ABSTRACT

Regardless of the knowledge structure lack about Resource Description Framework (RDF) data, difficulties, principally, occur in specifying and answering queries. Approximate querying is the solution to find relevant information by getting a set of sub structures (e.g. sub graphs) matching the query. Approaches based on the structure and others based on semantic, marginalized the common meaning between concepts in its computing. In this paper in order to improve the approximation by introducing the meaning similarity between components in the query and RDF components is proposed, getting better need satisfaction. The meaning similarity measure can be calculated using WordNet and used in all steps of the query answering process. In addition, other important properties in the approximation level calculation between query paths and RDF paths are considered; besides indexing and optimizations strategies are performed. Answers are a set of sub graphs ranked in decreasing order on its matching degree. Experiments are conducted within real RDF dataset.

Keywords: Approximate Querying, Graph Matching, Information Retrieval, Regular Path Queries, Resource Description Framework (RDF), Semantic Queries, Semantic Search, Similarity, Wordnet

INTRODUCTION

In recent years, amount data of semantic web grows increasingly; RDF framework (Lassila, & Swick, 1999) is the standard modeling language having widespread usage. RDF is a graph based representation introducing the development of knowledge management systems that offer possibilities to store, manage, explore, extract information, etc. In the context of knowledge structure lack and lack of organization about RDF data, difficulties in specifying queries which getting satisfying answers is noticed. Researchers in the field of query answering discover that existing querying methods are not suitable in such situation. Thus, approximate querying imposed itself as one of the best solutions trying to find relevant information by getting a set of sub structures (e.g. sub graphs) that match (with degree of matching) the query.

Many approaches proposed in approximate querying over RDF graphs based on the structure (e.g. sub graph isomorphism, graph edit distances, etc) or based on semantic (e.g. taking into account types, attributes and the semantic distance of edges and nodes, etc). Those approaches are interesting but they marginalized
the common meaning between concepts in the computing. They are complex, hard to compute and do not consider many concerns in the matching process. RDF representation is a typed graph containing semantic relationship between nodes. Thus, in order to get better satisfaction, the semantic meaning must be considered and combined with structure.

In this paper, an approximation by introducing the similarity meaning between the query components and RDF components getting better need satisfaction is proposed. This is achieved by proposing a meaning similarity measure calculated using WordNet and used in every step of the approach. In addition, we consider other important properties in the approximation level value computation between query paths and RDF paths (like edges orders). An offline RDF indexing and optimizations strategies are performed to efficient execution. The query is represented as a graph pattern and a set of query paths is extracted, trying to find in RDF graphs the set of paths matching the query paths using the proposed approximation. The answers are a set of sub graphs generated from the discovered RDF paths and ranked in decreasing order of its degree of matching. Our experiments are conducted with real RDF dataset.

In the second section related works are discussed and the third section presents the RDF graph normalization and indexing process. The similarity measure based on WordNet is detailed in the fourth section and the fifth section is devoted to the proposed approximate querying approach. The sixth section is dedicated to the approach implementation and experimentation whereas conclusion and future work are introduced in the seventh section.

RELATED WORKS

Many approaches have been proposed to approximate querying over RDF data. The authors in De Virgilio, Maccioni, and Torlone (2012, 2013) compute the approximate solutions (sub graphs) using path alignment trying to find similar RDF paths in the query paths using a similarity measures based on edit distance operations of edges and nodes (deletion, insertion, etc). In (Zhang, Song, He, Shi, & Dong, 2012) a similarity oriented RDF graph matching is proposed and can be used in approximate querying (one of the two graphs may be a query graph), based on computing similarities between triples of the two RDF graphs. In Carroll (2002) an RDF matching work based on graph isomorphism is proposed that can be used in approximate querying i.e. the source graph is a query. An interesting work is presented in Hurtado, Poulovassilis, and Wood (2009) which is an automata based approach considering the edit distance operations and creating an approximate automata. The RDF graph explorations are performed with respect to that automata to finding the node answers.

Another category of works are based on approximate querying using query relaxations (Huang, Liu, Zhou, 2008, 2012; Poulovassilis, & Wood, 2010). The work in (Poulovassilis, & Wood, 2010) is an extension to Hurtado et al. (2009) by combining the approximation and relaxation in the semantic regular path queries, it’s also an automata based approach by creating a relaxed automata by means of RDF inference rules advantages. The answers searching are based on this automata. To getting best relevant answers, the works in Huang et al. (2008; 2012) compute the similarities of relaxed queries with respect to the original query and using them to rank the approximate answers. Similarities are measured based on hierarchical concepts and properties in the ontology. Authors in Han, Finin, and Joshi (2011) proposed an approach allowing writing graph based queries to DBpedia (Bizer et al., 2009) and using similarity measures to translating user’s query into SPARQL query. Other works treating the approximation described in Corby, Dieng-Kuntz, Faron-Zucker, and Gandon (2006), Dolog, Stuckenschmidt, Wache, and Dieiderich (2009), Elbassuoni, Ramanath, and Weikum (2011), Hurtado, Poulovassilis, and Wood (2008), Oren, Guéré, and Schlobach (2008), and Zhu, Zhong, Li, and Yu (2002) are cited and investigating approximation of semantic queries.
Related Content

Web Services Discovery with Rough Sets
[www.igi-global.com/chapter/web-services-discovery-rough-sets/37665?camid=4v1a](www.igi-global.com/chapter/web-services-discovery-rough-sets/37665?camid=4v1a)

An E-learning System Based on Secure Data Storage Services in Cloud Computing
[www.igi-global.com/article/an-e-learning-system-based-on-secure-data-storage-services-in-cloud-computing/89326?camid=4v1a](www.igi-global.com/article/an-e-learning-system-based-on-secure-data-storage-services-in-cloud-computing/89326?camid=4v1a)

A Conceptual Model for Cloud Computing Adoption by SMEs in Australia
[www.igi-global.com/chapter/a-conceptual-model-for-cloud-computing-adoption-by-smes-in-australia/140807?camid=4v1a](www.igi-global.com/chapter/a-conceptual-model-for-cloud-computing-adoption-by-smes-in-australia/140807?camid=4v1a)
Experiences with Software Architecture Analysis of Usability
www.igi-global.com/article/experiences-software-architecture-analysis-usability/2648?camid=4v1a