

Qanary – A Methodology for Vocabulary-driven Open Question Answering Systems

Andreas Both¹, Dennis Diefenbach², Kuldeep Singh³, Saadeeh Shekarpour⁴, Didier Cherix⁵, Christoph Lange³

¹Mercateo AG, Germany, ²Laboratoire Hubert Curien, Saint-Etienne, France, ³Fraunhofer IAIS, Sankt Augustin, Germany, ⁴University of Bonn, Bonn, Germany, ⁵FLAVIA IT-Management GmbH, Germany

2016-06-01, 13th European Semantic Web Conference

Qanary an initiative of the WDAqua project

“Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN

CONTEXT

Field: Question Answering

- embedded into WDAqua project
 - “Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN
 - <http://wdaqua.informatik.uni-bonn.de>



CONTEXT

Field: Question Answering

- embedded into WDAqua project
 - “Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN
 - <http://wdaqua.informatik.uni-bonn.de>

Observations

- state of the art not as advanced as expected
- see also QALD challenge



CONTEXT

Field: Question Answering

- embedded into WDAqua project
 - “Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN
 - <http://wdaqua.informatik.uni-bonn.de>

Observations

- state of the art not as advanced as expected
- see also QALD challenge

Reasons: How are question answering systems created?



CONTEXT

Field: Question Answering

- embedded into WDAqua project
 - “Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN
 - <http://wdaqua.informatik.uni-bonn.de>

Observations

- state of the art not as advanced as expected
- see also QALD challenge

Reasons: How are question answering systems created?

- in general: hard and complex task



CONTEXT

Field: Question Answering

- embedded into WDAqua project
 - “Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN
 - <http://wdaqua.informatik.uni-bonn.de>

Observations

- state of the art not as advanced as expected
- see also QALD challenge

Reasons: How are question answering systems created?

- in general: hard and complex task
- cumbersome and inefficient



CONTEXT

Field: Question Answering

- embedded into WDAqua project
 - “Answering Questions using Web Data” (WDAqua) is a Marie Skłodowska-Curie ITN
 - <http://wdaqua.informatik.uni-bonn.de>

Observations

- state of the art not as advanced as expected
- see also QALD challenge

Reasons: How are question answering systems created?

- in general: hard and complex task
- cumbersome and inefficient
 - *lack of methodology* for creating question answering systems



MOTIVATION



MOTIVATION

Observations

- limited compatibility
- use predefined QA process
- less reusable implementations
- limited semantics



MOTIVATION

Observations

- limited compatibility
- use predefined QA process
- less reusable implementations
- limited semantics

Derived demands

- + interoperable infrastructure
- + exchangeable components
- + flexible granularity
- + isolation of components



MOTIVATION

Observations

- limited compatibility
- use predefined QA process
- less reusable implementations
- limited semantics

Derived demands

- + interoperable infrastructure
- + exchangeable components
- + flexible granularity
- + isolation of components

Goals

1. easy-to-build QA systems on-top of reusable components
2. establish an ecosystem of components for QA systems



MOTIVATION

Observations

- limited compatibility
- use predefined QA process
- less reusable implementations
- limited semantics

Derived demands

- + interoperable infrastructure
- + exchangeable components
- + flexible granularity
- + isolation of components

Goals

1. easy-to-build QA systems on-top of reusable components
 2. establish an ecosystem of components for QA systems
- tackle the challenge of retrieving data from large data sets



MOTIVATION

Observations

- limited compatibility
- use predefined QA process
- less reusable implementations
- limited semantics

Derived demands

- + interoperable infrastructure
- + exchangeable components
- + flexible granularity
- + isolation of components

Goals

1. easy-to-build QA systems on-top of reusable components
 2. establish an ecosystem of components for QA systems
- tackle the challenge of retrieving data from large data sets
- **best-of-breed QA system and QA components**



IDEA: KNOWLEDGE-DRIVEN QA SYSTEM REPRESENTATION

Requirements of knowledge perspective



IDEA: KNOWLEDGE-DRIVEN QA SYSTEM REPRESENTATION

Requirements of knowledge perspective

1. abstract knowledge representation
 - o advantage: independent representation



IDEA: KNOWLEDGE-DRIVEN QA SYSTEM REPRESENTATION

Requirements of knowledge perspective

1. abstract knowledge representation
 - advantage: independent representation
2. align the input/output of the each component
 - on a logical and sound level



