Department of Data Science

Advanced Database Management System Group Assignment

Title Query Rewriting and Optimization for Ontological Databases

Group 9

Name	ID
1.Amanial Gebru	DS0050/15
2.Asmare Yenealem	DS0054/15
3.Wengel Zerihun	DS0082/15
4.Esubalew Dejene	DS0089/15
5.Solomon Guesh	DS0075/15
6.Amanuel Mindaye	DS0089/16

Submission Date: Sunday April 28, 2024
Submitted to: Dr. Tilahun Melak

Abstract

- Ontological queries are evaluated against a knowledge base consisting of an extensional database and an ontology (i.e., a set of logical assertions and constraints that derive new intentional knowledge from the extensional database), rather than directly on the extensional database.

Two key aspects of this problem are query rewriting and query optimization.

- In this article, we discuss two important

aspects of this problem: query rewriting and query optimization. Query rewriting consists of the compilation of an ontological query into an equivalent first-order query against the underlying extensional database.
One of the strengths of this article lies in its thorough exploration of query rewriting, which involves translating ontological queries into equivalent first-order queries against underlying extensional databases.

We present a novel query rewriting algorithm for rather general types of ontological constraints that is well suited for practical implementations. In particular, we show how a conjunctive query against a knowledge base, expressed using linear and sticky existential rules, that is, members of the recently introduced datalog family of ontology languages, can be compiled into a union of conjunctive queries (UCQ) against the underlying database. Ontological query optimization, in this context, attempts to improve this rewriting process soas to produce possibly small and cost-effective UCQ rewritings for an input query.

Introduction

Ontological Database Management Systems

- The use of ontological reasoning in companies, governmental organizations, and other enterprises has become widespread in recent years.
- An ontology is an explicit specification of a conceptualization of an area of interest and consists of a formal representation of knowledge as a set of concepts within a domain, as well as the relationships between instances of these concepts.

Moreover, ontologies have been adopted as high-level conceptual descriptions of the data contained in data repositories that are sometimes distributed and heterogeneous in the data models.

This amalgamation of different technologies stems from the need for semantically enhancing existing databases with ontological constraints.

Indeed, database technology providers have recognized this need and have recently started to build ontological reasoning modules on top of their existing software with the aim of delivering effective database management solutions to their customers. For example, Oracle, Inc.offers a system, called Oracle Database 11g, enhanced by modules performing ontological reasoning tasks1. Also, Ontotext offers a family of semantic repositories, called OWLIM2, and Semafora Systems develops an inference machine, called Ontobroker3, for processing ontologies that support all of the World Wide Web Consortium (W3C) recommendations.

First-Order Rewritability

Polynomial-time tractability is often considered not good enough for efficient query processing. Ideally, one would like to achieve the same complexity as for processing first-order queries, or, equivalently, (nonrecursive) SQL queries.

An ontology language

L guarantees the *first-order rewritability* of conjunctive query answering if, for every conjunctive query *q* and ontology _ expressed in *L*, a positive first-order query *q*_ called *perfect rewriting*6 can be constructed such that, given a database *D*, *q*_ evaluated over *D* yields exactly the same result as *q* evaluated against the ontological database *D* ∪ _ [Calvanese et al. 2007]. Since answering first-order queries is in AC0 in data complexity [Vardi 1995], it immediately follows that query answering under ontology Languages that guarantee the first-order rewritability of the problem is also in AC0 in data complexity.

First-order rewritability is a most desirable property since it ensures that the query answering process can be largely decoupled from data access. In fact, as depicted in Figure 1, to answer a query *q* over an ontological database *D*∪ _, a separate software can compile *q* into *q*_, then translate *q*_ into a standard SQL query *q*_, and finally submit it to the underlying relational database management system holding *D*, where it is evaluated and optimized in the usual way.

This amalgamation of different technologies stems from the need for semantically enhancing existing databases with ontological constraints. Indeed, database technology providers have recognized this need and have recently started to build ontological reasoning modules on top of their existing software with the aim of delivering effective database management solutions to their customers. For example, Oracle, Inc. offers a system, called Oracle Database 11g, enhanced by modules performing ontological reasoning tasks1. Also, Ontotext offers a family of semantic repositories, called OWLIM2, and Semafora Systems develops an inference machine, called Ontobroker3, for processing ontologies that support all of the World Wide Web Consortium (W3C) recommendations. Enhancing databases with ontologies is also at the heart of several research-based systems such as QuOnto [Acciarri et al. 2005] and Quest [Rodriguez-Muro and Calvanese 2012].

Ontologies are modeled using formal languages called ontology languages.
The introduction provides a comprehensive overview of ontological database management systems (ODBMS) and the role of ontology languages in modeling data repositories.

