QUASAR: Querying Annotation, Structure, and Reasoning

Luying Chen, Michael Benedikt
Evgeny Kharlamov
Oxford University
Free University of Bozen-Bolzano

QUASAR Query Language

Naive query over annotated documents:

SELECT ?b
FROM Corp1.Annotation
WHERE ?a.snippet.paraNum = 78
AND ?b.assertion.predicate = "Place"

1) Too many wrong answers:
   a) irrelevant places are found
   b) returned places are not geographic locations

2) Concept reasoning:

   AND ?b.assertion = Place(?y)
   [OntoFilter: PopulatedPlace(?y)]
   AND DisjointWith(?y, "Ocean")

What Places has Magellan Visited?

Paragraph 1: Ferdinand Magellan was a Portuguese explorer. He was born in Sabora, and served King Charles I of Spain in search of a westward route to the "Spice Islands" (modern Maluku Islands in Indonesia).

Paragraph 2: Magellan's expedition of 1519 became the first to sail from the Atlantic Ocean into the Pacific Ocean (via the Strait of Magellan), and the first to cross the Pacific. Magellan was killed during the Battle of Mactan in the Philippines.

Paragraph 3: Yermak was a Cossack who led the Russian conquest of Siberia in 1582.

Issues with blind access of annotated documents:

1a) Doc. navigation:
   1) Too many wrong answers: a) irrelevant places are found
   2) Too few right answers: a) relevant types of places are missing

1b) Querying KBs:
   1) Too many wrong answers: a) irrelevant places are found
   2) Too few right answers: a) relevant types of places are missing

Ontology:

Facts:

Legend:

Concept Annotations: City Country Person Place
Polarity Annotations: Positive Tone
Relation Annotations: same-as born-in

Linking classes and relations using URIs

Linking entities using URIs

Knowledge Base

Annotation Local Store

Annotator

Index Module

Query Planner

Query Processor

Query Executor

Result Composer

Access API

Annotation Wrapping API

Local Annotation Store

OWL Reasoner API

Motivation

Number of systems for automatic text annotation increases:
   - E.g.: OpenCalais, AlchemyAPI, Zemanta, etc.

Number of (Web) ontologies or Knowledge Bases increases:
   - enriching annotations with semantics: City Sub-class-of Place
   - collecting annotated facts: Moscow is a City
   - E.g.: Dbpedia, Yago, UniProt, etc.

Our Goal: to develop a system that allows
   - to do focused querying of annotated documents and
   - exploit Knowledge Bases to increase quality of answers

System Architecture

Future Work

- How to find URIs that disambiguate annotated entities?
- Hilton is labeled as a Person, which Hilton is meant?
- How to unify vocabularies?
- of Knowledge Bases and automatic annotators
- How to filter out wrong annotations?
- annotator says the entity is a City but it is a Person
- How to combine their different annotators?
- they have different vocabularies, opinions