IDEA: KNOWLEDGE-DRIVEN QA SYSTEM REPRESENTATION

Requirements of knowledge perspective

1. abstract knowledge representation
 - advantage: independent representation
2. align the input/output of the each component
 - on a logical and sound level

→ *Qanary* methodology for creating question answering systems



1. ABSTRACT KNOWLEDGE REPRESENTATION

Idea

Represent all the knowledge about a question



1. ABSTRACT KNOWLEDGE REPRESENTATION

Idea

Represent all the knowledge about a question

requirements for knowledge representation

- self-describing, sound knowledge representation
- represent provenance for (all) information
- represent trust for (all) information



1. ABSTRACT KNOWLEDGE REPRESENTATION

Idea

Represent all the knowledge about a question

requirements for knowledge representation

- self-describing, sound knowledge representation
- represent provenance for (all) information
- represent trust for (all) information

→ use RDF



1. ABSTRACT KNOWLEDGE REPRESENTATION

Idea

Represent all the knowledge about a question

requirements for knowledge representation

- self-describing, sound knowledge representation
- represent provenance for (all) information
- represent trust for (all) information

→ use RDF

→ **“qa” vocabulary already established**

K. Singh, A. Both, D. Diefenbach, and S. Shekarpour. “Towards a message-driven vocabulary for promoting the interoperability of question answering systems.” In Proc. of the 10th IEEE Int. Conf. on Semantic Computing (ICSC), 2016



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`
- `oa:hasTarget`



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- oa:Annotation
- oa:hasTarget
- oa:hasBody



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- oa:Annotation
- oa:hasTarget
- oa:hasBody
- oa:annotatedAt



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- oa:Annotation
- oa:hasTarget
- oa:hasBody
- oa:annotatedAt
- oa:annotatedBy



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`
 - `oa:hasTarget`
 - `oa:hasBody`
 - `oa:annotatedAt`
 - `oa:annotatedBy`
- `<myIRI> a oa:Annotation;`



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- oa:Annotation
- oa:hasTarget
- oa:hasBody
- oa:annotatedAt
- oa:annotatedBy

```
<myIRI> a oa:Annotation;  
      oa:hasTarget <questionIRI> ;
```



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- oa:Annotation
- oa:hasTarget
- oa:hasBody
- oa:annotatedAt
- oa:annotatedBy

```
<myIRI> a oa:Annotation;  
    oa:hasTarget <questionIRI> ;  
    oa:hasBody <TextSelector> ;
```



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`
- `oa:hasTarget`
- `oa:hasBody`
- `oa:annotatedAt`
- `oa:annotatedBy`

```
<myIRI> a oa:Annotation;  
    oa:hasTarget <questionIRI> ;  
    oa:hasBody <TextSelector> ;  
    oa:annotatedBy <DBpediaSpotlight> ;
```



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`
- `oa:hasTarget`
- `oa:hasBody`
- `oa:annotatedAt`
- `oa:annotatedBy`

```
<myIRI> a oa:Annotation;  
  oa:hasTarget <questionIRI> ;  
  oa:hasBody <TextSelector> ;  
  oa:annotatedBy <DBpediaSpotlight> ;  
  oa:annotatedAt "...^^xsd:date ;
```



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- oa:Annotation
- oa:hasTarget
- oa:hasBody
- oa:annotatedAt
- oa:annotatedBy

```
<myIRI> a oa:Annotation;  
    oa:hasTarget <questionIRI> ;  
    oa:hasBody <TextSelector> ;  
    oa:annotatedBy <DBpediaSpotlight> ;  
    oa:annotatedAt "...^^xsd:date ;
```

qa vocabulary



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`
- `oa:hasTarget`
- `oa:hasBody`
- `oa:annotatedAt`
- `oa:annotatedBy`

```
<myIRI> a oa:Annotation;  
  oa:hasTarget <questionIRI> ;  
  oa:hasBody <TextSelector> ;  
  oa:annotatedBy <DBpediaSpotlight> ;  
  oa:annotatedAt "...""^^xsd:date ;  
  qa:score "...""^^xsd:float.
```

qa vocabulary

- e.g., new property:
`qa:score`



KNOWLEDGE REPRESENTATION USING THE qa VOCABULARY

Web Annotation Data Model (WADM)

