Knowledge Inquiry for Information Foraging

Kyriaki D. Zafeiroudi, and Rebecca J. Passonneau

The Pennsylvania State University

kzafeiroudi@cse.psu.edu

Outline

* Motivation

* InK

* Use Cases

* Experiments

* Conclusion

Motivation

* Processing data to answer task-specific questions can be challenging especially given any level of uncertainty about what information is available.

* Most Question Answering systems accept questions in Natural Language as input and in order to provide a response, they assume the answer exists in the knowledge base.

* But what happens if there is no good answer available?

* A more appropriate tool for the human analyst would support open-ended exploration for relevant information from structured data sources.

Motivation

* We present **InK**, an initial inquiry system for knowledge graphs that focuses on formulating **helpful responses** even when no good answer is available.

* InK is agnostic to the knowledge base and its terminology, and also transparent to the knowledge it has access to. It queries RDF graphs that are of the form of triples, each triple consisting of a subject, a predicate, and an object.

* Knowledge acquisition from RDF graphs is performed through question-guided generic RDF queries that extract and present relevant information; no training or semantic parsing is needed.

* The resulting output consists of knowledge that is assumed to be relevant to a general information need that motivates the question, without a translation of the complete NL question into a query language.

Motivation

❋ As part of the NS CTA research project on *Information Processing Across Networks for Decision-Making*, which is currently in its 10th year, we have come to realize how important for the human analyst the process of multi-genre, multi-modal knowledge construction is for intelligence analysis and foraging.

❋ Our system aims to fulfill the standards and requirements set by this project, of which InK became part during the 10th year.

Q: In which country is Solaize located?

❋ InK identifies *country* and *Solaize*.

❋ Gradually matches the two entity mentions to the KB.

❋ Retrieves relevant information for each entity mention.

❋ Finds connections between the entity mentions.

❋ Returns a response.

Subject	Predicate	Object
http://dbpedia.org/resource/Solaize	http://dbpedia.org/ontology/country	**http://dbpedia.org/resouce/France**

Q: What were the casualties from the Malaysia Airlines flight 17?

* InK identifies *Malaysia Airlines, Malaysia Airlines Flight 17, the casualties from Malaysia Airlines Flight 17, casualties*.

* Gradually matches these entity mentions to the KB.

* Retrieves relevant information for each entity mention.

* Tries to find connections between the entity mentions.

* Returns a response.

Subject	Predicate	Object
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/fatalities	**298**

Q: What were the casualties from the Malaysia Airlines flight 17?

※ InK identifies *Malaysia Airlines, Malaysia Airlines Flight 17, the casualties from Malaysia Airlines Flight 17, casualties.*

※ Gradually matches these entity mentions to the KB.

※ Retrieves relevant information for each entity mention.

※ Tries to find connections between the entity mentions.

※ Returns a response.

Subject	Predicate	Object
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/fatalities	**298**

Q: What were the casualties from the Malaysia Airlines flight 17?

