

# Enterprise Knowledge Graph

## From Specific Business Task to Enterprise Knowledge Management

### I. ABSTRACT

Data driven Knowledge Graph is rapidly adapted by different societies. Many open domain and specific domain knowledge graphs have been constructed, and many industries have benefited from knowledge graph. Currently, enterprise related knowledge graph is classified as specific domain, but the applications span from solving a narrow specific problem to Enterprise Knowledge Management system. With the digital transform of traditional industry, Enterprise knowledge becomes more and more complicated, it involves knowledge from common domain, multiple specific domains, and corporate-specific in general. This tutorial provides an overview of current Enterprise Knowledge Graph(EKG). It distinguishes the EKG from specific domain according to the knowledge it covers, and provides the examples to illustrate the difference between EKG and specific domain KG. The tutorial further summarizes EKG into three types: Specific Business Task Enterprise KG, Specific Business Unit Enterprise KG and Cross Business Unit Enterprise KG, and illustrates the characteristics, steps, challenges, and future research in constructing and consuming of each of these three types of EKG .

**Length:**This is a Half day tutorial (3 hours)

#### **Tutor:**

**Rong Duan**, In person presenter, *Huawei Technology, Shenzhen, China*,

email:rong.duan@huawei.com, phone:18244975822;

**Yanghua Xiao**, In person presenter, *Fudan University, Shanghai, China*,

email:shawyh@fudan.edu.cn;

### II. TARGET AUDIENCE, PREREQUISITES, AND BENEFITS

**Target audience:** This tutorial is designed for practitioners and researchers who are interested in knowledge graph, and willing to explore the techniques, and challenges in Enterprise Knowledge Graph.

**Prerequisites:** Participants will be expected to have basic knowledge of machine learning and knowledge graph

**Benefits:**There are a few reasons that the participants will benefit from this tutorial. On one side, it provides a review to practitioners in CIKM community about what is EKG, what are the differences among open domain,specific domain, and enterprise KG, what are the differences among different types

of EKG, and how to design and construct different types of EKG. And on the other side, it introduces new challenges to researchers in the community to solve the EKG problem that different from open and specific domain. Knowledge graph has been a hot topic in the past few years in many societies, But not many focus on enterprise knowledge graph. With the digital transformation, many industries, especially traditional industry have collected a large volume of data. Besides volume increase, multiple data format, diverse information storage systems,new quality challenges, developing knowledge base, the interactive knowledge consumption, and the multiple user groups with multiple perspectives are the new challenges that EKM is facing. The challenges involve topics from heterogenous data sources, knowledge base system, knowledge management system, etc, which are the interesting topics of this community. The last reason is that there is no others proposed similar tutorial as presenters' knowledge.

### III. OUTLINE

A tentative outline of the tutorial looks as follows

- **Part 1.** Enterprise Knowledge and Enterprise Knowledge Graph (60 minutes)

Many efforts have been done on enterprise knowledge. From the types of enterprise knowledge to EKM construction. [1] discusses different enterprise knowledge models, [2] illustrates enterprise knowledge creation model, [3] defines the top layer of enterprise ontology, and many expert-driven ontology based EKM systems [4] [5] have been proposed. With the digital transformation and AI-enabled information system, enterprise knowledge becomes more and more complicated. Besides volume increase, multiple data format, diverse information storage systems,new quality challenges, developing knowledge base, the interactive knowledge consumption, and multiple perspectives from different user groups are the new challenges that modern EKM system is facing. The modern EKM should be powerful enough to deal with large volume data, robust enough to adopt information updates, rigorous enough to control the quality of the information, and intelligent enough to interact with different groups of users. The predefined-ontology based EKM could not fully satisfy modern EKM requirements. Data driven Knowledge Graphs have been receiving more and more attention in enterprise. [6] proposes a concept framework in

constructing EKM systems with EKG, [7] illustrates three layers in constructing KG in large organization. Many industries have benefit from EKG [8] [9] [10]. To emphasize the characteristic of enterprise knowledge, and base on the knowledge coverage, this tutorial distinguish EKG from open, and specific domain [11] [8], and category EKG into three types: Specific Business Task EKG, Specific Business Unit EKG and Cross Business Unit EKG. This section discusses the characteristics of these three types of EKG, and challenges to construct different types of EKG through examples.

#### Breaks 15 minutes

- **Part 2.** Construction of Enterprise Knowledge Graph (60 minutes)

The section provides review of EKG construction techniques, which includes entity recognition [12] [13] [14] [15] [16], relation extraction [17] [18] [19] [20] [21] [22], and KG refinement [23]. The entity recognition includes Phrase Mining, Abbreviation Extraction, Synonyms Mining and Named Entity Recognition; The relation extraction covers general relation extraction and isA relation extraction; Knowledge refinement includes KG correction, and KG completion. Besides the general techniques involved in constructing EKG, this section will emphasize more on advantage and limitation of each technique in constructing different types of EKG, and techniques specialize in constructing EKG. For example, the techniques to integrate rich expert knowledge with data driven method in KG expansion.