(W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>)

- `oa:Annotation`
- `oa:hasTarget`
- `oa:hasBody`
- `oa:annotatedAt`
- `oa:annotatedBy`

```
<myIRI> a oa:Annotation;
  oa:hasTarget <questionIRI> ;
  oa:hasBody <TextSelector> ;
  oa:annotatedBy <DBpediaSpotlight> ;
  oa:annotatedAt "...^^xsd:date ;
  qa:score "...^^xsd:float.
```

qa vocabulary

- e.g., new property:
`qa:score`
- many new Annotation classes

```
qa:Question
  rdfs:subClassOf oa:Annotation.

qa:Answer, ...
qa:Dataset, ...
qa:AnnotationQuestion, ...

...
```



FROM KNOWLEDGE REPRESENTATION TO METHODOLOGY

Advantages of using an ontology



FROM KNOWLEDGE REPRESENTATION TO METHODOLOGY

Advantages of using an ontology

- agnostic to question format (text, structured, audio, ...)



FROM KNOWLEDGE REPRESENTATION TO METHODOLOGY

Advantages of using an ontology

- agnostic to question format (text, structured, audio, ...)
- agnostic to question answering processing steps



FROM KNOWLEDGE REPRESENTATION TO METHODOLOGY

Advantages of using an ontology

- agnostic to question format (text, structured, audio, ...)
- agnostic to question answering processing steps
- agnostic to implementation
 - programming language
 - component granularity



FROM KNOWLEDGE REPRESENTATION TO METHODOLOGY

Methodology

1. abstract knowledge representation
 - advantage: independent representation
2. align the input/output of the each component
 - on a logical and sound level



FROM KNOWLEDGE REPRESENTATION TO METHODOLOGY

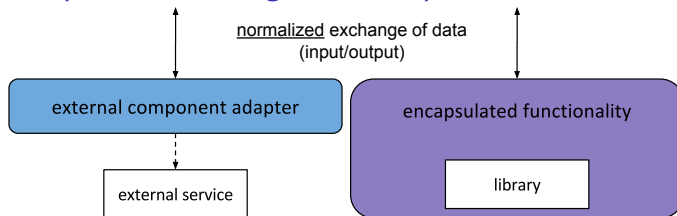
Methodology

1. abstract knowledge representation
 - advantage: independent representation
2. align the input/output of the each component
 - on a logical and sound level



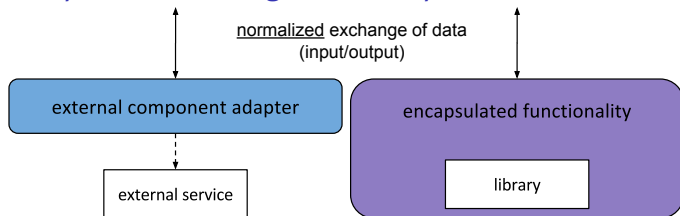
2. COMPONENT DATA ALIGNMENT

Component data alignment: 2 options



2. COMPONENT DATA ALIGNMENT

Component data alignment: 2 options

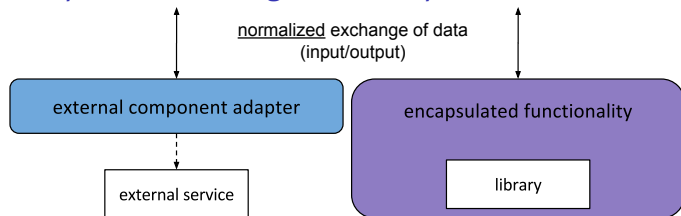


alignment of input/output of each component with qa



2. COMPONENT DATA ALIGNMENT

Component data alignment: 2 options



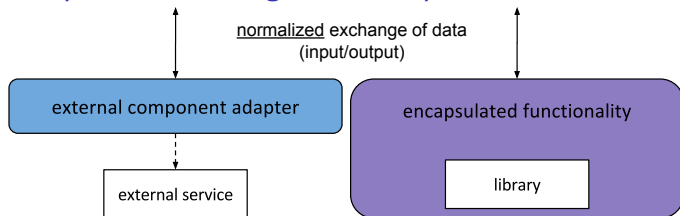
alignment of input/output of each component with qa