- Strengths:
- Clarity of Definitions: The introduction offers clear definitions of key concepts such as ontologies, ontology languages, and ODBMS, making it accessible to readers unfamiliar with the subject matter.
- Relevance and Context: The discussion of the evolution of database management systems towards incorporating ontological reasoning provides valuable context for understanding the motivations behind ODBMS development.
- Comprehensive Citations: The inclusion of citations to relevant research projects, grants, and existing systems adds credibility and situates the current work within the broader academic landscape.

Aims and Objectives

- Objective of ontological databases is to represent knowledge in a structured and semantically rich manner, allowing for more sophisticated querying, inference, and reasoning about the data.
- Here are some specific objectives of ontological databases:
  Knowledge Integration: Ontological databases facilitate the integration of heterogeneous data sources by providing a common semantic model.
  They enable the integration of data from different sources and formats, allowing users to query and analyze information across multiple domains.
  Inference and Reasoning: Ontological databases support inference and reasoning capabilities, allowing users to derive new knowledge from existing data through logical deductions. By leveraging the formal semantics encoded in the ontology, they can infer implicit relationships, classify entities, and perform complex reasoning tasks.
  Flexibility and Extensibility: Ontological databases are designed to be flexible and extensible, allowing for the representation of evolving knowledge structures. Users can easily modify and extend the ontology to accommodate new concepts, relationships, and constraints as the domain evolves over time.

Advanced Querying: Ontological databases enable more sophisticated querying capabilities compared to traditional databases. Users can express complex queries that involve semantic relationships and constraints, allowing for more precise and context-aware retrieval of information.

Domain-Specific Applications: Ontological databases are particularly well-suited for domain-specific applications where understanding complex relationships and semantics is crucial. They find applications in areas such as scientific research, healthcare, bioinformatics, knowledge management, and semantic web technologies.

Overall, the ontological databases is to provide a powerful framework for representing, integrating, and reasoning about knowledge, enabling more effective management and utilization of data in various domains.

Apart from designing a practical rewriting algorithm for linear and sticky TGDs, we would also like to investigate the possibility of improving the computation of the perfect rewriting on multicore architectures commonly available in modern database servers. constraints natively, as done today for traditional data dependencies such as primary and foreign keys.

A key difference is that ontological constraints are not supposed to be enforced by the DBMS as classical integrity constraints, but rather to be taken into consideration during the evaluation of a query. This article is a significant step towards this direction.

Critical Analysis

- Ontological databases are

it's evident that the article provides valuable insights into addressing the challenges of querying ontological databases.

A critical analysis of ontology databases involves examining their strengths, weaknesses, and implications for various applications.

Here's a breakdown:

Strengths:

Semantic Representation: Ontology databases excel at representing knowledge in a structured and semantically meaningful way.

Support for Semantic Web: Ontology databases play a crucial role in the Semantic Web vision by providing a foundation for organizing and sharing data on the web in a machine-readable format.

Weaknesses

Complexity: Designing and managing ontologies can be complex and challenging, especially for large and dynamic domains.

Scalability: Some ontology databases may face scalability limitations, particularly when dealing with large volumes of data or complex ontologies. Performance issues may arise as the size of the ontology and the dataset increases.

Query Complexity: Querying ontology databases can be more complex compared to traditional databases, especially for users who are not familiar with ontology languages such as OWL or RDF. Constructing queries that leverage the full expressive power of the ontology requires expertise in ontology modeling and querying.

Maintenance Overhead: Ontology databases require ongoing maintenance to keep the ontology up-to-date and consistent with the evolving domain knowledge. This includes updating the ontology structure, resolving conflicts, and ensuring data quality

Statement of the problem

Developing an ontology database system to effectively capture, organize, and represent domain-specific knowledge in a structured and coherent manner. The database should facilitate efficient retrieval, inference, and reasoning processes, enabling users to navigate and query the ontology to derive meaningful insights, make informed decisions, and support various applications within the targeted domain.

Key challenges to address include:

Ontology Design: Designing a comprehensive ontology schema that accurately reflects the concepts, relationships, and entities within the domain, ensuring semantic clarity and interoperability.

Data Integration: Integrating heterogeneous data sources and formats into the ontology database while maintaining consistency and resolving semantic conflicts.

Scalability: Ensuring the scalability of the database to accommodate large volumes of data and evolving domain knowledge without sacrificing performance or usability.

Querying and Inference: Developing efficient querying mechanisms and reasoning engines to support complex queries, infer implicit knowledge, and derive meaningful insights from the ontology.

User Interface: Designing an intuitive and user-friendly interface for accessing, browsing, and interacting with the ontology database, catering to the needs of both domain experts and non-expert users.

