



Talk 1: Introduction to OKBQA framework

Jin-Dong Kim

Database Center for Life Science (DBCLS)

Contact: jdkim@dbcls.rois.ac.jp

22, July, 2016

Agenda

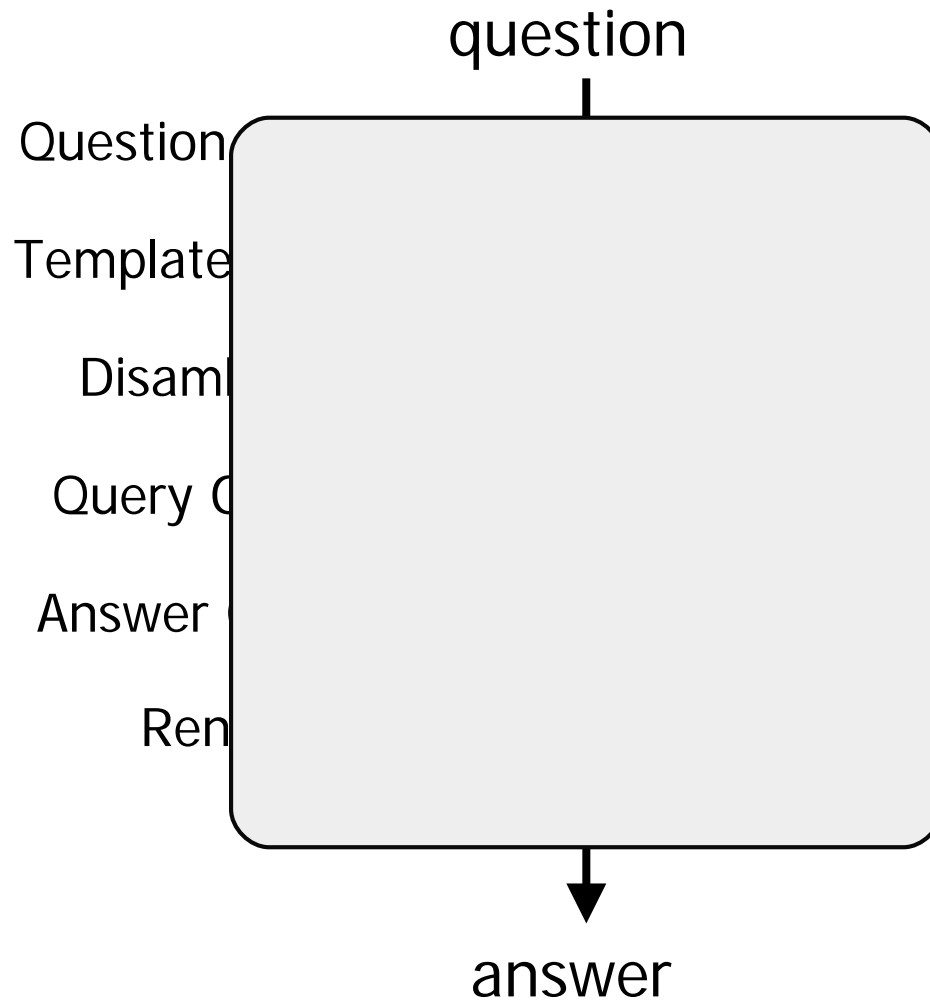
1. Introduction
2. Components
3. Principles
4. Resources
5. Conclusions



