BNodeDelta: A tool for efficient blank node matching

Overview

The ability to compute the differences that exist between two Knowledge Bases is an important step to cope with the evolving nature of the Semantic Web. RDF Deltas can be employed to aid humans understand the evolution of knowledge, and to reduce the amount of data that need to be exchanged and managed over the network in order to build Semantic Web synchronization, versioning and replication services. However, a significant percentage of datasets in Linked Open Data (LOD) contain a high number of blank nodes (e.g. the opencalais.com domain has 45% blank nodes). Because of their anonymity, these resources cannot be matched automatically as URIs and literals do. The inability to match them increases the Delta size and does not assist in detecting the changes between subsequent versions of a Knowledge Base (See Fig1).

The problem of finding the blank node mapping that minimizes the Delta size, when comparing two Knowledge Bases, is proved to be NP-Hard (See Fig2). Therefore, *BNodeDelta* takes advantage of the special cases that this problem can be solved optimally and implements two approximation algorithms for the general case.
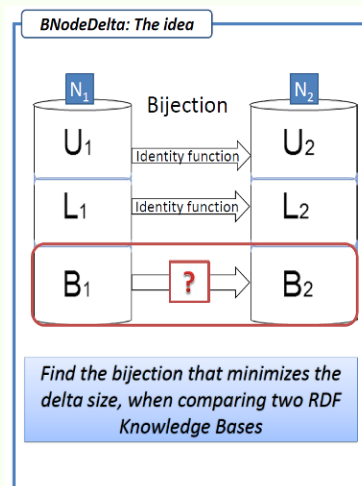

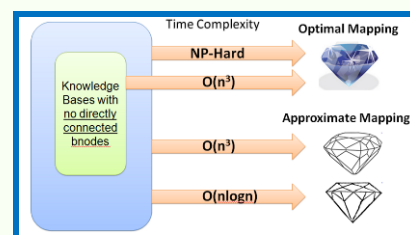*Fig1. Motivation for the blank node matching problem*

Fig2. Tractability map of the blank node matching problem

Target Domains

BNodeDelta can be used by Semantic Web data management frameworks and triplestores to help either in checking equivalence between RDF Knowledge Bases or in versioning and synchronization for reducing their generated Delta.

Description

BNodeDelta provides two novel algorithms for matching blank nodes. The **Hungarian-based algorithm** provides minimum (when blank nodes are not directly connected) or very close to minimum Deltas. For very large datasets, the **Signature-based algorithm** produces reduced Deltas in logarithmic time.

 BNodeDelta
i) takes as input two RDF Knowledge Bases
ii) computes a mapping between their blank nodes
iii) exports their difference (Delta), while taking into account the blank node mapping
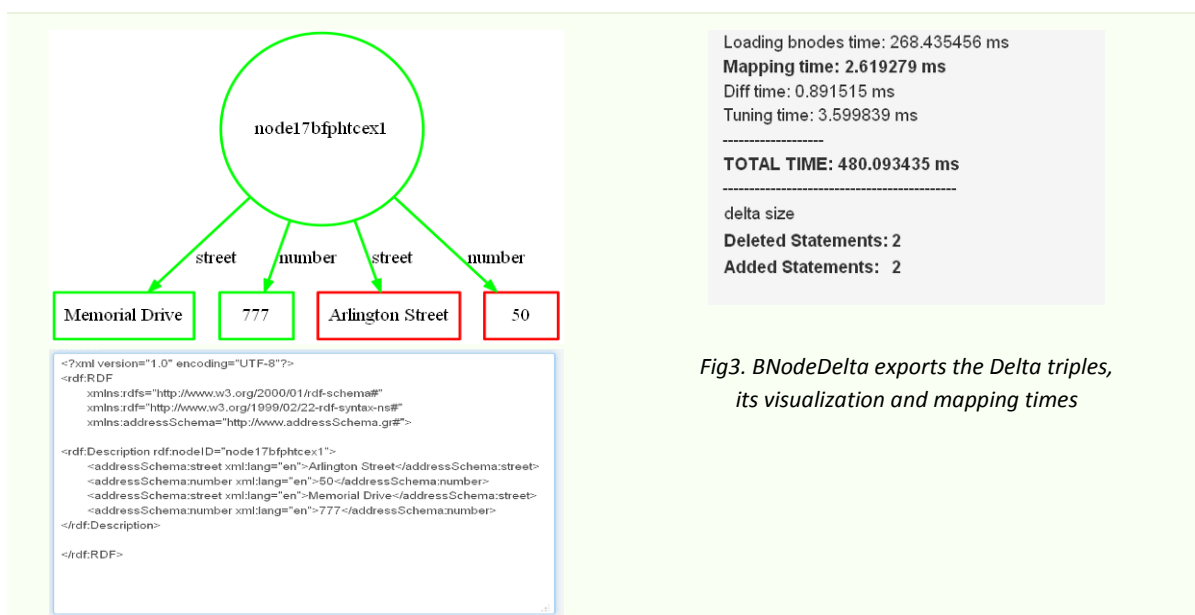iv) provides a quick visualization of the Delta (on the condition it is quite small in size)


```
Loading bnodes time: 268.435456 ms
Mapping time: 2.619279 ms
Diff time: 0.891515 ms
Tuning time: 3.599839 ms
-------------------
TOTAL TIME: 480.093435 ms
-------------------------------------------
delta size
Deleted Statements: 2
Added Statements:  2
```

*Fig3. BNodeDelta exports the Delta triples,
its visualization and mapping times*

Additional Information

BNodeDelta was partially supported by the APARSEN Network of Excellence.

More information is available at the website of BNodeDelta:
http://www.ics.forth.gr/isl/BNodeDelta/

BNodeDelta website

Contact details: **Yannis Tzitzikas**

tzitzik@ics.forth.gr

www.ics.forth.gr/isl