Maintenance and Evolution: Establishing mechanisms for ontology maintenance, versioning, and evolution to keep the database up-to-date with changing domain requirements and knowledge advancements.

Interoperability: Ensuring interoperability with existing systems, standards, and ontologies to facilitate data exchange, integration, and collaboration across different platforms and domains.

Addressing these challenges will contribute to the development of a robust ontology database system that serves as a valuable resource for knowledge representation, sharing, and utilization within the targeted domain.

Approaches to Addressing Challenges in Ontological Databases

Ontological databases represent a unique intersection of database management systems and ontological reasoning, posing distinct challenges and opportunities for research and development. Several approaches have emerged to tackle these challenges and unlock the potential of ontological databases.

Here are some key strategies:

Integration with Emerging Technologies:

Explore synergies between ontological databases and emerging technologies such as machine learning, natural language processing, and knowledge graphs.

Investigate how ontological databases can leverage these technologies to enhance data analysis, inference, and decision support capabilities.

User Interfaces and Interaction Paradigms:

Design intuitive user interfaces and interaction paradigms that enable users to interact with ontological databases effectively.

Explore visualization techniques, natural language interfaces, and intelligent query assistants to enhance user experience and usability.

Scalability and Efficiency:

Address scalability challenges associated with large-scale ontological databases by developing distributed and parallel processing techniques.

Investigate strategies for optimizing storage, indexing, and query processing to improve the efficiency of ontological database systems.

Query Rewriting Algorithms:

Design efficient algorithms for rewriting queries expressed in standard languages (e.g., SQL) into ontological query languages, ensuring semantic coherence with ontological constraints.

Investigate techniques for optimizing query rewriting to minimize computational overhead and improve query performance.

Ontology Languages and Formalisms:

Develop and refine ontology languages and formalisms to express complex domain concepts and relationships effectively.

Explore formalisms such as Description Logics (DLs), Web Ontology Language (OWL), and Tuple Generating Dependencies (TGDs) to model ontologies with varying degrees of expressiveness and computational complexity

Conclusion

- In conclusion, the study has delved into the challenge of designing a practical query rewriting algorithm for arbitrary Tuple Generating Dependencies (TGDs). The proposed resolution-based query rewriting algorithm, XRewrite, specifically targets linear and sticky TGDs. Additionally, various optimization techniques have been explored to enhance the efficiency of the rewriting process.

Here are the key takeaways:

Challenges and Solutions: We have identified and addressed various challenges encountered in ontological databases, including query rewriting, reasoning efficiency, semantic integration, and scalability. Through innovative approaches such as advanced query rewriting algorithms, reasoning mechanisms, and semantic integration techniques, researchers are paving the way for more efficient and effective ontological database systems.

Integration of Ontological Reasoning: The marriage of ontological reasoning with database technology has led to the emergence of ontological database management systems (ODBMS), enabling organizations to leverage explicit conceptualizations of their data domains for enhanced understanding and decision-making

Future Directions: Looking ahead, the evolution of ontological databases is poised to continue, driven by advancements in ontology languages, reasoning algorithms, and integration with emerging technologies. Future research directions may include the exploration of distributed ontological databases, integration with machine learning and natural language processing, and the development of intelligent ontological query assistants.

Impact and Implications: Ontological databases have the potential to revolutionize data management practices across various domains, including healthcare, finance, manufacturing, and beyond. By providing a unified framework for representing and reasoning about complex data domains, ontological databases empower organizations to extract actionable insights, facilitate decision-making, and drive innovation.

REFERENCES

Serge Abiteboul, Richard Hull, and Victor Vianu. 1995. *Foundations of Databases*. Addison-Wesley.

Andrea Acciarri, Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Mattia Palmieri, and Riccardo Rosati. 2005. QuOnto: Querying ontologies. In *Proceedings of the 20th National Conference on Artificial Intelligence and the 17th Innovative Applications of Artificial Intelligence Conference*. 1670–1671.

Miklos Ajtai and Yuri Gurevich. 1989. Datalog vs. first-order logic. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*. 142–147.

Hajnal Andreka, Johan van Benthem, and Istvan Nemeti. 1998. Modal languages and bounded fragments of predicate logic. *J. Philos. Logic* 27, 217–274.

Franz Baader. 2003. Least common subsumers and most specific concepts in a description logic with existential
restrictions and terminological cycles. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*. 319–324.

Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider,
Andrea Cali, Georg Gottlob, and Michael Kifer. 2008. Taming the infinite chase: Query answering under
expressive relational constraints. In *Proceedings of the 11th International Conference on Principles of
Knowledge Representation and Reasoning*. 70–80.