Unit 6 exercise – Ontology development methods

Ontology development is a critical discipline within knowledge engineering, directly shaping the representation, integration, and retrieval of knowledge in artificial intelligence systems. Recent literature highlights three predominant strategies for ontology construction: manual, automatic, and semi-automatic, each suited to distinct scenarios and accompanied by their own advantages and limitations.

Manual ontology engineering, as discussed by Yun et al. (2021), is exemplified by methods such as the Seven-step process and METHONTOLOGY. These approaches demand significant input from domain experts at every stage. For instance, Yun et al. refer to METHONTOLOGY's application in chemical safety, where ontological accuracy is essential due to regulatory and practical considerations. The primary strength of these methods lies in the creation of rich, precise, and context-sensitive concepts; however, such methods are inherently time- and resource-intensive, hindering their use in rapidly evolving or expansive data environments.

By contrast, automated approaches scale effectively to large or dynamic datasets, using advances in natural language processing and machine learning. Ciroku et al. (2024) introduce the RevOnt system, which extracts competency questions from knowledge graphs (e.g., Wikidata) by abstracting information through language models. A notable example from RevOnt reinterprets "Michael Jackson is a member of the Michael Jackson discography" as a generic statement like "Human is a member of the discography" and generates domain-independent competency queries. This paradigm allows for rapid bottom-up ontology specification and is particularly suitable where existing structured or semi-structured data can provide a foundation. Nevertheless, automatically generated ontologies risk issues such as semantic drift or lack of alignment with expert intuition, potentially compromising relevance in domains requiring subtle expert knowledge.

Semi-automatic strategies emerge as a middle ground. These approaches typically rely on an initial structure defined by experts, which is then extended with concepts and relationships identified through automated analysis—as seen in the workflow described by Yun et al. (2021) and by the customizable, expert-curated filtering in RevOnt (Ciroku et al., 2024). This hybrid scheme offers improved throughput over pure manual methods and greater accuracy than reliance on automation alone, making it especially suitable for interdisciplinary fields or scenarios where some expert oversight remains vital.

An illustrative large-scale use case is provided by Blagec et al. (2022) with the Intelligence Task Ontology and Knowledge Graph (ITO). Unlike more traditional hand-crafted or purely automated models, ITO adopts a process-centric framework, integrating both curated taxonomies from the Papers With Code initiative and automated imports of AI benchmarks, models, and metrics. The project demonstrates extensive manual curation to ensure clarity and accuracy, supported by reusable standards from the broader semantic web community. However, the underlying architecture is explicitly designed for extensibility, allowing future enrichment through further automation and community-driven collaboration. As Blagec et al. emphasize, such an approach is essential for keeping pace with the scale and dynamism inherent in the evolving AI research landscape. The ITO's blend of expert-guided structure with automated data ingestion and continual collaborative curation typifies the semi-automatic model, supporting both depth and breadth required for high-quality, up-to-date coverage of a complex domain.

The choice between these strategies thus depends on the specific requirements of the domain and the characteristics of the underlying data. In highly regulated or safety-critical sectors such as healthcare or chemical safety, manual approaches are preferable for their precision and trustworthiness. For domains with vast, quickly evolving datasets—such as benchmarking in AI, described in Blagec et al.—automated or semi-automatic frameworks are indispensable, balancing the need for timely updates with the assurance of quality through structured expert involvement.

In conclusion, contemporary ontology development increasingly benefits from approaches that balance the efficiency of automation with the rigor and insight of expert curation. The selected strategy should always reflect the domain's complexity, the available resources, and the anticipated rate of knowledge change. The practical scenarios provided by Yun et al. (2021), Ciroku et al. (2024), and Blagec et al. (2022) collectively demonstrate that the effectiveness of a chosen ontology engineering method is best evaluated in light of concrete domain requirements and the evolving nature of real-world knowledge systems,

References:

Blagec, K. *et al.* (2022) 'A curated, ontology-based, large-scale knowledge graph of artificial intelligence tasks and benchmarks', *Scientific Data*, 9(1), p. 322. Available at: https://doi.org/10.1038/s41597-022-01435-x.

Ciroku, F. *et al.* (2024) 'RevOnt: Reverse engineering of competency questions from knowledge graphs via language models', *Journal of Web Semantics*, 82, p. 100822. Available at: https://doi.org/10.1016/j.websem.2024.100822.

Yun, W. *et al.* (2021) 'Knowledge modeling: A survey of processes and techniques', *International Journal of Intelligent Systems*, 36(4), pp. 1686–1720. Available at: https://doi.org/10.1002/int.22357.