AI LEADERS PODCAST
EP 69 KNOWLEDGE GRAPHS IN THE AGE OF GEN AI
TRANSCRIPT

Navin Sharma [00:00:00] At the end of the day, we and all the graphs are all about, like I said, encapsulating information about things, things we understand on our day-to-day basis.

Teresa Tung [00:00:20] Hi, welcome to another Accenture A.I. Leaders podcast. My name is Teresa Tung. I am Accenture's Global Data Capability Lead. And this session is about knowledge graphs. This topic is very important as we talk about enterprise data and especially in the age of Gen AI, I'm so pleased to be joined by our guests who are experts not just in knowledge graphs, but what it means to look at knowledge graphs with enterprise data. So please join me in welcoming them. But I would like each guest to introduce themselves in their role. So, Tony, let's start with you.

Tony Romito [00:00:56] Thanks, Teresa. Hi, everyone. Tony Romito. I lead our global digital and AI asset portfolio for Accenture's Products Industry Group and been in the industry over 25 years and really excited about this new horizon of technology that's upon us. Thanks for having me.

Teresa Tung [00:01:14] Thanks, Tony. Tony has been applying knowledge graphs at some of his clients for a while now. Also joined by Navin, one of our leads, one of our strategic minority investments, Stardog. And Navin, describe a little bit more about your role.

Navin Sharma [00:01:32] Thank you, Teresa and hello, everyone. Navin Sharma, the head of product at Stardog. I've spent my career plus 20 plus years in the enterprise software and B2B domain, 15 of which have this has been in the product management space centered around data management holistically. Excited to be here and look forward to the conversation.

Teresa Tung [00:01:57] Thank you both so much for joining. So, I wanted to start off with just what are knowledge graphs, right? So, I think, Tony, let's start with you. Right? I mentioned that you've been working with a lot of companies in the data space and specifically with knowledge graphs. What do companies need to know about knowledge graphs?

Tony Romito [00:02:18] Yeah, I think as you set up in the beginning, Teresa, the concept of knowledge graphs has been around for decades. But I think what's making it hyper relevant right now is a convergence of three things. First, there's new technology that's enabling us to rapidly create and use knowledge graphs at an enterprise scale. The second thing is that the scope of data and sophistication of hardware makes the use of those knowledge graphs at scale much more attainable.

And the third thing is with advanced A.I. knowledge graphs are fundamental to creating more accurate and relevant context that just isn't inherent in large language models. And so, what we see organizations as are trying to harness the value of that untapped data in their enterprise is that they have a growing need to relate data to one another. They need to do that in a way that's constantly changing without re architecting data structures or frameworks. And they need to increasingly apply reasoning to the use of that data. So that requires a deeper understanding of what the data means in the context of the industry and the business function. And that's fundamentally where knowledge graphs become so powerful.

Teresa Tung [00:03:37] Yeah, I agree. Right. So, you're saying that any company who has siloed data and that's pretty much everybody, right? You have data that you can't access because it's siloed. And a lot of it is because you don't know how to apply it. So, I think knowledge graph is one for those sorts of use cases. And this is increasingly important in the age of generative AI, where it's the generative AI model really needs to be able to add your domain context to be relevant for your use case at all. And so being able to have the knowledge graph allow an automated system like generative AI, find what's relevant is key. I think. Navin, could you share a definition of knowledge graphs and then share some examples about why there needed in enterprise data?

Navin Sharma [00:04:25] Yeah, in fact, Google popularized knowledge graphs, even though the knowledge graphs have been around for some time in the early 2000 when they evolved their search engine to not only just look at keyword searches but also represent things that are related, related ideas or related concepts or people. And so, when you looked at the Google website, you would see on the right-hand side, it would manifest the knowledge graph, and the information connected to the actual information you are looking for. And there's lots of definitions out there.