- input represented using **qa** (RDF)
 - input required for the component C



2. COMPONENT DATA ALIGNMENT

Component data alignment: 2 options



alignment of input/output of each component with qa

- input represented using qa (RDF)
 - input required for the component C
- output from the component C
 - output represented using qa (RDF)



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa
if component provides output using a presentation as



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa
if component provides output using a presentation as

... semantic data (RDF)



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa

if component provides output using a presentation as

... semantic data (RDF)

- logical representation of alignment
 - ontology alignment (OWL, DOL)

- NER/NED

- DBpedia Spotlight (NIF)

- P. N. Mendes, M. Jakob, A. Garcia-Silva, and Ch. Bizer: "DBpedia Spotlight:shedding light on the web of documents." In I-SEMANTICS, 2011



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa

if component provides output using a presentation as

... semantic data (RDF)

- logical representation of alignment
 - ontology alignment (OWL, DOL)
 - SPARQL query

- NER/NED

- DBpedia Spotlight (NIF)

- P. N. Mendes, M. Jakob, A. Garcia-Silva, and Ch. Bizer: "DBpedia Spotlight:shedding light on the web of documents." In I-SEMANTICS, 2011

- relation detection

- PATTY

- N. Nakashole, G. Weikum, and F. M. Suchanek. PATTY: "A taxonomy of relational patterns with semantic types." In EMNLP-CoNLL, 2012



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa

if component provides output using a presentation as

... semantic data (RDF)

- logical representation of alignment
 - ontology alignment (OWL, DOL)
 - SPARQL query

... non-semantic data (API, JSON, XML, CSV, ...)

- NER/NED

- DBpedia Spotlight (NIF)

P. N. Mendes, M. Jakob, A. Garcia-Silva, and Ch. Bizer: "DBpedia Spotlight:shedding light on the web of documents." In I-SEMANTICS, 2011

- relation detection

- PATTY

N. Nakashole, G. Weikum, and F. M. Suchanek. PATTY: "A taxonomy of relational patterns with semantic types." In EMNLP-CoNLL, 2012



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa

if component provides output using a presentation as

... semantic data (RDF)

- logical representation of alignment
 - ontology alignment (OWL, DOL)
 - SPARQL query

... non-semantic data (API, JSON, XML, CSV, ...)

- SPARQL query

● NER/NED

- DBpedia Spotlight (NIF)

P. N. Mendes, M. Jakob, A. Garcia-Silva, and Ch. Bizer: "DBpedia Spotlight:shedding light on the web of documents." In I-SEMANTICS, 2011

● relation detection

- PATTY

N. Nakashole, G. Weikum, and F. M. Suchanek. PATTY: "A taxonomy of relational patterns with semantic types." In EMNLP-CoNLL, 2012

● query construction

- SINA

S. Shekarpour, E. Marx, A.-C.N. Ngomo, and S. Auer. SINA: "Semantic interpretation of user queries for question answering on interlinked data." Web Semantics: Science, Services and Agents on the WWW, 2015



2. COMPONENT DATA ALIGNMENT

alignment of input/output of each component with qa

if component provides output using a presentation as

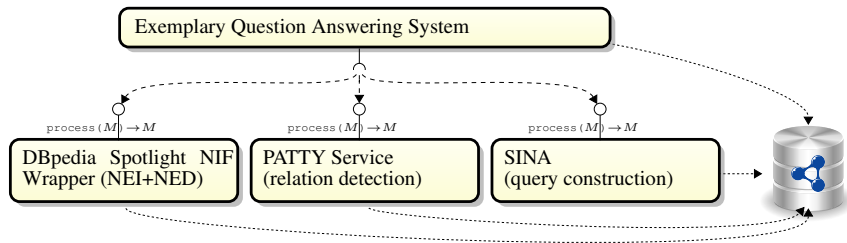
- ... semantic data (RDF)
 - o logical representation of alignment
 - ontology alignment (OWL, DOL)
 - SPARQL query
- ... non-semantic data (API, JSON, XML, CSV, ...)
 - o SPARQL query