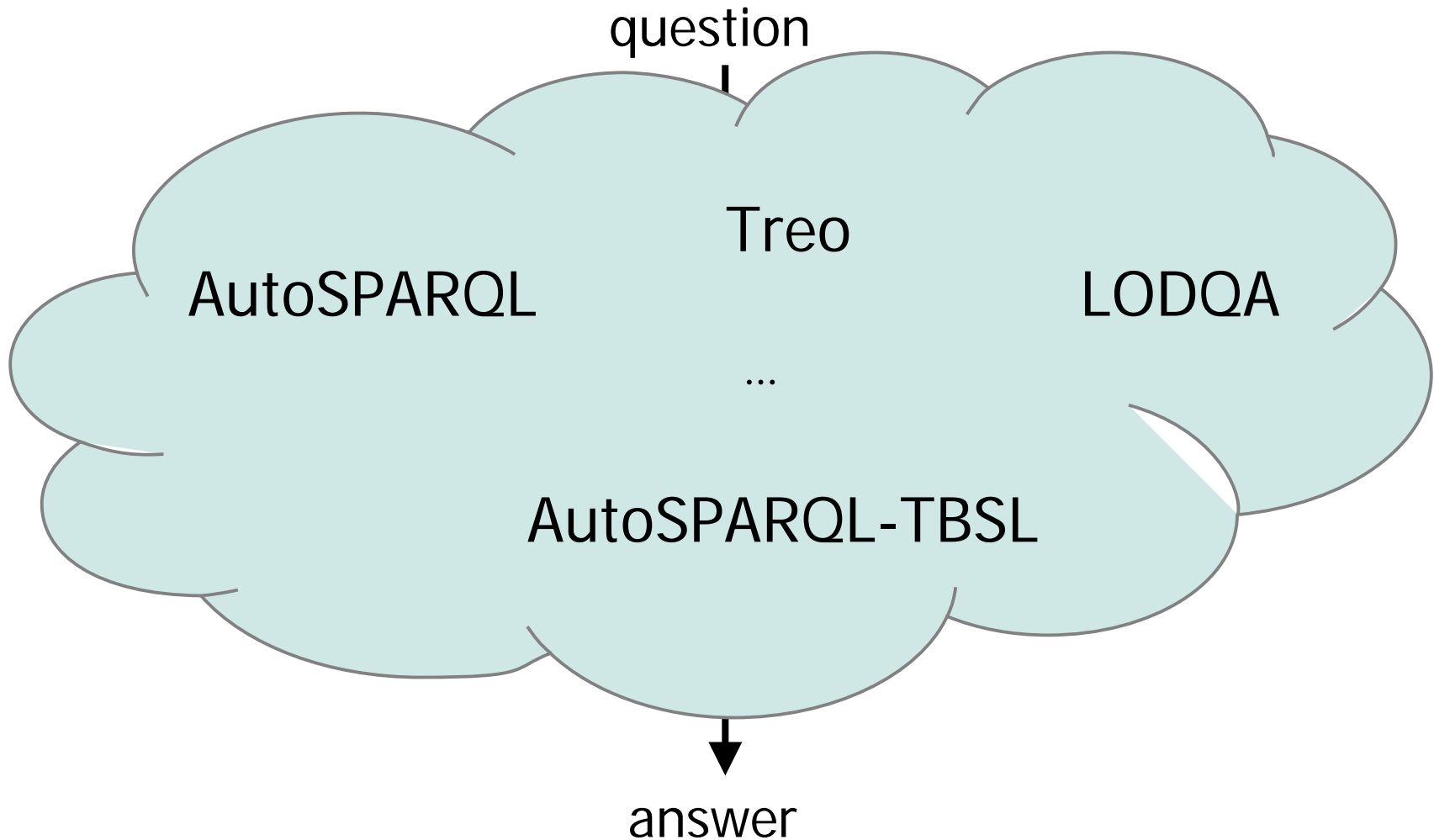
Introduction



Question-Answering

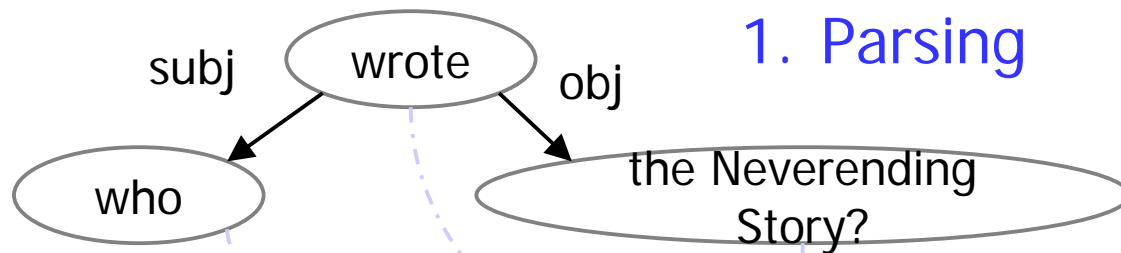


Existing Systems



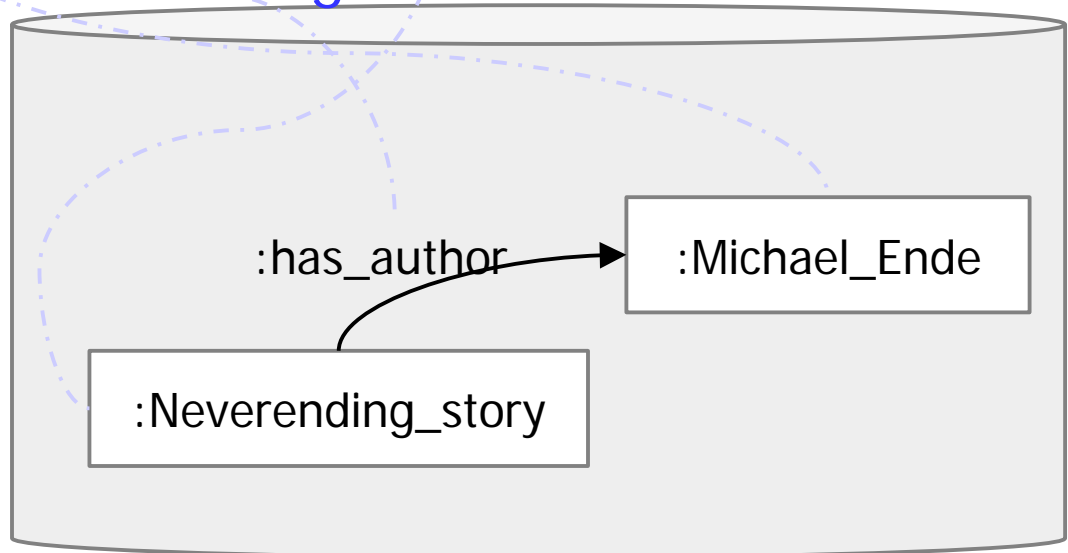
Typical Problems

Who wrote the Neverending Story?



