Demonstrating Blank Node Matching and RDF/S Comparison Functions

Christina Lantzaki, Yannis Tzitzikas and Dimitris Zeginis

Institute of Computer Science, FORTH-ICS, and
Computer Science Department, University of Crete, GREECE

ISWC2012, Boston, Nov. 2012
Several RDF/S Knowledge Bases rely heavily on blank nodes

- Bnodes are convenient for representing complex attributes or resources whose identity is unknown but their attributes (either literals or associations with other resources) are known.

- We show how to exploit blank node anonymity in order to reduce the delta size when comparing RDF/S Knowledge Bases.

- We approach the problem as an optimization problem:
  - Find the mapping that gives the minimum in size delta

Blank node prevalence *

Opencalais.com  44.9%
hi5.com foaf    87.5%

*[On blank nodes ISWC 2011]
Finding the Optimal Blank Node Mapping

All KBs (general case)

- KBs with no directly connected bnodes
- Time Complexity:
  - NP-Hard: $O(n^3)$
  - $O(n \log n)$

Optimal Mapping

Approximately Optimal mapping

Approximately Optimal mapping

Deviation from optimal

$|\Delta x| - |\Delta_{opt}|$

_mapping of 150,000 bnodes ~11 sec

Web system:
http://www.ics.forth.gr/isl/BNodeDelta

FORTH-ICS, ISWC 2012
Thank you for your attention

Web system available in:
http://www.ics.forth.gr/isl/BNodeDelta

Work done in the context of SCIDIP-ES, APARSEN and i-Marine