Q&A
24 November 2021
The Universal Digital Twin – accessing the world of chemistry
Professor Markus Kraft (University of Cambridge)

Q1: Please can you show the line with the list of ontologies again? Like the DEXPI, the equipment ontology for pipetting, etc.?

We have just published a paper where we talk about these things. This is my group's website (https://como.ceb.cam.ac.uk). If you click on "Research" and then "Knowledge graphs", you can see all the stuff that we publish. In particular, you can see preprints that are related to this, and you will see only a few things that are related to actual Chemistry. I told you it's a world model, but what is probably of interest is we've just submitted the review article on "Evolution of Laboratory Automation" and that would contain most of these links directly. Another article that you can look at is "From data base to knowledge graph - using data in chemistry" which is appeared in Current Opinion in Chemical Engineering and has a list of ontologies.

Q2: WolframAlpha shows the user a rephrased version of their query, so the user knows what WolframAlpha thinks they asked. Does your natural language query system provide something like that?

No, we cannot do that, not yet.

Q3: What does SPARQL do that SQL cannot do?

The way we use SPARQL so far you could also use SQL, but that's not the point. I'd say one of the key differences is this open-world assumption of the knowledge graph that allows you to constantly change the structure of the underlying database. If you have something like SQL and a fixed database, and you constantly change items names, then you really end up with a lot of trouble. And of course, you can then also use this to query in a logical sense. So, this is a bit more powerful. But as I said in our case we've been using this mainly for the open-world assumption and our ability to connect things without having to restructure the database every time.

You can traverse down the entire graph model, with SPARQL whereas with SQL you're bound by how the tables are linked together.

Q4: Is there any quantum chemistry software that checks in the OntoCompChem knowledge base to see if a calculation has already been done before, possibly repeating the calculation? And if not, do you think this would be a good idea?

Of course, one reason why we have OntoCompChem is to avoid repeating calculations as they are very expensive. The idea is that in the future you can ask Marie a question, and age restrict it and then it goes and does the quantum calculation for you, using Gaussian at the specified level of theory, or just looks it up. At the moment, with Marie, we're just experimenting with it: (https://kg.cmclinnovations.com/explore/marie). So, you can have these queries for example "the chemical formula of alkanol with heat capacity less than 15", and then it would find it. However, speed is a problem, and our ontologies are quite complex. Another example is, "what's the symmetry number of C8H14"? It can do something like that, but what we still need to do is to make it faster, make it more robust. We are in the process of doing this. Ideally, this thing is then linked to PubChem so that any result you'd find in PubChem, can you also get through Marie and more because you can link it to additional information, for example quantum chemistry results that are already in Marie.

Q5: What you tell us about using chemical ontologies for chemistry is very impressive, but I want to ask about an introductory point. If I understood you correctly, your general aspiration is to make a single digital twin of the entire universe?

I would almost say the universe is not enough. When I say "The World Avatar", then I mean basically anything that is conceptualisable. So just to give you a flavour, you can have a look at the UK Gas Grid: (https://kg.cmclinnovations.com/explore/digital-twin/gas-grid). Here you see the UK Gas Grid, you can look at Bacton Great Yarmouth Gas Terminal, and you can even check the real-time data in that gas line. And of course, the interesting thing is that the composition of the gas in the gas line is uses OntoSpecies representation. You can click here and then you get the real-time outcome. What I'm trying to say is that this is the beauty about knowledge graphs: open-world assumption. Of course, our ambition is to model the whole world, but it doesn't have to be complete, it's just growing and getting better. It incorporates live data and if you go to the website I mentioned earlier, we see different ideas and different areas where we've tried this out.

There are some follow-up points in the Q&A

The point is, I am not trying to build a strong AI or an Artificial General Intelligence, that's not the point here. The point is to enable interoperability, and for that, it is necessary to not concentrate on a specific use case. So, you have to be completely open to achieve interoperability. Then you implement within this very general framework specific use cases where you basically create domain ontologies, and then link them to on one another. They are then able to cross these domains – this is what it's all about. I'm not waiting for the superintelligence to arrive, this is not the point here. The point is that with this framework we are able to step over domain borders. I showed this in the very last video where we basically crossed several domain borders using real-time information. And the beauty about the knowledge graph is that it can be completely distributed; so you're not stuck on one computer, you can have it all over the place. In fact, what we do at the moment is work on a flood risk project where we look at: what infrastructure and what land use is affected if we have different types of flooding in the Kings Lynn area (https://kg.cmclinnovations.com/explore/digital-twin/flood-risk). And this is where we have the gas, electricity, flood information, even agriculture land use, all in one system. This is the reason why we use this world model, we are not planning to create Skynet.