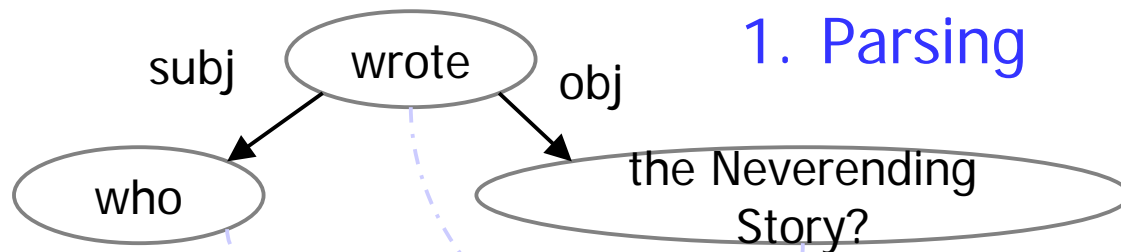
2. Lexical matching

3. structural matching



Typical Problems

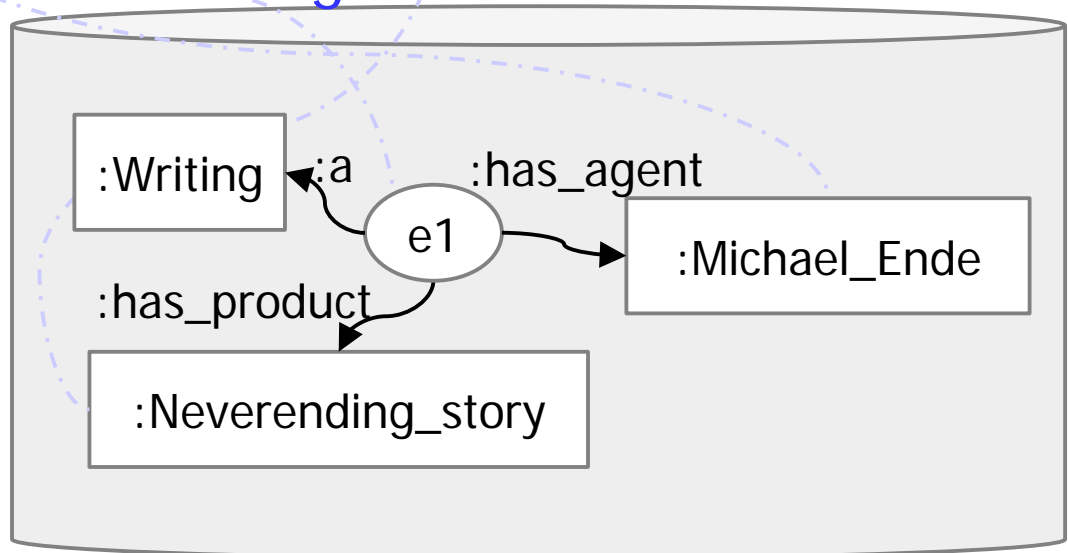
Who wrote the Neverending Story?