My favorite one tends to be thinking about knowledge graphs is a representation of how we think about the world as things and not necessarily how data is stored and structured as strings. And a lot of these information repositories inside of organizations. So, the idea of concepts, business concepts that we can all relate to, like people, places, order suppliers, fonts, SKUs that representation and the relationship across those business concepts is key to how you start to describe data and attach meaning to it. And that's really the manifestation of what a knowledge graph is in a lot of these organizations. And so, knowledge graphs really fit that mold where the focus of knowledge graphs tend to be on how information needs to be consumed rather than how information needs to be produced or structured or stored. And so representing data at a business level, at a business concept level, that all users inside the enterprise, they don't have to have necessarily a technical background, understand what they need and what they're looking for and have that data represented in that way only makes it possible for them to be in this position where we can finally talk about democratizing data, making data available from a self-service perspective, which I know when we get into the whole this discussion is just even more is almost closing that last mile even further.

Teresa Tung [00:06:36] Yeah, I think it's that elevation, as you mentioned, from data into knowledge, right? So, data itself, those are bits and bytes. The knowledge is really what's contextualizing what you need to see. So, an enterprise will have customer data, but the knowledge representation as to what do I actually need to see if I'm looking at from the perspective of a prospective customer that I'm trying to sell something to versus a customer that I'm doing support on.

It'll be a different view, right, and a different contextualization of that same data. So that's one of the hardest parts to capture in the knowledge graph captures that, right? It's capturing that human expertise that's able to make that determination that says this is the data I need for the problem at hand. And this is actually how I want to use it. So, Tony, why are knowledge graphs especially important now with Gen AI.

Tony Romito [00:7:34] This is one of the topics I'm probably most passionate about because we've all read stories about large language models and the issues with hallucination. And much has been done to fine tune models to adapt those models to improve accuracy, to use techniques like Bragg, to leverage private data and guide the paths to make the results available and more accurate. Yet we still see examples and headlines about large language model recommends using glue to stick cheese on pizza. Right? It's hilarious. But it's also frightening when you think about the implications in a more serious context. And so, it's important to understand why that is and how knowledge graphs to be a solution to that problem. And the reality is, most people should know large language models use statistics essentially to assemble combinations of words to generate a response. It doesn't have a semantic understanding of that context and that language. And so technically, it's accurate to say that glue increases adherence of substances to food. And so, in that respect, the models, right. However, you know, where knowledge graph comes in is it brings context. So, pizza and cheese or food, foods are edible, glue and cheese are both sticky, but glue is nonedible. And so just understanding that allows us to bind it together. And this is where answering your question, there's this notion of, you know, graph rag or knowledge graph assisted rag where you get the strength of both, and that ultimately creates more effective AI. because essentially the knowledge graph is the part taking artificial away from artificial intelligence is creating a knowledge led, responsible intelligence. That is what we need as we create those more advanced agentic use cases that you were mentioning in your example.

Teresa Tung [00:09:32] Yeah. Tony a great example with the glue and the gees are usually some people do eat glue they shouldn't. But you know, having that domain modeling is one way that companies can make sure that their use of LLMs and you would capture that within a knowledge graph is more accurate and more explainable. Navin, can you describe how you use knowledge graphs to improve accuracy with large language models?

Navin Sharma [00:10:01] Yeah. I mean, I think the stories, as scary as they are, I mean, they pale in comparison to how you, you look, you think about it from the response that comes back. And it's not just a here's a response to your question, but that you actually end up with some of these large language models citing sources that are nonexistent. Right. So just that area where you're now beginning to look at. Is that truly the case, that these are actual real sources that I can rely on and make decisions on? And that's, I think, where knowledge graphs certainly play a role beyond what large language models bring in terms of the sort of the general knowledge base. Right. You're now encapsulating your enterprise knowledge right within the domain model that you mentioned, Teresa, about a specific area of the business, about a specific functional area, domain, multiple domains that you want to be able to bring together and represent that at a level that encapsulates all the key business concepts and relationships. So that when a question does get asked, the grounding of the answers is based on what you've encapsulated within that domain model.

