

Workshop on Enterprise Knowledge Graphs using Large Language Models

Rajeev Gupta¹, Srinath Srinivasa²

¹Microsoft Search, Assistant, and Intelligence Microsoft India Pvt. Ltd, Hyderabad, India

²Professor and Dean (R&D), Web Science Lab, IIT-Bangalore, India

Abstract

Knowledge graphs are used for organizing and connecting individual entities to integrate the information extracted from different data sources. Typically, knowledge graphs are used to connect various real-world entities like persons, places, things, actions, etc. For the knowledge graphs created using the enterprise data, the knowledge graph entities can be of different types—static entities (e.g., people, projects), communication entities (e.g., emails, meetings, documents), derived entities (e.g., rules, definitions, entities from emails), etc. The graphs are used to connect these entities with enriched context (as edges and node attributes) and used for powering various search and recommendations applications.

With the advent of large language models, the whole lifecycle of knowledge graphs involving –information extraction, graph construction, application of graphs, querying knowledge graphs, using the graph for recommendations, etc., – is impacted. With large language models such as GPT, LLaMA, PALM, etc., entity and relationship extraction can be improved. Similarly, one can answer different types of queries with the help of LLMs which were very difficult without them. This workshop is about improving the enterprise knowledge graphs and its applications using large language models.

Enterprise graphs can be of different scopes—whether they contain data from individual users/customers, a sub-organization, or the whole enterprise. This workshop will also cover various privacy and access control related issues which are typical for any enterprise graph. These include privacy preserving federated learning, using LLMs to extract information from private data, querying the knowledge graph in a privacy preserving manner, etc.

Keywords

Knowledge Graph, Large Language Model, Entity Extraction, Relationship Extraction, Recommendations


1. Workshop Goals


Knowledge graphs can integrate diverse data sources and provide a holistic view to the downstream applications. By virtue of being structured, knowledge graphs offer transparency and interpretability to the search and recommendations applications. As per one prediction, this connected data with semantically enriched context applications and graph mining will grow 100% annually. This workshop is about creating and using knowledge graphs on the enterprise data. This data is the internal data of the enterprises—of their employees and/or their customers. Unlike the graphs of open-web entities, enterprise knowledge graphs (EKG) connect the entities specific to the enterprises. For example, all the employee emails, meetings, documents, projects, etc., can be used to create a graph and this graph can be used to summarize the interaction between two employees, identify close collaborators, identify documents which should be attached to an email, documents associated with a project, etc. Similarly, there can be a knowledge graph of items,

suppliers, teams, regions, etc., and the graph can be used to recommend suppliers for a particular requirement. In this workshop we will be covering how large language models (LLMs) can help with the construction and usage of these enterprise knowledge graphs (EKGs). This involves improving all the aspects of EKG workflow using large language models: entity extraction, entity enrichment, EKG construction, querying EKG for search and recommendations, scenario specific EKG, etc. Besides the well-known challenges associated with the knowledge graphs, EKGs have other issues—how to extract entities from private enterprise data? how to use large language models in a privacy aware manner? how to create relationships between different entities while preserving privacy? how to create EKG with internal (e.g., employees) data and external (e.g., suppliers) data? how is access control maintained in an EKG where data is from different divisions of the enterprise? how are the enterprise recommendations application different compared to, say, movie or a product recommendations? how these recommendations applications benefit from EKG with LLM? how can one integrate EKG with large language models for a particular application? etc. To ensure privacy and separation of access one may need to use federated graph learning while developing applications over EKGs. How to use federated learning in large lan-

Workshop on Enterprise Knowledge Graphs using Large Language Models, Oct 22, 2023, Birmingham, UK

✉ rajeev.gupta@microsoft.com (R. Gupta); sri@iiitb.ac.in (S. Srinivasa)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

guage models? Through this workshop we would like to highlight research issues specific to the integration of the enterprise knowledge graphs with large language models and associated applications. We will be bringing the researchers (from academia as well as industry) and practitioners (mainly from industry) together to achieve that.

2. Workshop Topics

- Designing Enterprise Knowledge Graph (EKG)
- Scalable extraction of enterprise entities using LLMs
- Relationship extraction using LLMs
- Natural Language Processing (NLP) algorithms to build EKGs
- Semantic reasoning based on deep learning on graph
- Explainable AI based on EKG
- EKG Implementation issues
- Federated graph learning with LLMs
- Privacy in graph algorithms
- Privacy preserving graph construction and mining
- Use of EKG and LLMs for search and recommendations
- Building EKGs for specific domains or applications
- Industrial applications of EKGs: finance, retail, healthcare, etc.

3. Workshop Program

The workshop includes five invited talks, which are listed below.