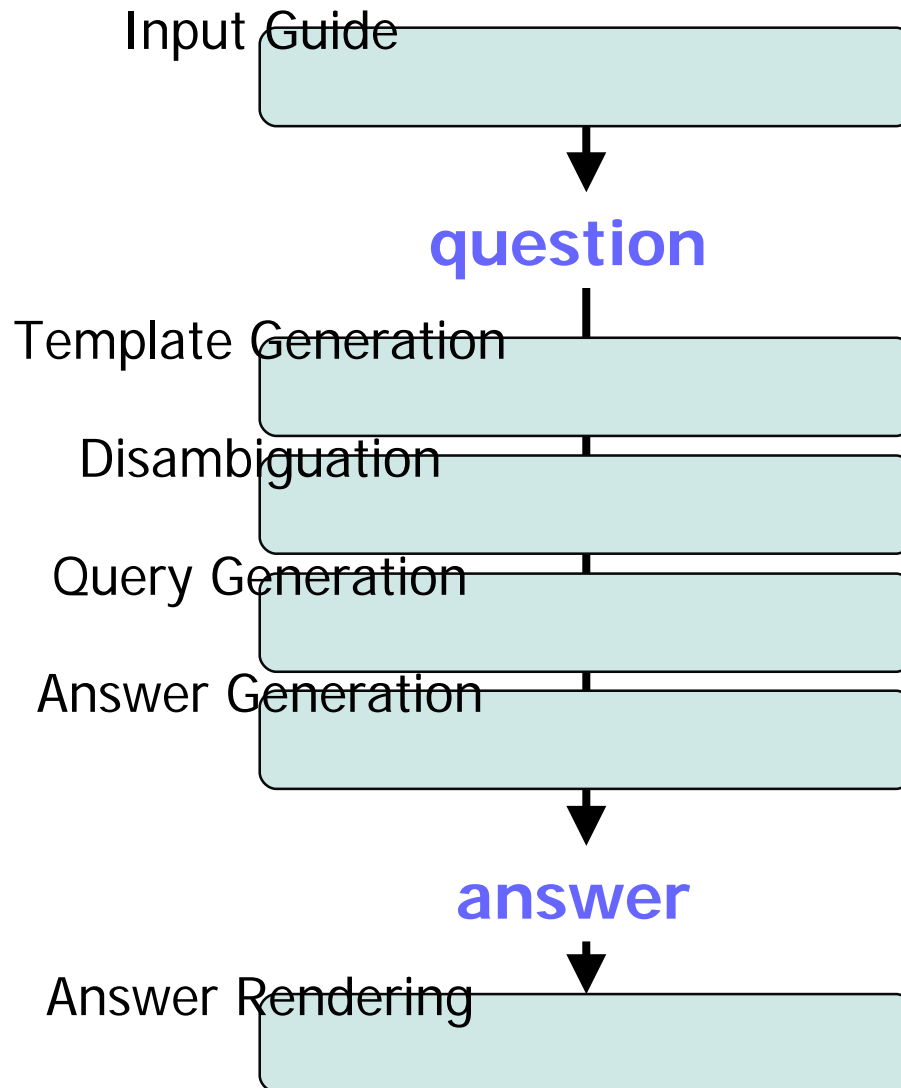
1. Parsing

2. Lexical matching

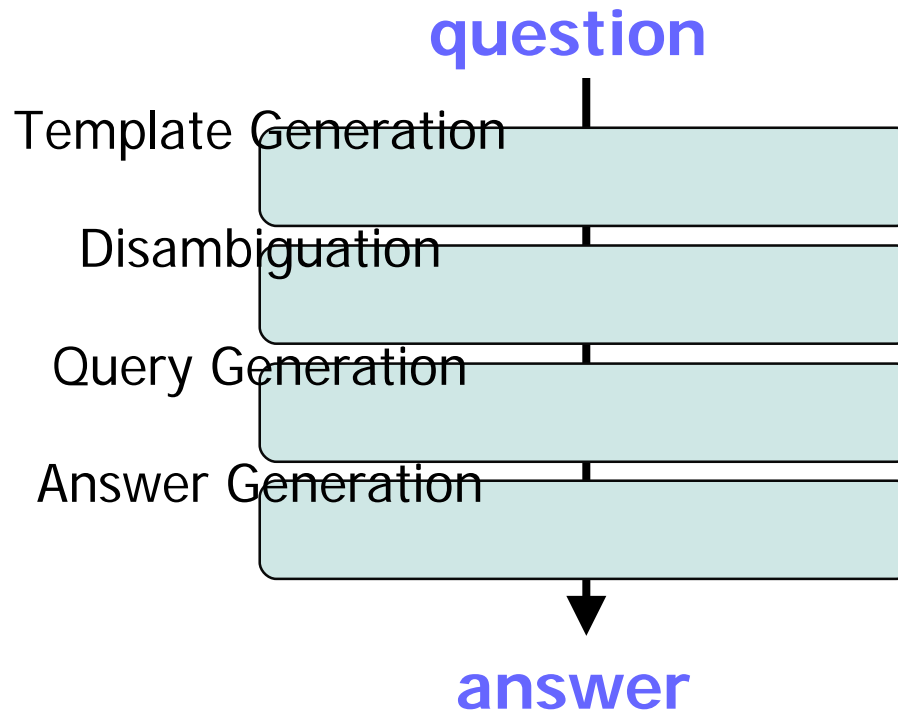
3. structural matching



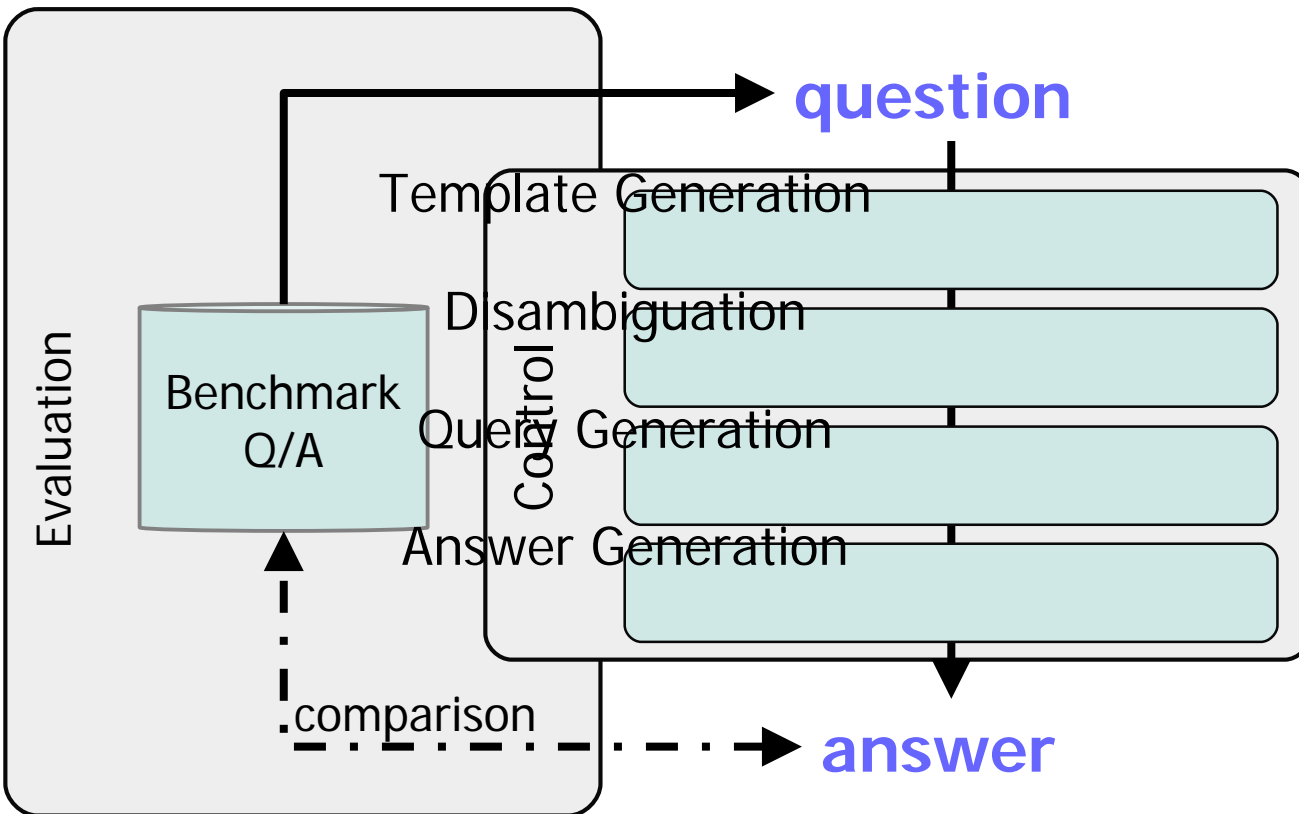
OKBQA framework



OKBQA framework (core)



OKBQA framework (Benchmark)



Components

Template Generation Module (TGM)

- Input: natural language query
- output: structured description of sparql

```
{  
  "language": "en",  
  "string": "which river flows in Seoul?"  
}
```

```
{  
  "query": "SELECT ?v4 WHERE { ?v4 ?v2 ?v6 ; ?v7 ?v3 . }",  
  "score": "1.0",  
  "slots": [  
    { "s": "v2", "p": "is", "o": "rdf:Property" },  
    { "s": "v2", "p": "verbalization", "o": "flows" },  