So, if a question gets asked that falls outside of that, that grounding of that domain model, then it's always good to say, sorry, I don't understand. I don't have information about that. It's because it's not represented within my domain knowledge that you have encapsulated within the knowledge graph. So that's one way we've seen knowledge graphs really begin to improve the accuracy. Then the second thing is there's a reliance on large language models to actually provide be the final arbiter of the answers that come back. Right. So, whether you talk to the large language model directly or you use an even a rag implementation approach, even within a rag where your retrieval augmented generation, you're still very much relying on the final answer to come from the large language models as being the final arbiter of the response. So, the answer is it has to be no hallucinations. And when you stop the reliance on large language model from final answer perspective and reference your actual enterprise sources of truth, which is another approach that we take that's slightly different than your traditional rag is we use the large language model to generate a structure query for us. But then at the end of the day, the answers come from your sources of truth, your data lakes, your data warehouses, your application that are, there's no source of truth for the answers. And then there's full traceability of where we got the response from. And then the third piece is sort of this this larger contextual knowledge. So that I think that was discussed in the in the previous question, the idea that we can look at a single entity and encapsulate and understand everything about that entity and the relationships are key component to that contextual understanding and knowledge graphs are great at capturing that dense information and the relationships that are at play within those key concepts.

Teresa Tung [00:13:26] Yes. So really good techniques, right? I think some of it we're saying is that you must have you're using the LLM almost just as an interface, right? The actual knowledge is still coming from the data and this knowledge representation from your company. And that is something you can govern. And then in the knowledge graph, if you didn't have a knowledge graph and then I was using a rag, I might have too much similarity, right?

Especially as I scale beyond the single use case, I'll have too much. Things that are possibly related that can lead to hallucinations just like a LLM could by itself. Right. It's finding these probabilistic matches, as Tony mentioned, whether it's within LLM or using just a vector database. So having the knowledge graph is actually able to contextualize and really refine. This is the set of data and the outcomes I care about. Right. And again, that's something I can test. And then finally, that ability to take that dense set of information and understand what's related and what's important to my use case. That's that final piece, right? So, the knowledge graph is that Rosetta Stone, right? That allows you to bring your data and knowledge right to bear.

Navin Sharma [00:14:39] That's right.

Teresa Tung [00:14:41] So, Tony, you know, you've worked a long time in life sciences, and life sciences has been one of the earliest adopters of knowledge graphs. Can you give an example of knowledge graphs at scale and the outcomes it's achieved?

Tony Romito [00:14:55] Yeah, sure, Teresa. I think it's useful maybe just to, to start with the notion that we're using the term knowledge graph and there's certainly a maturity spectrum in terms of how sophisticated the design and implementation of a knowledge graph may be. Not just in terms of the scope of business domains in an enterprise, but the capability within the knowledge graph itself. And so, we have this notion of basic property graphs or OPG where we can represent the relationship between entities and the arcs and in between those entities.

And essentially, when we think about the use, we mentioned Google, but something like a Netflix where or other social networks, we're just trying to understand the relationship between entities and create maybe statistical algorithms that help you recommend, you know, the next movie or show that you want to watch may be sufficient. But I think when you get into these more sophisticated and we mentioned regulated use cases, but even ones where you're trying to answer a question for a customer, getting the wrong answer matters even if it's not, you know, designing a new molecule. And so that's where the use of something like the RDF graph resource description, framework type graph becomes a lot more important. And I'll bring that to light through an example. And we're working with one of the large international biopharma companies. They've got dozens of research labs, hundreds of sources from different instruments, sensors, data simulations, Elan systems and all of which have their own data. Definitions need to measure categories of data from chemistry to protein, omics, etc. And what they were trying to understand was how can we compare and connect that data in a way that helps us inform better design of experiments, better definition of experimentation methods, safety profiling, testing processes. We're talking about over 7000 different types of compounds and chemicals, etc. And so, in this case, being able to implement a knowledge graph with standards-based ontologies that had phenotypes from bacteria and fungi to viruses were super important so that we could cross-pollinate the connection between those enterprise but labs and sensors and instruments. So that, as we mentioned earlier, we're talking about applying in an agentic approach. We can now layer on top of that knowledge graph the notion of a lab testing agent or a testing protocol design agent that is leveraging those data as Navin mentioned within the enterprise but also connected to these standard ontologies to really make sense of that and leverage it for both advanced search as well as more sophisticated agentic applications.