- **Part 3.** Challenges and future research in Enterprise Knowledge Graph (45 minutes). The last section will discuss the challenges and Potential future research directions in EKG field. The challenge includes strategic challenge and technical challenge. For technical challenge, it will discuss the challenges in the availability of corpora, training data issue, validation problem, cold start problem, expert knowledge integration and the multiple perspective problems.

#### IV. A LIST OF FORUMS

The tutors present a similar tutorial at the 24th International Conference on Database Systems for Advanced Application (DASFAA2019) at Chiang Mai, Thailand. The whole structure will keep the same, but new material will be added, the detail is as following:

- The enterprise knowledge characteristics and modern enterprise knowledge management part will be the same. Examples will be added when introduce three types of EKG and their challenges.
- The general technique review is similar, but will emphasize more on the challenges in different types of EKG,

and techniques to solve these problems. New techniques to integrate expert knowledge with data driven method will be added.

- In the future research part, the challenges of maintaining/updating KG with multiple user groups from multiple perspectives will be added.

There are many knowledge graph related tutorials in data mining, machine learning and AI conferences in these few years. Most of them are for open domain knowledge graph, which includes knowledge graph basic concepts and construction techniques. For example:

- **KDD 2014** "Constructing and mining web-scale knowledge graphs", Antoine Bordes, Evgeniy Gabrilovich
- **WWW 2015** "Knowledge Bases for Web Content Analytics", Johannes Hoffart, etc,
- **AAAI 2017** "Mining Knowledge Graphs from Text" JAY PUJARA, SAMEER SINGH, BHAVANA DALVI,
- **KDD 2018** "Building a Large-scale, Accurate and Fresh Knowledge Graph", Yuqing Gao, etc,

There are also tutorials focus on specific domain, but only focus on the corpora from web.

- **AAAI 2018** "Knowledge Graph Construction from Web Corpora" Mayank Kejriwal, Craig Knoblock, Pedro Szekely

By presenters knowledge, there is no tutorial focus on Enterprise Knowledge Graph.

#### V. SHORT BIO

**Dr. Rong Duan** currently is Chief Data Scientist of Corporate Data Management Department, Huawei Technologies Co., Ltd. She is guiding a group to build various AI applications in supporting internal users. Multiple enterprise knowledge graphs have been constructed under her supervision. Before joined Huawei, Rong was principal inventive scientist at AT&T Labs, big data research and adjunct professor at Stevens Institute of Technology. Has been working in AT&T Labs for more than 20 years, Rong has extensive experience on statistical learning, data mining, predictive modeling and data analysis for business data. Rong received her PhD and MSc in computer engineering and computer science respectively from Stevens Institute of Technology, US. Her research interests include data mining, statistical learning theory and methods, Spatial-temporal Risk Assessment and management, Data Integration and Data Quality Assessment.

Rong was former Chair for the Data Mining Section of INFORMS, and Data Mining cluster co-chair for INFORMS International Beijing. Rong also served as a program co-chair for the 1st and the 2nd International Symposium on System Informatics and Engineering, Panelist for ICDM, tutor for DSAA, DASFAA, INFORMS, etc.

**Dr. Yanghua Xiao** got his Ph.D. degree from Fudan University, Shanghai, China, in 2009. He now is a full professor of the School of Computer Science at Fudan University. His research interest includes big data management and mining,

graph database, knowledge graph. He was a visiting professor of Human Genome Sequencing Center at Baylor College Medicine, and visiting researcher of Microsoft Research Asia and Alibaba. He won 10+ research awards granted by governments or industries, including CCF Natural Science Award (second level), ACM(CCF) Shanghai distinguished young scientists and Alibaba Research Fellowship Award. Recently, he has published 100+ papers in international leading journals and top conferences, including TKDE, SIGMOD, VLDB, ICDE, IJCAI, AAAI. He is the PI or Co-PI of 30+ projects supported by 10+ national and local funding agency and big companies including Microsoft, IBM, HUAWEI, China Telecom, China Mobile, Baidu, XiaoI Robot etc. He regularly serves as the reviewer of 10+ national and local funding agencies and SPC or PC members of 50+ top conferences including IJCAI, AAAI, SIGKDD, ICDE, WWW, CIKM, ICDM, SDM etc. He is the Associate Editor of *Frontier of Computer Science*, and reviewers of 20+ leading journals. He is a member of ACM, IEEE, AAAI and senior member of CCF. He is the director of Knowledge Works4 Research Laboratory at Fudan University. His team at Fudan published a lot of Chinese knowledge graphs, which serve industries with 1 billions of API calls. He is the chief scientist or senior advisors of many top Chinese big data companies or AI companies. He has ever given more than 10 keynote speeches or tutorials in international leading conference including SIGMOD2012, IDEAL2017, ADMA2018, PIC2017, CCKS2018 etc.