- NER/NED
 - o DBpedia Spotlight (NIF)
P. N. Mendes, M. Jakob, A. Garcia-Silva, and Ch. Bizer: "DBpedia Spotlight:shedding light on the web of documents." In I-SEMANTICS, 2011
- relation detection
 - o PATTY
N. Nakashole, G. Weikum, and F. M. Suchanek. PATTY: "A taxonomy of relational patterns with semantic types." In EMNLP-CoNLL, 2012
- query construction
 - o SINA
S. Shekarpour, E. Marx, A.-C.N. Ngomo, and S. Auer. SINA: "Semantic interpretation of user queries for question answering on interlinked data." Web Semantics: Science, Services and Agents on the WWW, 2015

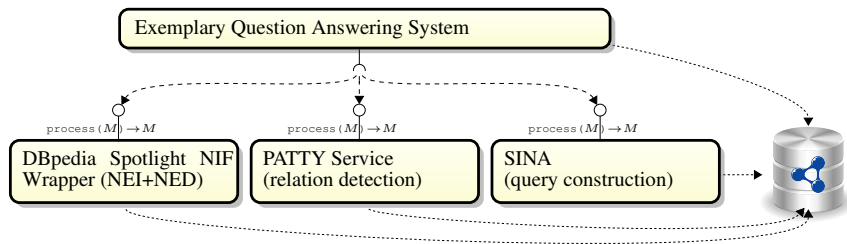
Note:
many options for alignment



CASE STUDY



CASE STUDY

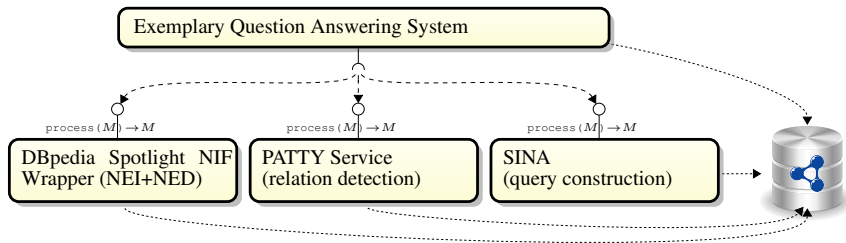


Component

1. DBpedia Spotlight
2. PATTY
3. SINA query execution



CASE STUDY



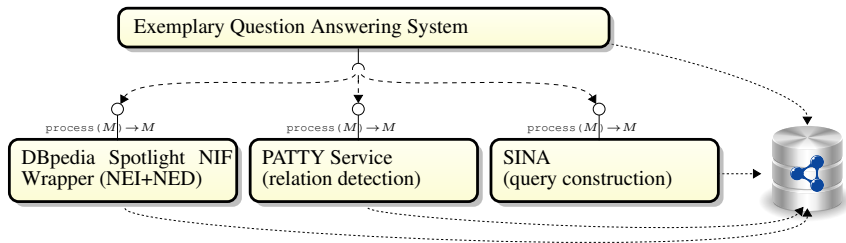
Component

1. DBpedia Spotlight
2. PATTY
3. SINA query execution

Process within components



CASE STUDY



Component

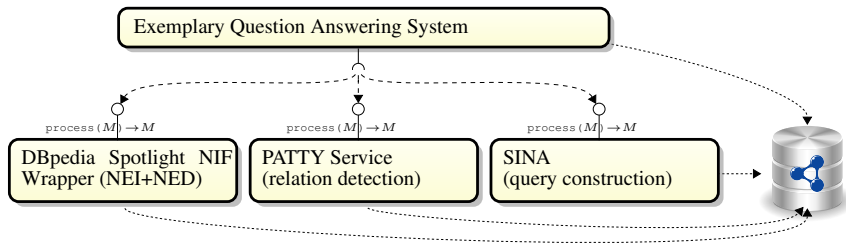
1. DBpedia Spotlight
2. PATTY
3. SINA query execution