    { "s": "v6", "p": "is", "o": "rdf:Resource|rdfs:Literal" },  
    { "s": "v6", "p": "verbalization", "o": "Seoul" },  
    { "s": "v7", "p": "is", "o": "" },  
    { "s": "v3", "p": "is", "o": "rdf:Class" },  
    { "s": "v3", "p": "verbalization", "o": "river" },  
  ]  
}
```

Disambiguation Module (DM)

Input: SPARQL templates

Output: disambiguation information for each slot

```
"ned": [
  {
    "classes": [
      { "var": "v3", "value": "http://dbpedia.org/ontology/River", "score": 0.57 }
    ],
    "entities": [
      { "var": "v6", "value": "http://dbpedia.org/resource/Seoul", "score": 1 }
    ],
    "literals": [
      { "var": "v6", "value": "Seoul", "score": 1 }
    ],
    "properties": [
      { "var": "v2", "value": "http://dbpedia.org/ontology/city", "score": 0.62 },
      { "var": "v2", "value": "http://dbpedia.org/property/flower", "score": 0.53 }
    ],
    "score": 1
  }
]
```

Query Generation Module (QGM)

Input: SPARQL templates with disambiguation

Output: corresponding SPARQL queries

```
[
  {
    "query": "SELECT ?v4 WHERE {
      ?v4 <http://dbpedia.org/ontology/city> <http://dbpedia.org/resource/Seoul> ;
      ?v7 <http://dbpedia.org/ontology/River> .
      FILTER (str(?v7) IN ("http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
                           "http://www.w3.org/2000/01/rdf-schema#subClassOf"))
    }",
    "score": 0.5
  },
  ...
]
```

