Research Interests:

Distributed SPARQL

Gregory Todd Williams
Interests

- Query answering over data that is:
  - Heterogeneous
  - Distributed
- Challenges: Unknown schemata, data size; latency; limited resources
Linked Data

- Musicbrainz
- Audio-Scrobbler
- QDOS
- Doap-space
- Flicker exporter
- SIOC profiles
- ECS Southampton
- Sem-Web-Central
- SW Conference Corpus
- Open-Guides
- Jamendo
- BBC Later + TOTP
- BBC John Peel
- Magnatune
- DBpedia
- US Census Data
- riese
- World Factbook
- Geonames
- Gov-Track
- Wiki-company
- Eurostat
- Open Cyc
- Onto-world
- Revyu
- RDF Book Mashup
- DBLP Berlin
- RKB Explorer
- DBLP Hannover
- W3C WordNet
- Project Gutenberg
- flickr wrappr
- lingvoj
- QDOS
Related Work

- ARQ and Sesame extensions for explicit remote execution of sub-queries by Seaborne and Zemánek, resp.
- Join selectivity estimation heuristics by Stocker, et. al.
- Quilitz's DARQ federated SPARQL extension to ARQ.
Past Research

• SPARQL Extensions
  • On-the-fly extension functions
  • Temporal Queries
  • Intermediate Result Passing
Current Research

- Evaluation of join cardinality estimation heuristics
- Distributed Query Answering
  - Pose queries to a set of endpoints
  - Optimize for:
    - Co-located data
    - Early results
Optimization Strategy

• Co-located data

  • Send optimistic sub-queries to individual endpoints

  • Join sub-query results to produce early query results

• Early results

  • Expect sub-queries to be answerable quickly (data provider knows more about data than we do)
Questions?