Teresa Tung [00:17:52] Yeah. So great examples of being able to take that data that you have and using the knowledge graph to multiply the uses of it. I think that's one of the best descriptions of the differences between label property graphs, like things that you might use for a social network or recommending a movie versus a resource description framework or an RDF that takes a bit more work. But it's for that reasoning. You couldn't read all the literature; you couldn't see all the data changes. So, it's really magnifying that knowledge. So, Navin, one way that knowledge graphs can be applied is also how we manage our data systems. Can you share how we've used knowledge graphs to improve data management?

Navin Sharma [00:18:44] Yeah, absolutely. So, yeah. Believe it or not, Accenture is a client of ours as much as you guys are a partner. We love you just as much as a customer. In fact, this project initially started off as an enterprise search use case where you are obviously it's not lifesaving as we were talking about life sciences. But the emotion here was all around productivity of your teams, sales teams, your client account leads, your consultants, your SMEs who are engaging in various projects. And this was specific to what you call your intelligent engineering and manufacturing platform. The idea was based on challenges that were presented by your clients. You wanted to quickly enable those users on the front line of your engagement opportunities to identify what the right set of solutions might be based on the use cases that were being offered and being able to sort of up level that information and knowledge across previous experiences or previous projects, data and KPIs from those previous projects.

Pulling that together and, you know, ensuring that those were part of the knowledge graph that was being built so that those could be shared based on the user input, which again, was a specific use case. So, the challenge really was how could we improve that productivity for those teams in ways that a given a user challenge, our client challenge that you could quickly formulate a response that brought together all of that information together from people, expertise and knowledge to actual specific use cases and accelerators that have specific KPIs, and metrics attached to them as well.

Teresa Tung [00:20:47] I think most companies will have something similar, just finding the data that you need. And it's something that I usually ask. I'll probably ask Tony. Hey Tony, give me the best proposal or give me the best thing that satisfies X, Y, and Z conditions. And every time I ask, it'll be slightly different, right? And then in his mind, he's probably walking through some logic that is figuring out, well, based off of that, I'm kind of pulling from this one because it has the Geo factor, this one because it's using that technology stack, this one because it has the right financial models so that knowledge representation is really that key, that is what's needed to replace what human's expert would have done.

Navin Sharma [00:21:26] Correct. Absolutely.

Teresa Tung [00:21:29] So, you know, we talked a lot about why and what. But how do you get you know, how is it different than other data architectures? Right. So, you know, there's a common confusion between semantic layers, right? We talked about that meaning and MDM master data management. Tony, how do knowledge graphs complement these existing data architectures?

Tony Romito [00:21:54] Yeah, I think there's a couple of things worth noting here, Teresa. Firstly, of course. Platforms like MDM and knowledge graphs, both in some respects organize and connect data. When you think about MDM or even traditional data warehouses, data lakes are really about pulling data and collecting it and storing it so it can be consumed by other applications, by tools, etc. Whereas in, in you references earlier in the discussion, the knowledge graph is much more about mapping complex relationships across entities and data sources. So, it's more about the interconnectivity versus the collection and storage. I think that's the main difference. And so, you know, as a result, the knowledge graph can reference your existing data enterprise of technologies. And we talk a lot, Teresa, with our clients about this notion of the data mesh and becomes super important because the comparison when I talk to other CIOs and CDOs is I don't have to throw away the data state that I have established and invested significantly in over decades, perhaps I can leverage that as sources and connected into building the knowledge graph that then accomplishes the goals that the team was mentioning in terms of leveraging the enterprise knowledge much more effectively and efficiently than if I had to restructure it into an entirely new data architecture.