1. Nishanth Sastry. *Lessons from the Age of User-Generated Content for the Age of AI-Generated Content*
2. Fabio Petroni. *Can Machines Discover New Knowledge?*
3. Liana Mikaelyan. *Enhancing Enterprise Knowledge Base Construction with Fine-Tuned Generative Language Models*
4. Bojan Babic. *LLMs for Social Networks: Applications, Challenges and Solutions*
5. Manoj Agarwal. *Building Knowledge Graph for Products at Scale and Infusing it in to LLMs*

Here is the list of 5 papers presented at the workshop:

1. Anurag Mohanty. *EduEmbedd – A Knowledge Graph Embedding for Education.*
2. Pranav Subramaniam, Udayan Khurana, Kavitha Srinivas and Horst Samulowitz. *Related Table Search for Numeric data using Large Language Models and Enterprise Knowledge Graphs.*
3. Apurva Kulkarni, Chandrashekar Ramanathan and Vinu E Venugopal. *Cognitive Retrieve: Empowering Document Retrieval with Semantics and Domain Specific Knowledge Graph.*
4. Mohit Sewak, Vamsi Emani and Annam Naresh. *CRUSH: Cybersecurity Research using Universal LLMs and Semantic Hypernetworks.*
5. Elan S Markowitz, Aram Galstyan. *StATIK+: Structure and Text for Inductive Knowledge Graph Modeling and Paths towards Enterprise Implementations*

4. Program Committee

Members of the Program Committee are listed below,

- **Manoj Agarwal**
(Senior Researcher, Discovery Intelligence, Uber Research)
- **Manish Bhide**
(CTO, AI Governance, IBM)
- **Mukesh Mohania**
(Professor, CSE, IIT Delhi)
- **Prasad Deshpande**
(Senior Staff Software Engineer, Databricks)
- **Qi He**
(Head of AI, Nextdoor)
- **Ranganath Kondapally**
(Principal Applied Scientist, Microsoft)
- **Rushi Bhatt**
(Partner, ML Systems and Services, Microsoft)
- **Sauvik Ghosh**
(Director of AI, LinkedIn)

5. Target Audience

We will involve people from Industry as well as academia. The practitioners and researchers from industry are likely to present their domain, graphs they are building using LLMs, for various applications, whereas folks from academia are likely to identify research problems of common interest and advice appropriately.

6. Workshop Organizers

6.1. Workshop Chairs

- **Rajeev Gupta**
Rajeev Gupta is a Principal Applied Scientist at Microsoft Search Assistant & Intelligence (MSAI), India. He got his PhD from Indian Institute of Technology (IIT) Mumbai (Bombay) in the area of distributed data management. He has more than 35 publications and 20 patents in the areas of data management, information extraction, machine learning, and distributed computing in reputed conferences and journals such as VLDB, KDD, WWW, SIGMETRICS, TKDE, ICDE, etc. He has given tutorials, in various reputed conferences like KDD, ECIR, AIML, etc., in the areas of streaming data management, privacy

aware information management, enterprise entity extraction, etc. Currently he is involved in building various graph technologies involving graph walk algorithms, representation learning, edge prediction in private data, etc., to power a number of search and recommendations applications. He has been involved in organizing various conferences and workshops, e.g., ICDE, DASFAA, BDA, COMAD, etc., and programing committee member for KDD, IJCAI, ICDE, VLDB, etc., during various years.

- **Srinath Srinivasa**

Srinath Srinivasa heads the Web Science lab and is the Dean (R&D) at IIT Bangalore, India. Srinath holds a Ph.D. (magna cum laude) from the Berlin Brandenburg Graduate School for Distributed Information Systems (GkVI) Germany, an M.S. (by Research) from IIT-Madras and B.E. in Computer Science and Engineering from The National Institute of Engineering (NIE) Mysore. He works in the area of Web Science – that models the impact of the web on humanity. Technology for educational outreach and social empowerment has been a primary motivation driving his research. He has participated in several initiatives for technology enhanced education including the VTU Edusat program, The National Program for Technology Enhanced Learning (NPTEL) and an educational outreach program in collaboration with Upgrad. He is a member of various technical and organizational committees for international conferences like International Conference on Weblogs and Social Media (ICWSM), ACM Hypertext, COMAD/CoDS, ODBASE, etc. He has served as a technical reviewer for various journals like the VLDB journal, IEEE Transactions on Knowledge and Data Engineering, and IEEE Transactions on Cloud Computing. He is also the recipient of various national and international grants for his research activities.

6.2. Website and Proceedings Chairs

- **Bhoomika A P**
PhD Scholar, Web Science Lab, IIT-Bangalore
- **Aparna M**
M.S. by Research Scholar, Web Science Lab, IIT-Bangalore

References

- [1] S Pan, L Luo, Y Wang, C Chen, J Wang, X Wu. *Unifying Large Language Models and Knowledge Graphs: A Roadmap*, arXiv preprint arXiv:2306.08302, 2023.
- [2] M Trajanoska, R Stojanov, D Trajanov. *Enhancing Knowledge Graph Construction Using Large Language Models*. arXiv preprint arXiv:2305.04676, 2023.
- [3] S Carta, A Giuliani, L Piano, AS Podda, L Pompianu, SG Tiddia. *Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction*. arXiv preprint arXiv:2307.01128, 2023.
- [4] L Yang, H Chen, Z Li, X Ding, X Wu. *ChatGPT is not Enough: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling*. arXiv preprint arXiv:2306.11489, 2023.
- [5] H Khorashadizadeh, N Mihindukulasooriya, S Tiwari, J Groppe, S Groppe. *Exploring In-Context Learning Capabilities of Foundation Models for Generating Knowledge Graphs from Text*. arXiv preprint arXiv:2305.08804, 2023.