A Deep Dive Into Benchmarking Ontology Reasoners: Techniques, Tools, and Insights

Gunjan Singh[1,*], Raghava Mutharaju[1]

[1]*Knowledgeable Computing and Reasoning Lab, IIIT-Delhi, New Delhi, India*

Abstract

Ontology-based reasoners are crucial for knowledge representation and reasoning across various domains, including healthcare and finance, as they facilitate informed decision-making. Despite their importance, evaluating and comparing reasoners remains challenging due to differences in ontology expressivity, dataset characteristics, and the nature of reasoning tasks. This tutorial will guide participants through key performance metrics, experimental design strategies, and data considerations required for effective benchmarking, equipping them with the knowledge to evaluate and enhance the capabilities of the reasoners.

Keywords

Ontology, OWL 2, Reasoner, Benchmarking

1. Goals and Objectives

1.1. Overall Goal

To equip researchers, practitioners, and students with the knowledge and skills required to perform rigorous and meaningful benchmarking of ontology-based reasoners. This will advance the development of high-performance reasoning systems and enhance their application in diverse real-world scenarios.

1.2. Concrete Objectives

- **Introduction to Ontology-Based Reasoning**: Provide an overview of ontology-based reasoning, including fundamental concepts, types of reasoners (static and streaming), their algorithms, use cases, and performance characteristics.
- **Review Benchmarking Methodologies**: Examine current methodologies for benchmarking ontology reasoners, including experimental setups, performance metrics, and evaluation criteria. Identify the strengths and limitations of existing approaches.
- **Hands-On Experience**: Offer practical sessions where participants can use a provided GitHub repository containing a Docker container pre-configured OWL 2 reasoners. Participants will run different benchmarks on these reasoners and visualize the datasets using tools such as WebVOWL to better understand the differences in structures of datasets, thereby gaining insights into performance variations.
- **Interactive Discussions**: Create an environment conducive to active discussion and knowledge sharing. Encourage participants to ask questions, share experiences, and engage in collaborative problem-solving.

2. Audience

- **Intended Audience:** Researchers, practitioners, and students in the fields of artificial intelligence, machine learning, and knowledge representation who are interested in ontology-based reasoning

ER2025: Companion Proceedings of the 44th International Conference on Conceptual Modeling: Industrial Track, ER Forum, 8th SCME, Doctoral Consortium, Tutorials, Project Exhibitions, Posters and Demos, October 20-23, 2025, Poitiers, France

*Corresponding author.

gunjans@iiitd.ac.in (G. Singh); raghava.mutharaju@iiitd.ac.in (R. Mutharaju)

0000-0003-3171-9088 (G. Singh); 0000-0003-2421-3935 (R. Mutharaju)

and benchmarking techniques.

- **Level:** Basic to Advanced. The tutorial is designed to be accessible to participants with a foundational understanding of Knowledge Graphs (KG) or Artificial Intelligence (AI) concepts while also providing advanced insights and practical knowledge for experienced researchers and practitioners.
- **Prerequisites:** A basic understanding of knowledge representation and engineering concepts is recommended. Familiarity with ontology tools and reasoning algorithms will be beneficial but not required.

3. Topic Relevance and Novelty

- **Relevance to the Scope of ER:** Ontologies are one of the central mechanisms for conceptual modeling. Reasoning over ontologies to derive inferences is an important task often performed on ontologies. Benchmarking of ontology reasoners provides insights into the performance of the reasoning system and the structure of the ontology. So, the proposed tutorial is very much within the scope of the ER conference.
- **Relevance to Practice:** Participants will acquire practical skills in benchmarking techniques, enabling them to evaluate and select appropriate reasoners for their domain specific applications. This knowledge is essential for developing efficient AI systems that depend on accurate and reliable reasoning mechanisms over ontologies.
- **Novel Aspects:** To the best of our knowledge, there has not been a tutorial on this theme in the recent past at ER and related venues such as ISWC, ESWC, KR, AAAI, and IJCAI. So, we hope that the ER participants will find this theme novel and interesting.