Teresa Tung [00:23:30] Yeah, I guess it's horses for courses. So, they coexist, right? You would still have MDM, especially when you must have that single canonical view and then you would have knowledge graph that fills in all the gaps where we can't ever get to a single canonical view because we just might disagree, or I have already invested in my system.

You've invested in your system and especially as we're sharing data across silos of the business, but even with partners, right, I'm going to force you to create my use my standard or me to use your standard, that that's kind of a nonstarter, right? So that knowledge graph is really good to deal with at the conceptual level. We as humans can really tell even if you have something that's biweekly versus monthly, we know how to relate that in ways there could be many ways to actually relate those concepts. But you know, just something as simple as that could take a lot of, of just fighting with a master data management approach. But sometimes you do need to make sure that you are canonical so that you can answer a question properly.

Navin Sharma [00:24:32] Yeah, no, absolutely. And then look, at the end of the day, it's, it's still very much you're still collecting a lot of the idea of a master data is it's still somewhat static in nature not as dynamic when we start to think about transactional data. Right. That really starts to inform that context about what someone is buying or what products are being supplied. And those are changing fast. And where we are in the delivery route of these, these parts are products that transactional data is, is the way you start to bring that larger context. And that's where knowledge graphs can connect a dynamic element of the information landscape with the somewhat static information that needs to be mastered to your point, so that you at least have a view, a perspective about an individual or an entity, but then you have the broader context you bring together with the knowledge graph.

Teresa Tung [00:25:24] And even more so with unstructured data, right? So, let's talk about getting started. Right? So, Tony, can you share your experiences about getting started with knowledge graphs and particularly regarding the talent and skills needed? We've been talking about a lot of terms like how do you even how do you begin?

Tony Romito [00:25:44] Yeah, sure. I think I mean, first, obviously we've talked about other. Technology leaders Google, Netflix and Meta having used knowledge graph for years. But clearly, it's not deeply rooted in the competency of most enterprise IT Organizations. So, I mean a couple of areas that I think are primed to start with. First, I would highlight would be in enterprise architecture. This is no longer just about data architecture. You really need new skills, not just implement. Knowledge graph as a tool, but to incorporate it into the fabric of the digital core for the enterprise. Because the knowledge graph really does need to become a building block of how applications are surface. And that's very different from building a data warehouse and then putting the AI on top. It's not really in the flow of how business applications and business workflow is happening. And so, I think that's one area that organizations need to kind of redefine what that digital core looks like with the notion of this knowledge layer. The semantic layer accounted for. I think the other thing I would highlight would be this notion of the knowledge domain owners. And I think this is actually a great catalyst. The technology in and of itself for organizations to enhance their depth of understanding of the functional domain, the know-how and then representing that in the graph framework becomes an output. But having that accountability and bring that to light through kind of an example, we start to see organizations that are implementing knowledge graph in this way. It's fundamentally changing the ways of working between the business and IT. Because there's a different level of accountability associated with the knowledge graph content. There's more ownership from the business standpoint, but also from the digital organization to understand how to use that and build the types of applications and agenda solutions that are meaningful for the business.

Teresa Tung [00:27:49] I love that domain knowledge owner. That's I think something we're going to see a lot more of. Navin, what are some practical tips and resources for those looking to get started with knowledge graphs? I know in some cases you can jump start even where we don't have enough good quality data.

Navin Sharma [00:28:10] Yeah. No, absolutely. The best way in. And you started talking about domain knowledge individuals. And we've seen the best way companies can embark on this journey is focus put you know, bring your focus down to where are you going to drive maximize value for the organization and what are those functional use cases that are important for you to make informed decisions? It somewhat ties to this notion of decision intelligence, where it matters and. For us. We've seen that, believe it or not, in some of these regulated industries we've been talking about. Right. So, whether it's supply chain and building some sort of a digital twin of a particular product or product area where you can track it all the way down to the manufacturing parts, to the component parts of that supply chain to post-market delivery. And that may have an impact, for example, from a business perspective, a stock shipment use case. So, there was an impact on a particular material that was attached to a particular product, which of course is multiple hierarchies involved in what makes up an even a door on an automobile. And then there's those components that have been impacted by that. How do you have full traceability back all the way to the actual manufacturer or the suppliers? What orders are in transit that need to be stopped from a shipment perspective? You can apply that in the life sciences domain, as Tony mentioned as well, right? We get all these issues where someone's reported a bad reaction to a particular drug. How do you trace that back down to the molecules? Right.

