



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

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 ⇨ Intermediate Structure ⇨ SPARQL

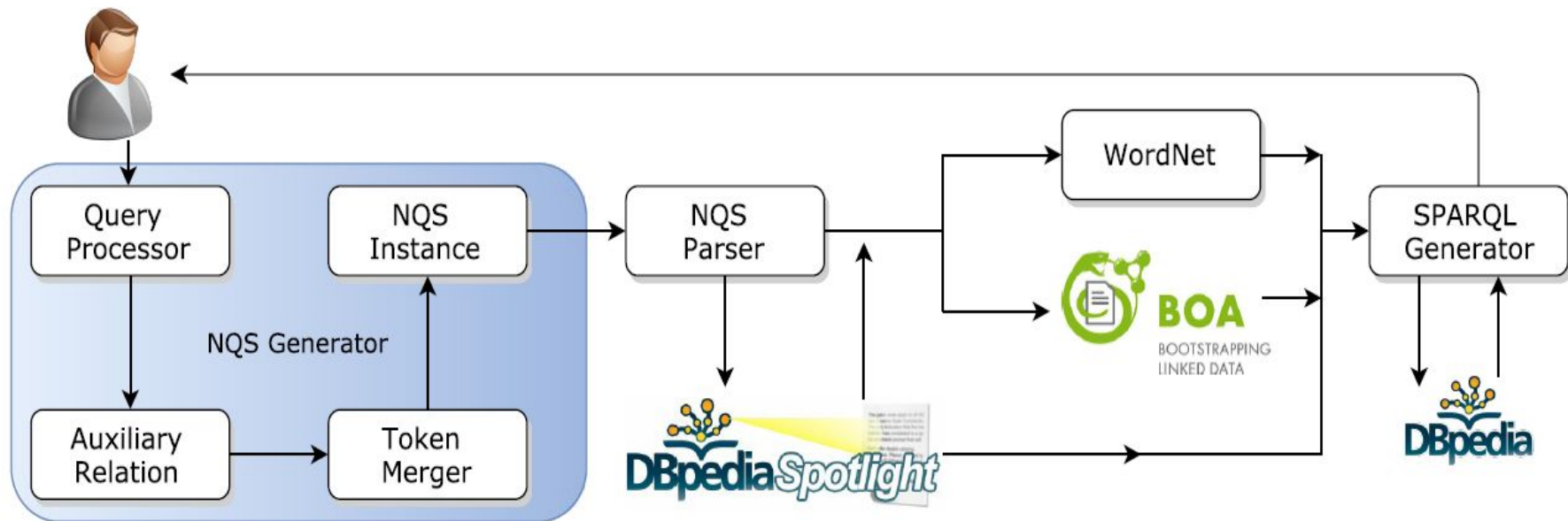
Key Contributions

- 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



NL to NQS

Query Reframe



In which country is Bonn located ?

which country is Bonn located In ?

Auxiliary Relation

NER Merger


Quantify Merger

POS tag Merger

NQS Instance

NL to NQS

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

NL to NQS

Query Reframe

Auxiliary Relation

NER Merger



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

Quantify Merger

What/WP, is/REL1, the/DT, birth/NN, place/NN, of/IN,
Christiano Ronaldo/NNP-NER, ?/.

POS tag Merger

NQS Instance

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, ?/.

NL to NQS

Query Reframe

Auxiliary Relation

NER Merger

Quantify Merger

POS tag Merger



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

NQS Instance

What/WP, is/REL1, the birth_place/NN, of/IN, Cristiano Ronaldo/NNP-NER, ?/.

NL to NQS

Query Reframe

Auxiliary Relation

NER Merger

Quantify Merger

POS tag Merger

NQS Instance



[WH] = What, [R1] = are, [[DQ] = some [DM] = fresh water [D] = lakes], [R2] = in, [[IM] = lower [I] = Himalayas]

NQS parser

DBpedia Spotlight

WordNet

BOA pattern library

SPARQL Generator

NQS to SPARQL

NQS to SPARQL

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

← NQS parser

DBpedia Spotlight

WordNet

BOA pattern library

SPARQL Generator

NQS to SPARQL

NQS parser

← DBpedia Spotlight

WordNet

BOA pattern library

SPARQL Generator

Query Input (I) maps to DBpedia equivalent
[[I] = Himalayas] ← → <http://dbpedia.org/resource/Himalayas>

↑
DBpedia Spotlight
Annotation

NQS to SPARQL

NQS parser

DBpedia Spotlight

← WordNet

BOA pattern library

SPARQL Generator

To get synonym of Relation or Desire

[[D]=writer] → author ↔ http://dbpedia.org/ontology/author

writer



WordNet Synonyms

NQS to SPARQL

NQS parser

DBpedia Spotlight

WordNet

← BOA pattern library

SPARQL Generator

[wife of] ↔ <http://dbpedia.org/ontology/spouse>

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:

SC-Precision	98.53
SC-Recall	99.45
SC- F1	98.99

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:

Variational-Recall (VR)	96.51
Variational-Precision (VP)	85.18

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:

Semantic-Precision (SP)	81.03
Semantic-Recall (SR)	91.81

4. Accuracy of the AskNow System

Dataset: QALD-5

Parameters: as stated by QALD-5 benchmark

Result:

	Recall	Precision	F1	F1 Global
Xser	0.72	0.74	0.73	0.63
AskNow	0.63	0.60	0.61	0.33
QAnswer	0.35	0.46	0.40	0.30
APEQ	0.48	0.40	0.44	0.23

Result Discussion

1. Fail case of NQS:

Which animals are critically endangered?

[WH] = Which, [R1] = are, [D] = animals, [R2] = endangered, [I] = critically

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 ;-)