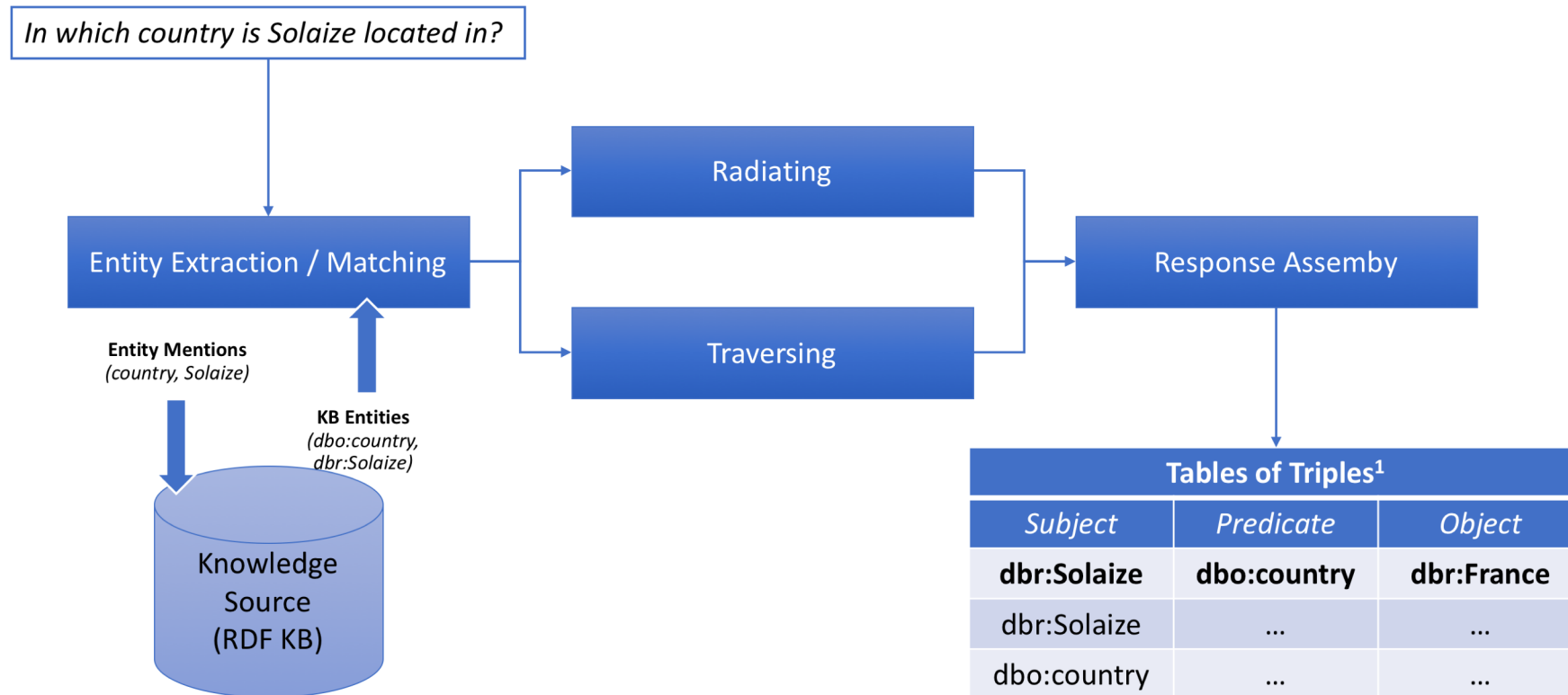⁜ A QA system relying on semantic parsing would generate the following query:

> *SELECT ?O WHERE {*
>
> > *<dbr:Malaysia_Airlines_Flight_17>* *<dbp:casualties>* *?O*
>
> *}*

which would return no results.

⁜ InK overcomes the issue of the arbitrariness of the KB terminology, as well as that of the linguistic structure of a question.

Subject	Predicate	Object
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/fatalities	**298**

InK

In which country is Solaize located in?

Entity Extraction / Matching

Entity Mentions
(country, Solaize)

KB Entities
*(dbo:country,
dbr:Solaize)*

Knowledge
Source
(RDF KB)

Radiating

Traversing

Response Assemby

Tables of Triples[1]

Subject	Predicate	Object
dbr:Solaize	**dbo:country**	**dbr:France**
dbr:Solaize	...	...
dbo:country	...	...

[1]*Abbreviations used:*
dbr *for* http://dbpedia.org/resource/
dbo *for* http://dbpedia.org/ontology/

InK

* First, incremental grounding of concepts to KB entities is performed.
* Then, two query procedures, radiating and traversing, are applied, following Gricean principles to be relevant, clear, and to balance the tradeoff between being informative and being concise.
* Radiating assembles a subgraph of triples where each candidate KB entity occurs as subject or object.
* Traversing searches for short sequences of triples (paths) that connect the candidate KB entities.
* An entity mention can be matched to multiple KB entities; traversing also serves as a procedure to prefer one candidate KB entity over another by comparing which pairs of candidate matches return the most informative and most concise response.