4. Projected Benefits

Targeted Knowledge Outcomes:

- **Benchmarking Requirements**: Understand the essential requirements for benchmarking ontology-based reasoners, including key performance metrics, evaluation criteria, and the unique challenges associated with different systems.
- **Existing Methodologies**: Gain insights into current benchmarking methodologies, recognize their strengths and limitations, and understand how to apply these methodologies effectively.
- **Practical Application**: Develop the ability to apply benchmarking methodologies to evaluate reasoners in diverse scenarios using practical tools and frameworks.
- **Tool Selection and Configuration**: Enhance skills in selecting and configuring appropriate benchmarking tools and frameworks to improve the accuracy and relevance of evaluation results.
- **Addressing Gaps**: Identify gaps in existing benchmarking practices and explore potential solutions to address these shortcomings, contributing to the advancement of benchmarking techniques.

5. Detailed Outline and Timetable

1. **Introduction and Motivation (15 minutes):**
    - Outline the goals and objectives of the tutorial.
    - Provide an overview of ontology-based reasoning and its significance.
    - Discuss the importance of benchmarking in advancing reasoning systems.
    - Highlight existing RDF benchmarks such as LUBM, BSBM, SP^{2}Bench, and WatDiv, and identify gaps and the need for OWL 2 benchmarks.
2. **Ontology Reasoners (15 minutes):**

- Explore different types of ontology reasoners and their characteristics, including commonly used ones such as Konclude (a tableau-based reasoner) and ELK (a rule-based reasoner), discussing their strengths, limitations, supported profiles (e.g., ELK focuses on $\mathcal{EL}^{++}$) and performance evaluation requirements for static reasoners.

3. **Existing Benchmarking Methodologies (15 minutes):**
   - Conduct a detailed examination of the existing OWL2Bench [1] benchmark and another OntoGen benchmark, discussing its suitability for evaluating OWL 2 reasoners.
4. **Hands-on Session (30 minutes)** : Participants will work on the following.
   - Use a provided GitHub repository having a Docker container pre-configured with static reasoners and run several ontologies generated using different benchmarks on various reasoners to observe and compare performance metrics. Participants will explore how different reasoners handle the same ontology and how performance varies with different benchmarks.
   - Custom build their ontologies using our tool, OntoGen [2]. Participants will learn how to create ontologies that fit their specific needs and see how these ontologies interact with different reasoners.
   - Visualize several ontologies using visualization tools such as WebVOWL to gain insights into performance variations.
   - Analyze and interpret the results, leading to a discussion of common pitfalls and best practices for accurate and meaningful evaluations.
5. **Open Challenges and Conclusion (15 minutes):**
   - Briefly introduce neurosymbolic ontology reasoning.
   - Discuss future directions and challenges in benchmarking static, streaming, and neurosymbolic reasoners.
   - Share additional resources and references for further exploration.
   - Get feedback from the participants on the tutorial.

6. Tutorial Method

6.1. Teaching Methods

- **Lectures and Presentations**: Use slide presentations to explain key concepts, methodologies, and benchmarking techniques.
- **Hands-on Sessions**: Provide a GitHub repository with a Docker container that includes pre-configured static and streaming reasoners, along with sample ontologies. Participants can easily use their ontologies and run the reasoners without having a complex installation setup.
- **Interactive Discussions**: Foster engagement through discussions and Q&A sessions, addressing participant questions and encouraging knowledge exchange.

6.2. Technology Requirements

- **Standard Equipment**: PC projector for presentations and visual aids.
- **Additional Requirements**: Internet access for live demonstrations and interactive tools.

Declaration on Generative AI

In preparing this tutorial proposal, generative AI tools—specifically ChatGPT—were used solely for grammar checking, spelling correction, and improving the readability of certain sentences. All AI-suggested edits were carefully reviewed and refined by the authors. The conceptualization, structure, and content of the proposal were developed entirely by the authors, and the use of ChatGPT was limited to enhancing the clarity and presentation of the text.

References

[1] G. Singh, S. Bhatia, R. Mutharaju, OWL2Bench: A Benchmark for OWL 2 Reasoners, in: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II, volume 12507 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 81–96.

[2] G. Singh, A. Kumar, K. Bhagat, S. Bhatia, R. Mutharaju, OWL2Bench: Towards a Customizable Benchmark for OWL 2 Reasoners, in: Proceedings of the ISWC 2020 Demos and Industry Tracks: 19th International Semantic Web Conference (ISWC 2020), Globally online, November 1-6, 2020 (UTC), volume 2721 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 344–349.