Answer Generation Module (AGM)

Input: SPARQL queries

Output: Answers (URIs, Literals)

```
http://dbpedia.org/resource/Han_River_(Korea)
http://dbpedia.org/resource/Cheonggyecheon
http://dbpedia.org/resource/Cheonggyecheon
http://dbpedia.org/resource/Ara_Canal
http://dbpedia.org/resource/Jungnangcheon
http://dbpedia.org/resource/Jungnangcheon
http://dbpedia.org/resource/Han_River_(Korea)
http://dbpedia.org/resource/Yanghwa_Bridge
http://dbpedia.org/resource/Hannam_Bridge
http://dbpedia.org/resource/Banghwa_Bridge
http://dbpedia.org/resource/Jamsu_Bridge
http://dbpedia.org/resource/Hannam_Bridge
http://dbpedia.org/resource/Banghwa_Bridge
http://dbpedia.org/resource/Jamsu_Bridge
http://dbpedia.org/resource/Andrew_Kim-Taegon
http://dbpedia.org/resource/Andrew_Kim-Taegon
http://dbpedia.org/resource/Andrew_Kim-Taegon
...
```

Principles

JSON for I/O

- “a simplified XML”
- Elements
 - Object {}, a.k.a. hash, dictionary, key-value pairs
 - Array []
- Example

```
{
  "query": "SELECT ?v4 WHERE { ?v4 ?v2 ?v6 ; ?v7 ?v3 . } ",
  "score": "1.0",
  "slots": [
    { "s": "v2", "p": "is", "o": "rdf:Property" },
    { "s": "v2", "p": "verbalization", "o": "flows" },
    { "s": "v6", "p": "is", "o": "rdf:Resource|rdfs:Literal" },
    { "s": "v6", "p": "verbalization", "o": "Seoul" },
    { "s": "v7", "p": "is", "o": "" },
    { "s": "v3", "p": "is", "o": "rdf:Class" },
    { "s": "v3", "p": "verbalization", "o": "river" },
  ]
}
```

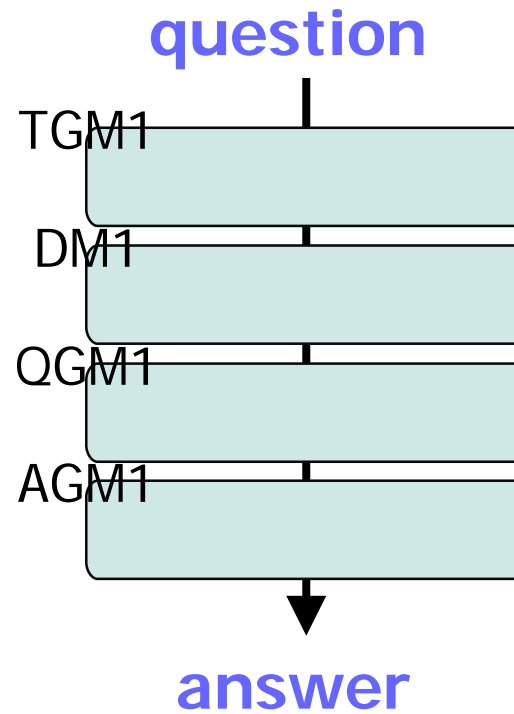
Components as REST services

- HTTP POST (or GET) and response
- Example (Post a JSON object in the body of req.)

