

Benchmarking HDT and GraphRePair

Philip Frerk

May 23, 2019

1 Fragen an Michael

- compressor model: input immer N-triples und turtle/N3 dann schon als compressor ansehen
- baseline compressor: wie sollen die evaluiert werden (DBPedia, Wordnet Teilgraphen)?
- Dekompression evtl auch bei baseline compressors evaluieren?
- HDT star pattern: warum triples kleiner?
- Problem: materialisierung bringt auf Graph Ebene Verbesserung, aber auf Kodierungsebene nicht immer
Grund wahrscheinlich: durch stärkere Kompression mehr Hyperkanten und dadurch wird permutations file größer
Außerdem: Kodierung der Startregel evtl auch manchmal schlechter

2 Experiments

Ontology	Dictionary
Entity-Teilgraph (wenige Vorkommen der relevanten Properties)	blank nodes (Polygon-Daten)
Teilgraph mit vielen Vorkommen	literals (DBPedia abstracts and SWDF)

What the different datasets contain:

	symmetric	inverse	transitive
DBPedia	yes	yes	yes
Wordnet	yes	(hypernym <-> hyponym) only for nouns	yes

All together: persondata graph

3 Dictionary vs Graph

dict / graph compr	yes	no
yes	Normal HDT or GRP+HDT-Dict	HDT with plain triples
no	(HDT or GRP) with HDT's plain dictionary	N-triples

HDT plain triples: possible

HDT Plain dictionary: not possible

4 Notes

warum hypernym und hyponym nicht invers? (nur nomen Teil) Wordnet: (hypernym und hyponym inverse), antonym symmetrisch

open data: URIs aufrufen

DBPEdia: persondata (nur personen rausnehmen: da kommen die relevanten Properties vor)

Eine person nehmen und von dort aus teilgraphen erzeugen

persondata erweitern mit abstracts und birthplace koord.