

AKSW Colloquium

PhD progress & CubeQA

Konrad Höffner

April 29, 2016

Section 1

PhD progress

February 2015–March 2016

Thesis Core Progress

| | |
|----------------|-------------|
| LinkedSpending | ✓ |
| QA Survey | revising |
| CubeQA Short | ✓ |
| CubeQA Long | revising |
| CubeQA Journal | future work |
| ProcQA(?) | future work |

Thesis Core Progress

| | | |
|----------------|-------------|----------|
| LinkedSpending | ✓ | |
| QA Survey | revising | April 14 |
| CubeQA Short | ✓ | |
| CubeQA Long | revising | April 30 |
| CubeQA Journal | future work | 2016 |
| ProcQA(?) | future work | 2016 |

2015–now Authorship (Published & Pending)

First Author

- ▶ QA Survey
- ▶ CubeQA long

Coauthor

- ▶ kOre
- ▶ AskNow
- ▶ *path

2015–now Other Work

- ▶ Supervisor for Software Engineering Student Internship “Interactive Budget Calculator for the City of Leipzig”
- ▶ Creation of QALD 6 Task 3 Benchmark (QA on RDF Data Cubes)
- ▶ GeoKnow service administration (Website, Wiki, Blog)
- ▶ GeoKnow documentation
- ▶ Reviews for KESW, SWJ, . . .
- ▶ aksw.org Website Data Officer MOLE

2015–now

What went well

- ▶ QA Survey (thanks for all the help!)
- ▶ Coauthorships
- ▶ Other Work

What didn't

- ▶ publication of CubeQA long version

Lessons Learned

- ▶ focus on formalization
- ▶ get more authors on board early to combine strenghts and offset weaknesses
- ▶ projects need promotion to become noticed

Section 2

CubeQA

CubeQA

- ▶ Question Answering on RDF Data Cubes
- ▶ New research sub-field
- ▶ CubeQA project: Question Corpus, Benchmarks, TCQA algorithm
- ▶ Short Paper Preprint
<http://svn.aksw.org/papers/2014/cubeqa/short/public.pdf>
- ▶ Long Paper (rejected, revising)
http://svn.aksw.org/papers/2016/IJCAI_cubeqa/submission.pdf

CubeQA

- ▶ Question Corpus: realistic statistical questions from volunteers
- ▶ Benchmarks
 - ▶ single-dataset
 - ▶ choice-among-many-datasets, now QALD 6 Task 3 (current evaluation)
 - ▶ combination-among-many-datasets (in progress)
 - ▶ hard challenges: implied aggregations, subqueries,
- ▶ TCQA: Tree-based Cube Question Answering, first algorithm that can answer these Questions
- ▶ $F_1 \approx 0.4$ on QALD 6 Task 3 train

Problems of CubeQA

- ▶ long version of CubeQA not published yet
- ▶ new research sub-field, needs groundwork
 - ▶ needs to solve too many things at once, lack of focus
 - ▶ tension between completeness and understandability
- ▶ $F_1 = 0.4$ not deemed high enough (even though only CubeQA can do it at all)
- ▶ formalization

CubeQA Paper Current State

http://svn.aksw.org/papers/2016/IJCAI_cubeqa/submission.pdf