NPCS: Native Provenance Computation for SPARQL

Zubaria Asma1,2, Daniel Hernández3, Luis Galárraga4, Giorgos Flouris1, Irini Fundulaki1 and Katja Hose5

(1) FORTH-ICS, University of Crete, Heraklion, Crete, Greece, (2) University of Crete, Heraklion, Crete, Greece, (3) University of Stuttgart, Stuttgart, Germany, (4) Inria, Rennes, France, (5) TU Wien, Wien, Austria

Introduction

❖ Knowledge Graphs (KGs)
  Important in both academia and industry
  Ideal for integrating data from various sources
❖ Provenance
  Critical for trust assessment and dynamic data
❖ Our Contribution (NPCS)
  Enriches results with how-provenance annotations
  Supports monotonic and non-monotonic SPARQL queries

NPSCS rewrites SPARQL query \( \mathcal{Q} \) into \( \mathcal{Q}' \), generating how-provenance polynomials in the spm–semiring \( K = (K, \oplus, \otimes, \odot, 0, 1) \) for \( K \)-graph \( G \).

Evaluation of NPCS

❖ Tested on different reification schemes using two engines (Stardog, GraphDB) with datasets of 10M, 100M, 200M from WatDiv and 15 billion triples from Wikidata
❖ Observed trends: NPCS consistently outperforms SPARQLprov in 48 out of 50 studied cases
❖ Applicable to any standard RDF/SPARQL engine with a significant performance margin

Evaluation of NPCS

❖ Novel SPARQL-based method for computing how-provenance annotations
❖ Outperforms existing solutions
❖ Enables efficient computation of provenance for millions of query results on large knowledge graphs
❖ Ideal for ETL processes in multi-source KG construction

Acknowledgment

This work was supported by the ITN KnowGraphs project, under the EU H2020 Marie Skłodowska-Curie grant agreement No 860801

References