Ontologies in Artificial Intelligence

One requirement of Artificial Intelligence (AI) is the ability to understand, reason, and learn with new data sets in a manner that produces expected and reproducible results. That is, AI requires machines to correctly apply existing knowledge to new and potentially unexpected inputs. The adaptation to new domains requires use of various machine learning techniques, such as supervised learning, where examples of known solutions are used to guide the learning process. In support of these learning processes, Deep Learning algorithms allow computers to develop effective neural networks that can collaborate to recognize complex patterns in large data sets.

Ontologies are ideally suited as repositories of machine processable representations of existing knowledge. They can be used to provide the context to direct algorithmic behavior towards generalizing its knowledge when presented with future data. The use of ontologies as tools to help machines learn offers the promise of eventually requiring less input by people and better understanding of AI results.

While it is true that ontologies are designed for knowledge representation, it is not the case that simply building ontologies to describe particular domains will make them suited for advancing the state of the art in AI unless those representations are integrated into a consistent, semantically interoperable whole.

Developing Ontologies for Artificial Intelligence

It is a common practice among data modelers to create a representation with a scope limited to a particular domain or some portion of the domain used by an application. While sensible, clearly every representation should not model the entire universe, such efforts are short-sighted if they do not enable performative linking of data from the domain they describe to data from other domains. To provide those linkages, the Linked Open Data approach is to build ontologies in isolation and then create mappings between ontologies as needed. Figure 1 depicts the Linked Open Data Cloud showing 1,639 data sets linked by 16,147 mappings.

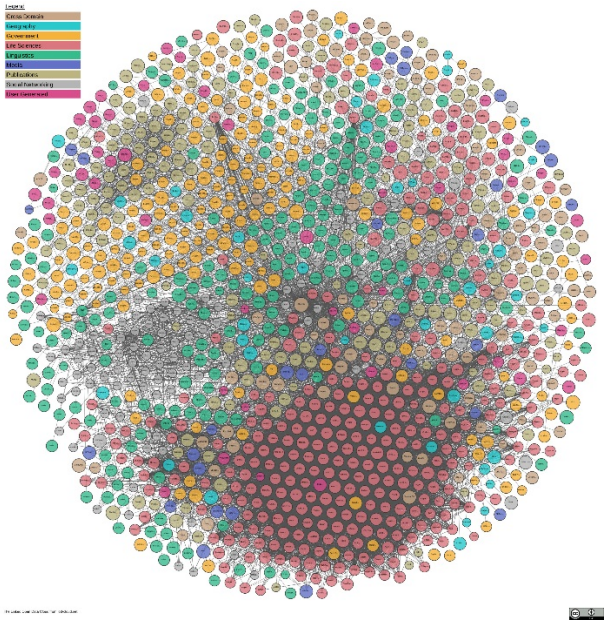

Figure 1 The Linked Open Data Cloud from lod-cloud.net

We maintain that such an approach is not scalable to the challenges presented by integrating large numbers of domains. *N* ontologies requires N^2-*N* mappings and the maintenance needed to keep such a large number of mappings current would be costly. Even with up-to-date mappings, the number of different ontologies that contain the same content creates obstacles to the precision and recall of queries. The complexity and number of mappings leads to contradictions in the resulting semantics.[1]

In contrast to the Linked Data approach, CUBRC's Common Core Ontologies (CCO) provide the rigor, domain coverage, flexibility, expandability, scalability, modularity, and interoperability that scalable AI requires. CCO has been carefully developed according to a methodology based on the principles from the OBO Foundry[2], which promotes an approach based on best-practices in ontology development. The purpose of the CCO is to provide a robust, well-vetted upper and mid-level framework that facilitates the expedited development of domain-level extension ontologies that can keep pace with the new domains encountered in AI applications. By extending from CCO, the rigor and interoperability of the domain representations is ensured, thereby facilitating the repeated production of high-quality ontologies.

The CCO have been in development by CUBRC since 2008, which has received over $7M of funding to mature the core ontologies and further develop extension ontologies for a wide array of domains, examples of which include: space situational awareness, undersea warfare, aircraft maintenance, joint doctrine, cyber warfare, sensor assignment to missions, the product lifecycle, and homemade explosive devices. External users of CCO include the National Geospatial Agency, Johns Hopkins University Applied Physics Laboratory, the Institute for Defense Analyses, and the University of Toulouse. The CCO were open-sourced in 2017 and are available on GitHub[3].

The CCO are submitted as a candidate standard for mid-level ontologies in field of Artificial Intelligence. The breadth and depth of CCO make it applicable as a cross-cutting base terminology for any domain. As a standard, the CCO would serve as a central repository for this base terminology, but would evolve as needed to serve the needs of the AI community both in terms of content and serializations. In addition to this brief overview, the response to the RFI includes three other documents: Best Practices of Ontology Development, An Overview of the Common Core Ontologies 1.1, and Modeling Information with the Common Core Ontologies.

[1] Rudnicki R, Cox AP, Donohue B, Jensen M, "Towards a methodology for lossless data exchange between NoSQL data structures," Proc. SPIE 10635, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX, 106350R (4 May 2018).

[2] http://www.obofoundry.org/principles/fp-000-summary.html

[3] https://github.com/CommonCoreOntology/CommonCoreOntologies