Motivation

- Currently users need to be aware of the structure of datasets in order to query Linked Data.
- Natural language queries as an intuitive way to query Linked Data.

Objective

- A vocabulary independent natural language query mechanism for Linked Data.

Challenges

- How to query heterogeneous, complex and distributed Linked datasets using natural language?
- How to allow users query data without prior knowledge of its representation?
- How to address the expressivity-usability trade-off of existing approaches?

Proposed Approach

- Investigation of a best-effort query processing approach based on the combination of 3 elements:
  - Entity search.
  - Wikipedia-based semantic relatedness measure.
  - Spreading activation search.
- Investigation of the suitability of a Wikipedia-based semantic relatedness measure (Wikipedia Link Measure) for semantically matching query terms to dataset terms.

Results

Query quality evaluated using a variation of the QALD Workshop DBpedia training query set containing 50 natural language queries over DBPedia.

- Approach was capable of answering 70% of the queries.
- Mean reciprocal rank = 0.614, avg. precision = 0.487, avg. recall = 0.572.
- The relatedness measure:
  - Showed high discrimination node selection (2.81 σ above average of node relatedness measures).
  - Deliveries a precise and flexible semantic matching between query and vocabulary terms.

Acknowledgements

This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/11380 (Lion-2).