Podcast Name: *ACM ByteCast*
Episode: Xin Luna Dong

Welcome to the *ACM ByteCast* podcast, a series from the Association for Computing Machinery! The podcast features conversations with researchers, practitioners, and innovators at the intersection of computing research and practice about their experiences, lessons learned, and visions for the future of computing. In this episode, host Bruke Kifle is joined by Dr. Xin Luna Dong, a principal scientist at MetaReality Labs, who has significantly contributed to the development of knowledge graphs, a tool essential for organizing data into understandable relationships. Luna has been a leader in machine learning applications, working on intelligent personal assistants, search, recommendation, and personalization systems. Her journey spans roles at Amazon, Google, AT&T Labs, and academia, earning her recognition as an ACM and IEEE Fellow.

To begin, Luna shares how her early experiences in China sparked her interest in computing. Her first encounter with a computer was through her mother, who introduced her to basic games on a COM35 and an Apple II. This led her to experiment with simple coding tasks, despite language barriers. As she progressed, she learned about foundational algorithms, such as the A* algorithm, which deepened her understanding and fascination with computational intelligence. By high school, she was committed to studying computer science, eventually completing her bachelor's, master's, and PhD in the field.

During her PhD, Luna worked with a professor specializing in the field of data integration. The work aligned with her interest in structuring and organizing data, laying the groundwork for her future work with knowledge graphs. Her experience in data integration naturally extended into knowledge graphs, particularly when she joined Google, where she first encountered large-scale knowledge organization. Luna explains the relevance and structure of knowledge graphs, which consist of entities and the relationships between them, mimicking how humans understand the world. Knowledge graphs are valuable because they present information in a structured, highly accurate form. In Google and Bing search engines, knowledge panels provide easily accessible information, and at Amazon, they improve product information for digital media and shopping. Luna emphasizes that a fundamental aspect of knowledge graphs is the necessity for clean, high-quality, and high-fidelity data.

Next, Luna outlines the progression of data integration methodologies over the past two decades. Initially, runtime data integration was prevalent, exemplified by web searches where users input queries and receive results from multiple sources in real-time. The advent of knowledge graphs introduced a second generation of data integration, focusing on offline data assembly and normalization. With the rise of machine learning and AI, a third generation of data integration has emerged, which Luna calls "knowledge internalization." While large language models have revolutionized natural language processing with their ability to generate coherent and contextually relevant responses, they possess limitations. Knowledge graphs can enhance LLMs by supplying up-to-date and detailed data, thereby improving the quality and reliability of responses in applications such as search engines and personal assistants.

Looking ahead, Luna envisions knowledge graphs playing a pivotal role alongside LLMs in the evolving tech landscape. She highlights several emerging technologies and trends that hold promise for enhancing the creation and application of knowledge graphs. One such area is Retrieval-Augmented Generation (RAG), which integrates knowledge graphs with LLMs to retrieve relevant information from structured data sources, thereby improving answer generation and recommendation systems. Luna mentions her work on the Comprehensive RAG Benchmark (CRAC) and the KDD CUP competition, which aim to evaluate and advance the state-of-the-art in RAG systems. Luna also shares her passion for making information access effortless. She envisions solutions that allow non-technical users, such as small business owners, to easily serve their data through knowledge graphs without the need for complex QA systems. Emerging technologies like multimodal AI, which can handle various data types and contexts, are also of great interest to her. These advancements will facilitate more personalized and accurate information delivery, aligning with her mission to provide the right information at the right time. She concludes by emphasizing the ongoing need to improve knowledge graph creation and integration, suggesting that continued progress in these areas will significantly enhance how people access and interact with information in the digital age.

Key Takeaways:
1:52 - Luna's early Interest in computing.
6:25 - Path to knowledge graphs and data integration.
10:35 - Impact and uses of knowledge graphs.
16:13 - Challenges in data extraction and integration.
16:56 - Evolution of data integration methods.
19:41 - Complementarity between knowledge graphs and Large Language Models.
24:13 - The future of knowledge graphs in the tech landscape.
29:05 - Improving the way humans interact with data and technology.
33:40 - Exciting upcoming developments in the tech space.

Links
Learn more about Dr. Xin Luna Dong.
Learn more about the ACM ByteCast podcast at https://learning.acm.org/bytecast