## REFERENCES

- [1] E. Simperl, S. W?lger, B. K?mpgen, S. Hangl, D. Vrandecic, and K. Siorpaes, *Enterprise Knowledge Structures*, P. W. et al., Ed. Berlin Heidelberg: Springer-Verlag, 2011.
- [2] I. Nonaka, R. Toyama, and N. Konno, "SECI, ba and leadership: a unified model of dynamic knowledge creation," *Long Range Planning*, vol. 33, no. 1, pp. 5–34, feb 2000. [Online]. Available: <https://doi.org/10.1016%2Fs0024-6301%2899%2900115-6>
- [3] M. Uschold, M. King, S. Moralee, and Y. Zorgios, "The enterprise ontology," *Knowledge Engineering Review*, vol. 13, no. 1, pp. 31–89, 1998.
- [4] P. Bernus and L. Nemes, "A framework to define a generic enterprise reference architecture and methodology," *Computer Integrated Manufacturing Systems*, vol. 9, no. 3, pp. 179–191, 1996.
- [5] A. Maedche, B. Motik, L. Stojanovic, R. Studer, and R. Volz, "Ontologies for enterprise knowledge management," *IEEE Intelligent Systems*, vol. 18, no. 2, pp. 26–33, March 2003.
- [6] M. Galkin, S. Auer, M.-E. Vidal, and S. Scerri, "Enterprise knowledge graphs: A semantic approach for knowledge management in the next generation of enterprise information systems," in *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)*, vol. 2. Porto, Portugal: SciTePress, April 2017, pp. 88–98.
- [7] J. Z. Pan, G. Vetere, J. M. Gomez-Perez, and H. Wu, *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, 1st ed. Springer Publishing Company, Incorporated, 2017.
- [8] P. Cotroneo, W. Haak, G. Oscars, E. Presani, A. Rohatgi, and P. T. Groth, "From data search to data showcasing: The role of semantic technologies in a new service," in *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018)*, Monterey, USA, October 8th - to - 12th, 2018., 2018. [Online]. Available: <http://ceur-ws.org/Vol-2180/paper-80.pdf>
- [9] T. Hubauer, S. Lamparter, P. Haase, and D. M. Herzig, "Use cases of the industrial knowledge graph at siemens," in *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018)*, Monterey, USA, October 8th - to - 12th, 2018., 2018. [Online]. Available: <http://ceur-ws.org/Vol-2180/paper-86.pdf>
- [10] J. Aasman and P. Mirhaji, "Knowledge graph solutions in healthcare for improved clinical outcomes," in *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018)*, Monterey, USA, October 8th - to - 12th, 2018., 2018. [Online]. Available: <http://ceur-ws.org/Vol-2180/paper-78.pdf>
- [11] M. Kejriwal, *Domain-Specific Knowledge Graph Construction*. Springer, 2019.
- [12] K. Frantzi, S. Ananiadou, and H. Mima, "Automatic recognition of multi-word terms: the c-value/nc-value method," *International Journal on Digital Libraries*, vol. 3, no. 2, pp. 115–130, Aug 2000.
- [13] J. Shang, J. Liu, J. Meng, R. Xiang, C. R. Voss, and J. Han, "Automated phrase mining from massive text corpora," *IEEE Transactions on Knowledge & Data Engineering*, vol. PP, no. 99, pp. 1–1, 2017.
- [14] J. Liu, J. Shang, C. Wang, and X. R. J. Han, "Mining quality phrases from massive text corpora," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 2015, pp. 1729–1744.
- [15] A. Elkishky, Y. Song, C. Wang, C. Voss, and J. Han, "Scalable topical phrase mining from text corpora," *Proceedings of the Vldb Endowment*, vol. 8, no. 3, pp. 305–316, 2014.
- [16] Z. Huang, W. Xu, and K. Yu, "Bidirectional lstm-crf models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.
- [17] E. Agichtein and L. Gravano, "Snowball: Extracting relations from large plain-text collections," in *Proceedings of the fifth ACM conference on Digital libraries*. ACM, 2000, pp. 85–94.
- [18] D. Liu, "Dependency-based evaluation of minipar," in *Treebanks*. Springer, 2003, pp. 317–329.
- [19] H. Jiang, C. Zhang, D. Yang, Y. Xiao, J. Liu, J. Chen, C. Wang, C. Li, J. Liang, B. Liang, and W. Wang, "Hierarchical conceptual labeling," in *Database Systems for Advanced Applications - 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22-25, 2019, Proceedings*, 2019, pp. 230–234.
- [20] J. Chen, A. Wang, J. Chen, Y. Xiao, Z. Chu, J. Liu, J. Liang, and W. Wang, "Cn-probase: A data-driven approach for large-scale chinese taxonomy construction," in *35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China, April 8-11, 2019*, 2019, pp. 1706–1709.
- [21] L. Chen, J. Liang, C. Xie, and Y. Xiao, "Short text entity linking with fine-grained topics," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, 2018, pp. 457–466.
- [22] J. Liang, Y. Xiao, Y. Zhang, S. Hwang, and H. Wang, "Graph-based wrong isa relation detection in a large-scale lexical taxonomy," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, 2017, pp. 1178–1184.
- [23] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic web*, vol. 8, no. 3, pp. 489–508, 2017.