Process within components

1. retrieve data from KB



CASE STUDY



Component

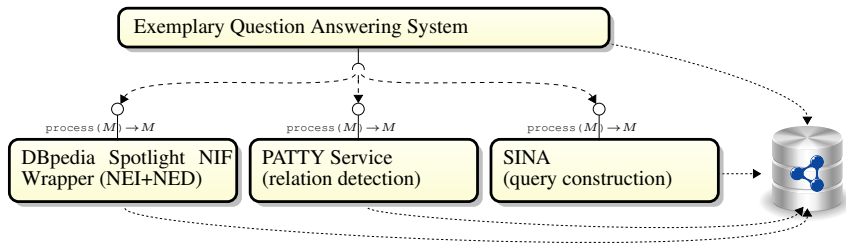
1. DBpedia Spotlight
2. PATTY
3. SINA query execution

Process within components

1. retrieve data from KB
2. process data



CASE STUDY



Component

1. DBpedia Spotlight
2. PATTY
3. SINA query execution

Process within components

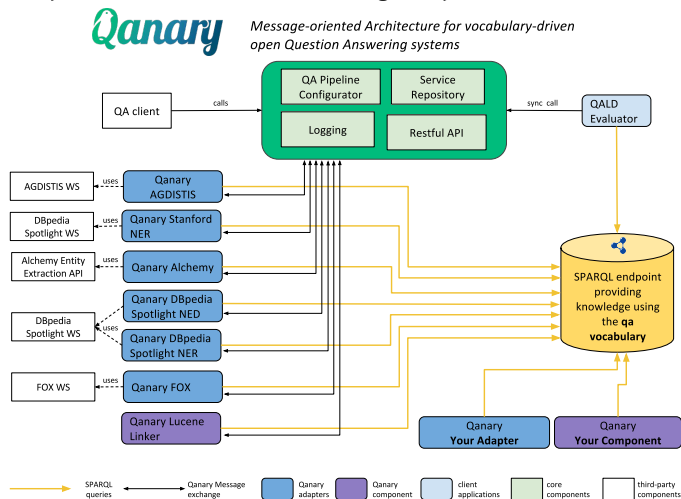
1. retrieve data from KB
2. process data
3. extend KB

→ vocabulary-driven, component-oriented QA system possible



OUTLOOK: ONGOING WORK

- goal: easy-to-use framework for creating QA systems



TAKE AWAY: QANARY METHODOLOGY



TAKE AWAY: QANARY METHODOLOGY

- Qanary: knowledge-driven methodology for QA systems
- build on-top of the qa vocabulary



TAKE AWAY: QANARY METHODOLOGY

- Qanary: knowledge-driven methodology for QA systems
- build on-top of the **qa** vocabulary
- agile approach for creating QA systems
 - interoperable infrastructure, exchangeable components, flexible granularity, isolation of components



TAKE AWAY: QANARY METHODOLOGY

- Qanary: knowledge-driven methodology for QA systems
- build on-top of the **qa** vocabulary
- agile approach for creating QA systems
 - interoperable infrastructure, exchangeable components, flexible granularity, isolation of components
- first step towards an *ecosystem of QA components* enabling best-of-breed approaches for future QA systems



TAKE AWAY: QANARY METHODOLOGY

- Qanary: knowledge-driven methodology for QA systems
- build on-top of the **qa** vocabulary
- agile approach for creating QA systems
 - interoperable infrastructure, exchangeable components, flexible granularity, isolation of components
- first step towards an *ecosystem of QA components* enabling best-of-breed approaches for future QA systems

Join **Qanary** at
Github!

github.com/WDAqua/Qanary

Visit the **Qanary**
demo!

Thursday

Visit WDAqua at the
project networking
session! **Wednesday**



Andreas Both
contact@andreasboth.de

[xing.com/profile/Andreas_Both6](https://www.xing.com/profile/Andreas_Both6)
[linkedin.com/in/andreas-both-9426722](https://www.linkedin.com/in/andreas-both-9426722)