Use Case 1: What were the casualties from Malaysia Airlines Flight 17?

http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/aircraftType	http://dbpedia.org/resource/Boeing_777
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/audio	Intercepted phone calls, verified with voice recognition by the National Security Agency, between rebels discussing which rebel group shot down the aircraft and initial reports that it was a civilian aircraft. Audio released by Security Service of Ukraine with English subtitles.
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/crew	15
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/destination	http://dbpedia.org/resource/Kuala_Lumpur_International_Airport
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	**http://dbpedia.org/property/fatalities**	**298**
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/operator	http://dbpedia.org/resource/Malaysia_Airlines
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/origin	http://dbpedia.org/resource/Amsterdam_Airport_Schiphol
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/passengers	283
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/site	Near Hrabove, Donetsk Oblast, Ukraine
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/summary	http://dbpedia.org/resource/List_of_airliner_shootdown_incidents
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/survivors	0
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/tailNumber	9
http://dbpedia.org/resource/Joep_Lange	http://dbpedia.org/ontology/deathCause	http://dbpedia.org/resource/Malaysia_Airlines_Flight_17
http://dbpedia.org/resource/Liam_Davison	http://dbpedia.org/ontology/deathCause	http://dbpedia.org/resource/Malaysia_Airlines_Flight_17
http://dbpedia.org/resource/Shuba_Jay	http://dbpedia.org/ontology/deathCause	http://dbpedia.org/resource/Malaysia_Airlines_Flight_17
http://dbpedia.org/resource/United_Nations_Security_Council_Resolution_2166	http://dbpedia.org/property/subject	http://dbpedia.org/resource/Malaysia_Airlines_Flight_17

Use Case 2: Who is the leader of the Islamic State of Iran?

* Incorrect assumption:
  * "Islamic Republic of Iran" or
  * "Islamic State of Iraq"?
* InK overcomes this problem:
  * No candidate entity for "Islamic State of Iran".
  * Searches for more specific entity mentions: "Islamic State" and "Iran".
  * Finds "Islamic State of Iraq", "Islamic State of Iraq and the Levant", "Islamic State of Iraq and the Levant in Libya", "Islamic State of Iraq and the Levant – Caucasus Province", "Islamic State of Iraq and the Levant – Khorasan Province", "Islamic State of Iraq and the Levant – Sinai Province" and "Islamic Republic of Iran".
* The user will find out that there is no Islamic State in Iran, and additionally will receive information about all the leaders in every Islamic State, but also Iran, skipping a step that standard QA systems require: letting the user know the question does not have an answer and that the user would have to rephrase or change the question completely.

Experiments on QALD [240 Questions]

❋ InK's performance was evaluated using recall as a metric, computed for each question Q following the formula below:

$$Recall(Q) = \frac{\text{\# of correct answers for Q occuring in InK response}}{\text{\# of gold standard answers for Q}}$$

Item	Recall=1.0	Recall>0.5	Recall>0.0
Radiating	54.58%	66.25%	72.92%
Traversing	33.75%	40.00%	50.42%
InK	62.50%	71.25%	78.75%

Dataset	InK Recall
QALD-1	0.68
QALD-2	0.68
QALD-7	0.75
Overall	0.70

System	Questions	Recall	InK Matched Recall
QALD-1			
PowerAqua	23	**0.79**	0.77
FREyA	23	0.72	**0.77**
QALD-2			
SemSeK	54	0.56	**0.64**
MHE	53	0.49	**0.63**
QAKis	26	0.42	**0.73**

Conclusion

* InK grounds Natural Language questions in knowledge graph concepts through a set of procedures, agnostically to the semantics of the knowledge graph.

* Contentful phrases and words are independently matched to concepts in the knowledge graph, and a succession of queries searches for triples that contain these concepts.

* A set of triples is assembled into a response that aims for relevance, informativeness and concision.

* In the process, potential ambiguity of concept grounding is often resolved.

* An evaluation of the recall on InK responses to factual questions shows that the responses often contain the correct answer.

Thank you for your time!

Questions?