AskNow
A Framework for Natural Language Query Formalization in SPARQL

Mohnish Dubey
University of Bonn

Sourish Dasgupta
DA-IICT, India

Ankit Sharma
University at Buffalo

Konrad Höffner
AKSW, Leipzig

Jens Lehmann
University of Bonn

13th International Conference, ESWC 2016, Crete, Greece
Outline

1. Introduction
2. Key Contribution
3. Architectural Pipeline
4. NL to NQS
5. NQS to SPARQL
6. Evaluation
Introduction

- Natural language query formalization (NL-QF)
- Based on linguistic analysis of natural language query
- NL query $\Rightarrow$ Intermediate Structure $\Rightarrow$ SPARQL
Key Contributions

- NQS: novel chunker-styled query re-rephrasing structure.
- Intermediary model for translating NL queries into formal queries.
- NQS to SPARQL converter algorithm.
- Evaluation is also performed at intermediate language level.
Definition of NQS and its Importance?

- “Normalized Query Structure” is a **Linguistic analyzer**
  - Identifies and distinguishes between:
    - Desired query output information (D) and
    - Query input information (I)
  - Establishes mutual semantic relationship (R)
  - Detects main entity

- Adapts to query paraphrasing by performing syntactic normalization

- Enables translation of NL query to Formal Language query (SPARQL)
Architecture Pipeline
NL to NQS

Query Reframe
Auxiliary Relation
NER Merger
Quantify Merger
POS tag Merger
NQS Instance

Token Merger
In which country is Bonn located?

which country is Bonn located In?
Query Reframe

Auxiliary Relation: Marks the auxiliary relation such as “is”, “are” and other similar forms of relation

NER Merger

Quantify Merger

POS tag Merger

NQS Instance
What/ WP, is/ REL1, the/ DT, birth/ NN, place/ NN, of/ IN, Cristiano/ NNP, Ronaldo/ NNP ?/.

What/ WP, is/ REL1, the/ DT, birth/ NN, place/ NN, of/ IN, Cristiano Ronaldo/ NNP-NER, ?/.
NL to NQS

Query Reframe

Auxiliary Relation

NER Merger

Quantify Merger

POS tag Merger

NQS Instance

What/WP, are/REL1, some/DT, fresh/JJ, water/NN, lakes/NNS, in/IN, lower/JJR, Himalayas/NNPS, ?/.

What/WP, are/REL1, some fresh_NM water/NN, lakes/NNS, in/IN, lower_NM Himalayas/NN, ?/.
What/WP, is/REL1, the birth/NN, place/NN, of/IN, Christiano Ronaldo/NNP-NER, ?/.

What/WP, is/REL1, the birth_place/NN, of/IN, Christiano Ronaldo/NNP-NER, ?/.
NQS to SPARQL

NQS parser
DBpedia Spotlight
WordNet
BOA pattern library
SPARQL Generator
NQS to SPARQL

Decide query type: Boolean, Count, Ranking, List, Data property

NQS parser
DBpedia Spotlight
WordNet
BOA pattern library
SPARQL Generator
Query Input (I) maps to DBpedia equivalent


DBpedia Spotlight Annotation

NQS parser
DBpedia Spotlight
WordNet
BOA pattern library
SPARQL Generator
To get synonym of Relation or Desire
writer
WordNet Synonyms

NQS to SPARQL
NQS parser
DBpedia Spotlight
WordNet
BOA pattern library
SPARQL Generator

NQS to SPARQL

NQS parser
DBpedia Spotlight
WordNet
BOA pattern library
SPARQL Generator
Final SPARQL query based on above steps is generated
Evaluation
Research Hypothesis

1. Syntactic Robustness: there should not be any mismatch between the POS tag of a linguistic constituent and its corresponding NQS cell.

2. Sensitivity to Structural Variation: With variation in syntactic structure of the query, NQS should still correctly identify the Query Desire (D).

3. Semantic Accuracy: Query Desire(D) and its Relation(R) with Query Inputs(I) has been correctly identified.

4. Correct Answer to the Query: NQS to SPARQL provide the correct answer.
1. **Syntactic Robustness**

Dataset: Microsoft Encarta 98, OWL-S TC, QALD 5

Parameter: Structuring Coverage(SC): SC-Precision, SC-Recall, SC- F1

Result:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-Precision</td>
<td>98.53</td>
</tr>
<tr>
<td>SC-Recall</td>
<td>99.45</td>
</tr>
<tr>
<td>SC- F1</td>
<td>98.99</td>
</tr>
</tbody>
</table>
2. Sensitivity to Structural Variation

Dataset: OWL-S TC query dataset and QALD-4 dataset (both in three versions)

Parameter: Variational-Precision (VP), Variational-Recall (VR)

Result:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variational-Recall (VR)</td>
<td>96.51</td>
</tr>
<tr>
<td>Variational-Precision (VP)</td>
<td>85.18</td>
</tr>
</tbody>
</table>
3. Semantic Accuracy

Dataset: OWL-S TC query dataset and QALD-4 dataset (both in three versions)

Parameter: Semantic-Precision (SP), Semantic-Recall (SR)

Result:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic-Precision (SP)</td>
<td>81.03</td>
</tr>
<tr>
<td>Semantic-Recall (SR)</td>
<td>91.81</td>
</tr>
</tbody>
</table>
### 4. Accuracy of the AskNow System

**Dataset:** QALD-5

**Parameters:** as stated by QALD-5 benchmark

**Result:**

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
<th>F1 Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xser</td>
<td>0.72</td>
<td>0.74</td>
<td>0.73</td>
<td>0.63</td>
</tr>
<tr>
<td>AskNow</td>
<td>0.63</td>
<td>0.60</td>
<td>0.61</td>
<td>0.33</td>
</tr>
<tr>
<td>QAnswer</td>
<td>0.35</td>
<td>0.46</td>
<td>0.40</td>
<td>0.30</td>
</tr>
<tr>
<td>APEQ</td>
<td>0.48</td>
<td>0.40</td>
<td>0.44</td>
<td>0.23</td>
</tr>
</tbody>
</table>
Result Discussion

1. Fail case of NQS:
   Which animals are critically endangered?

2. Improvement Required in NQS to SPARQL in
   a. Relation annotation
   b. SPARQL generation
Conclusion

- NQS serves as a robust intermediary model for Query Formalization.
- NQS is a good linguistic analyzer.
- NQS is independent from the target Knowledge Base.
- Query are distinguished from a natural language perspective as *Simple, Complex and Compound*
Thank You

Contact:
Mohnish Dubey
dubey@cs.uni-bonn.de
http://sda.cs.uni-bonn.de/mohnish-dubey/
Looking forward for questions and discussion
You may AskNow ;-)