So, for us, where we've seen success is where organizations focused that the notion of knowledge drive connecting the digital thread and whether it's in manufacturing processes, supply chain processes, life sciences, processes that drive important critical decision. Same thing in the public sector domain, right? It's we're talking about military movement. We're talking about life and death, those sort of situational awareness types of decision diligence.

Teresa Tung [00:30:29] Okay, Let's go through one last section. It's a rapid Crossfire segment around debunking myths, around knowledge graphs. So, I'm going to give you a chance to give a short answer about how you would debunk some of these myths. So, I want to start. Right. So, the first one is knowledge graphs will become unnecessary as AI evolves. Tony, do you want to start?

Tony Romito [00:30:54] Sure. I think its clear air has become more sophisticated at a pace than we maybe otherwise would have never anticipated. But I think even recently, as of last week, co-founder from Open AI said we're reaching the end of the Pre-training era. We've kind of consumed all the data on the Internet with large language models. And so, the point was it's now going to be more about understanding and getting optimal use out of these models. And I think that's where the agent horizon comes into play. And more important than ever is going to be that institutional knowledge about the content, the semantics, the context. And that's where I think knowledge graphs have a prime role as a complement to what others may consider kind of the core of modern art.

Teresa Tung [00:31:47] Yeah. Great. Let's do another one. So, the myth is knowledge. Graphs are too complicated and too hard to use. Navin, do you want to share any thoughts on this misconception?

Navin Sharma [00:32:00] Absolutely. At the end of the day, what knowledge graphs are all about, like I said, encapsulating information about things, things we understand on our day-to-day basis. So, when you talk about modeling, we know how customers and products relate. We know how products and orders relate. And modeling information in a knowledge graph is by far the fastest, most easiest way companies can jump start a lot of their data initiatives than their traditional relational databases approaches where we just kind of like a waterfall approach where you go from a logical modeling exercise to conceptual, conceptual to logical to physical, and that requires a very deep skill set and data engineering or database management versus someone that's just knowledgeable on the domain, the functional domain knowledge of the person that we talked about earlier. That being said, yes, it's there's a lot more relational skill says a lot more skill sets. And the good news is that technology has evolved. And even in our case, it started all, you know, we've put a lot of effort behind sort of reducing that skillset gap.

Teresa Tung [00:33:11] Yeah, I think it's been easier with this age of Gen AI, right? So, we as data practitioners learn to think relationally, right? We've forced all knowledge to be relational and the world's not. And so now it's actually making it easier for people who haven't been trained that way, right. To model how things really are. So, I think the third myth we can implement Gen AI at scale without a knowledge graph. Tony, that's for you.

Tony Romito [00:33:36] Sure. I mean, I've said to others, Teresa, that, you know, I think everyone was fascinated with Gen AI, about being able to generate content. And so, it's great if you want to write a poem about cats that doesn't require a sophisticated approach. It doesn't require a precise approach. But as we move forward and again, whether it's a regulated industry use case or not, if I'm trying to use Gen AI to respond to a customer request, if I'm trying to use AI to write, underwrite an insurance policy, let alone do something like, you know, design lifesaving medicines, precision matters and accuracy matters. And that requires us to really harness all that industry and domain knowledge in a way that is more about effective AI than it is just efficient AI. And I think that's where knowledge graphs again, really play a strong role in enabling that that future reality.

Teresa Tung [00:34:39] Okay. Well, thank you so much. And hopefully we've convinced our listeners that knowledge graphs are more important now than ever. So, I want to thank Tony and Navin so much for joining me. But thank you all for listening.

Navin Sharma [00:34:53] Thank you, Teresa, and thank you to our listeners as well. Appreciate the opportunity.