

Building an Ontology with LLMs

Dave McComb, June 2025

We implement Enterprise Knowledge Graphs for our clients. One of the key skills in doing so is ontology modeling. One might think that with the onslaught of ChatGPT and the resulting death knell of professional services, we'd be worried. We're not. We are using LLMs in our practice, and we are finding ways to leverage them in what we do but using them to design ontologies is not one of the use cases we're leaning on.

A Financial Reporting Ontology

Last week Charlie Hoffman, who is an accomplished accountant and CPA, showed me the financial reporting ontology he had built with the help of an LLM. As so many of us are these days, he was surprised at the credible job it had done in so little time. It loaded into Protégé, the reasoner ran successfully (there weren't any real restrictions so that isn't too hard to pull off). It created a companion SHACL file. In the prompt he asked it to base it on gist, our upper ontology, and sure enough there was a gist namespace (an old one but still it was a correct one) with the requisite gist: prefix. It built a bunch of reasonable-sounding classes and properties in the gist namespace (technically namespace squatting, but we haven't gotten very far on ethical AI yet).

Now I look at this and think, while it is a clever trick, it would not have helped me build a financial reporting ontology at all (a task I have been working on in my spare time, so I would have welcomed the help if there was any). I would have tossed out every line. There wasn't a single line in the file I would have kept.

One Click Ontology Building

But here's where it gets interesting. A few months ago, at the KM World AI Conference, one of my fellow panelists, Dave Hannibal of Squirro, stated confidently that within a year there would be a one-click ontology builder. As I reflect on it, he was probably right. And I think there is a market for that. I overheard attendees saying, "even if the quality isn't very good, it's a starting point, and we need an ontology to get started."

An old partner and mentor once told me, "Most people are better editors than authors." What he meant was: give someone a blank sheet of paper and they struggle to get started, but give them a first draft and they tear into it.

The Zeitgeist

I think the emerging consensus out there is roughly as follows:

- GraphRAG is vastly superior to prompt engineering or traditional RAG (it's kind of hard for me to call something "traditional" that's only a year old), in terms of reigning in LLM errors and hallucinations.
- In order to do graphRAG you need a Knowledge Graph, preferably a curated Enterprise Knowledge Graph.
- A proper Enterprise Knowledge Graph has an Ontology at its core.
- Ontology modeling skills are in short supply and therefore are a bit of a bottleneck to this whole operation.
- Therefore, getting an LLM to create even a lousy ontology is a good starting point.

This seems to me to be the zeitgeist as it now exists. But I think the reasoning is flawed and it will lead most of its followers down the wrong path.

The flawed implicit assumption

You see, lurking behind the above train of thought is an assumption. That assumption is that we need to build a lot of ontologies. Every project needs an ontology.

There are already tens of thousands of open-source ontologies "out there" and unknowable multiples of that on internal enterprise projects. The zeitgeist seems to suggest that with the explosion of LLM powered projects we are going to need orders of magnitude more ontologies. Hundreds of thousands, maybe millions. And our only hope is automation.

The Coming Ontology Implosion

What we need are orders of magnitude *fewer* ontologies. You really see the superpowers of ontologies when you have the simplest possible expression of complex concepts in an enterprise. Small is beautiful. Simpler is better. Fewer is liberating.

I have nearly 1000 ontologies on our shared drive that I've scavenged over the years (kind of a hobby of mine). Other than gist, I'd say there are barely a handful that I would rate as "good." Most range from distracting to actively getting in the way of getting something done. And this is the training set that LLMs went to ontology school on.

Now I don't think the world has all the ontologies it needs yet. However, when the dust settles, we'll be in a much better place the fewer and simpler the remaining ontologies are. Because what we're trying to do is negotiate the meaning of our information, between ourselves and between our systems. Automating the generation of ontologies is going to slow progress down.

How Many Ontologies Do We Need?

Our work with a number of very large as well as medium-sized firms has convinced me that, at least for the next five years, every enterprise will need an Enterprise Ontology. As in 1. This enterprise ontology that some of our clients call their “core ontology” is extended into their specific sub-domains.

But let's look at some important numbers.

- gist, our starter kit (which is free and freely available on our web site) has about 100 classes and almost that many properties, for a cognitive load of 200 concepts.
- When we build enterprise ontologies, we often move many distinctions into taxonomies. What this does is shift a big part of the complexity of business information out of the structure (in the ontology and the shapes derived from the ontology) and into a much simpler structure that can be maintained by subject matter experts and has very little chance of disrupting anything that is based on the ontology. It is not unusual to have many thousands of distinctions in taxonomies, but this complexity does not leak into the structure or complexity of the model.
- When we work with clients to build their core ontology, we often double or triple the number of concepts that we started with in gist, to 400-600 total concepts. This gets the breadth and depth needed to provide what we call the scaffolding to include all the key concepts in their various lines of businesses and functions.
- Each department often extends this further, but it continues to astound us how little extension is often needed to cover the requisite variety. We have yet to find a firm that really needs more than about 1000 concepts (classes and properties) to express the variety of information they are managing.
- A well-designed Enterprise Ontology (a core and a series of well-managed extensions) will have far fewer concepts to master than even an average-sized enterprise application database schema. Orders of magnitude fewer concepts than a large packaged application, and many, many orders of magnitude fewer than the sum total of all the schemas that have been implemented.

We're already seeing signs of a potential further decrease. Most of the firms in the same industry share about 70-80% of their core concepts. Industry ontologies will emerge. I really mean useful ones; there are many industry ontologies out there, but we haven't found any useful ones yet. As they emerge, and as firms move to specializing their shared industry ontology, they will need even fewer new unique concepts.

What we need are a few thousand well-crafted concepts that information providers and consumers can agree on and leverage. We currently have millions of concepts in the many ontologies that are out there, and billions of concepts in the many database schemas that are out there.

We need a drastic reduction in quantity and a ramp up in quality if we are to have any hope of reigning in the complexity we have created.

LLMs used for ontology building promise a major distraction to that goal. Let's use LLMs instead for things they are good at, like extracting information from text, finding complex patterns in noise, and generating collateral content at wicked rates to improve the marketing department's vanity metrics.