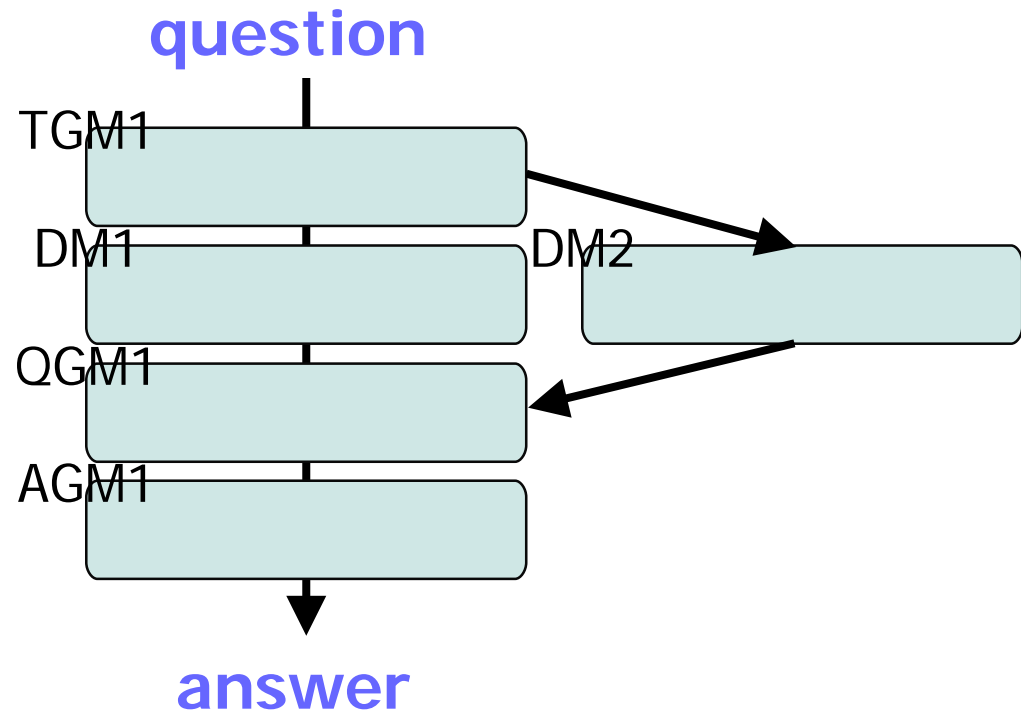
```
curl -H "Content-Type: application/json"  
http://ws.okbqa.org:1555/templategeneration/templator -d '  
{  
  "string": "Which rivers flow through Seoul?",  
  "language": "en"  
}  
'
```

```
[  
  {  
    "score": "1.0",  
    "slots": [  
      { "s": "v3", "p": "is", "o": "rdf:Class" },  
      ...  
    ]  
  }  
  ...  
]
```

Flexible Workflow



Flexible Workflow



Benchmark Evaluation

- To allow evaluation of improvements



Resources

Repository

- A central venue for collection of OKBQA modules
- URL
 - <http://repository.okbqa.org/>



List

New Component

Title	Category	Description	Maintainer
Sparqlator	Query generation	Description It is a wrapper to call the GraphFinder::sparqlator method, using ...	jindong.kim@gmail.com
Templator	Template generation	Description Templator is a module for dependency-driven SPARQL template ...	cunger@cit-ec.uni-bielefeld.de
AGDISTIS-KO	Disambiguation	# Description Disambiguator uses the AGDISTIS web service to disambiguate...	zakria.ai@gmail.com
SRDF 1.0	Other	Description SRDF is a Korean Open Information Extraction system. It is desi...	nam.sangha@gmail.com
C2K 2.0	Other	C2K is a knowledge acquisition system that extracts triples of DBpedia prope...	jjiseong@kaist.ac.kr
Disambiguator	Disambiguation	Description Disambiguator is a module to disambiguate class, properties and...	hahmyg@kaist.ac.kr
TGM JAVA WRAPP...	Template generation	Description It is a wrapper for Java Template Generation Module(TGM). If yo...	nam.sangha@gmail.com
DM JAVA WRAPPER	Disambiguation	Description It is a warpper for Java Disambiguation Module(DM). If you want...	nam.sangha@gmail.com

Controller

- To allow control and execution of the workflow
- URL
 - <http://ws.okbqa.org/~testuser02/>

Conclusions



Conclusions

- OKBQA framework enables
 - Modular development
 - Distributed development
 - Testing different workflows
 - Benchmarking performance of workflows
- Caveat
 - OKBQA framework is optimized for
 - development phase (collaborative development)
 - but not for
 - production phase (fast processing)
- License condition
 - OKBQA modules are recommended to be released under MIT license terms (the most liberal one)