At the invitation of the Ministry of Education and the education departments of Liaoning, Guangxi, Shaanxi, and other provinces (autonomous regions), the applicant gave five reports, sharing experience with more than 120,000 university teachers and experts nationwide.

III. Application and Deployment of the Research Results

The related results won first place worldwide in the Language and Intelligence Technology Competition jointly organized in 2019 by the China Computer Federation and the Chinese Information Processing Society of China. On this basis, the project further constructed and publicly released a number of large graphs, including the large-scale Chinese encyclopedic knowledge graph CN-DBpedia, supporting a great deal of research on Chinese knowledge graphs. As of July 2026, CN-DBpedia has been downloaded 53,852 times in total across multiple platforms, with more than 1.8 billion API calls. It has not only advanced the field of Chinese knowledge graphs but also reached an internationally leading level with lower update cost and greater timeliness, demonstrating China's technological strengths in graph construction and application. The knowledge graph construction results have been deployed in more than ten core business lines of

Huawei, strongly supporting the intelligent development of Huawei's business lines. An industry-first enterprise-level knowledge graph (containing tens of millions of entities) was built for Huawei, realizing autonomous updating of domain knowledge and a knowledge closed loop; its intelligent analysis model for network operations has been applied in scenarios such as Huawei Digital Finance and Digital Government, generating tens of millions of yuan in economic benefits; and in the group-wide document search and recommendation business of Huawei it achieved notable results, raising search accuracy by 10%, increasing knowledge search satisfaction by more than 30%, and reaching a knowledge recommendation satisfaction rate of 85%. For these contributions, the applicant received honors including the Huawei Outstanding Partner Award and the Huawei Spark Award (see Figure 4), and was selected as one of the "20 Benchmark Figures Empowering China's AI Knowledge Graphs in 2019."

The related results have also been deployed on Meituan's life services platform, where a concept graph of the life services domain at the ten-million scale was built (with an accuracy of 88%, the highest-quality product-level domain concept graph at that time). Conceptual commonsense was used to form cognitive capabilities in the life services domain, improving user search understanding and merchant tag understanding in Meituan's online business and significantly raising the accuracy of automated product tag matching; for this work the applicant received the Meituan Scientific Research Cooperation Innovation Award (see Figure 5). The results were likewise deployed on Alibaba's Fliggy e-commerce platform, where the concept graph guided analysis of platform users' behavior understanding and greatly increased the user click-through rate; an abbreviation generation model and a sentiment triple extraction model for travel scenarios were launched in the Fliggy app search system, increasing the product conversion rate by 2.03% in relative terms. The applicant won the first Alibaba Research Award and the Alibaba Academic Collaboration Best Partner Award (see Figure 6).

The work on large-graph partitioning has been applied in Microsoft's graph database system Trinity (the predecessor of Graph Engine, Microsoft's distributed large-scale graph data processing engine), providing infrastructure support for Microsoft's knowledge graph ProBase and receiving broad attention and recognition from academia and industry. The project team applied the related technologies in collaborative projects with leading enterprises in traditional industries such as AVIC and the State Grid Corporation of China, supporting intelligent services such as customer service question answering and knowledge recommendation for the industries and enterprises concerned. The related technologies attracted the attention of a number of enterprises and investors to the commercialization of the project results, and the project team, together with Shanghai AISHU

Information Technology Co., Ltd., established China's first university-enterprise joint laboratory for cognitive intelligence oriented toward intelligent content management.



Figure 1. Large-scale graph-structured data management received the Second Prize of the Natural Science Award of the Ministry of Education



Figure 2. Research on large-scale knowledge graph construction, reasoning, and application received the First Prize of the 2025 Science and Technology Award of the Shanghai Computer Society (Natural Science Award)



Figure 3. The textbook Knowledge Graph: Concepts and Technologies and its edition introduced and published in Taiwan, China



Figure 4. The intelligent analysis model for network operations applied in scenarios such as Huawei Digital Finance and Digital Government, winning the Huawei Outstanding Partner Award and the Huawei Spark Award



Figure 5. Work such as large-scale fine-grained concept extraction, abbreviation prediction, product hypernym-hyponym relation completion, and compound noun explanation based on the commonsense knowledge graph applied in Meituan's life services domain,



Figure 6. The abbreviation generation model and the sentiment triple extraction model for travel scenarios launched in the Fliggy app search system, winning the Alibaba Academic Collaboration Best Partner Award and Alibaba Research Fellow

winning the Meituan Scientific Research Cooperation Innovation Award	