Motivation

- Natural language is more accessible to the average user. It also relieves the user from having to know the underlying schema.
- Knowledge bases contain crisp facts from which complex information needs can be answered, possibly by joining these facts.

Goal

Automatically translate user’s natural language questions to structured queries over knowledge bases

Data

- A knowledge base with instance & ontological data

YAGO = Wikipedia + Wordnet

Triples of entities, classes & relations

- Rome isA city
- city subclassOf location
- Roberto_Rossellini marriedTo Ingrid_Bergman

- Surface forms for entities, classes & relations

  {'Rome', 'eternal city', 'Roma'} → Rome
  {'Casablanca', 'Ad Dar al Bayda'} → Casablanca

  {'play', 'star in', 'act', 'leading role'} → actedIn
  {'play', 'perform'} → hasMusicalRole
  {'married', 'spouse', 'wife'} → marriedTo
  {'film', 'movie'} → movie

Challenges

- Detection and mapping of entities, classes & relations
- Joint disambiguation of entities, classes & relations
- Structured query generation

Construct a Disambiguation Graph

1. Detect candidate phrases
2. Map candidate phrases to (potential) semantic items
3. Detect phrase-level triples (q-units) using dependency parsing
4. Joint disambiguation via a constrained dense subgraph computation using ILP.
5. Translate subgraph to structured query: classes are type constrained variables, joins expressed by common phrases between q-units.

More Information

Demo session 3: Thursday, 14:00-15:30 in Bellecour #1
Contact: Mohamed Yahya <myahya@mpi2.de>