Knowledge Inquiry for Information Foraging

Kyriaki D. Zafeiroudi and Rebecca J. Passonneau
The Pennsylvania State University
kzafeiroudi@cse.psu.edu rjp49@cse.psu.edu

Abstract[i]

Human analysts have a vital role in the task of sensemaking, the process of extracting information to reach conclusions and make decisions. Question-Answering (QA) is an existing natural language processing application that would appear to be relevant to the analyst's task, given information needs to address in a structured knowledge source. Standard QA systems, however, assume an input question can be interpreted in isolation, meaning that there is a single translation of language to a structured query, and that there is a unique correct answer. We assume that a more appropriate tool for an analyst would support open-ended exploration for relevant information from structured data sources, and would not commit too early to a single interpretation of the analyst's question. We provide the capability to pose natural language questions to knowledge graphs in RDF format where information that is relevant to the question can be visualized, making the knowledge source more transparent to the user. This paper presents InK, an inquiry system for knowledge graphs where the input is a NL natural language (NL) question and the output consists of knowledge assumed to be relevant to a general information need that motivates the question.

Introduction

Visualization tools can have a huge impact in the process of sensemaking for analysts (1). Processing data to answer task-specific questions is a difficult task that becomes even harder when there is any level of uncertainty about what information is available. For instance, for the question ***What were the casualties from the Malaysia Airlines Flight 17?*** the right answer lies in DBpedia Knowledge Base (KB) in the form of the RDF triple showing in Table 1.

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

Table 1: The DBpedia RDF triple presenting the information about the flight MH17 and its casualties.

DBpedia is one of many structured knowledge sources on the web encoded in RDF[ii][1], a framework that consists of a graph of triples, where a triple consists of three entities in a subject, predicate, object relation. Note that the DBpedia predicate "fatalities" connects the relevant flight to the number of casualties. DBpedia also includes the entity "http://dbpedia.org/property/casualties", but not connected to the incident concerning the flight MH17, making the process of translating the question into a query and retrieving the

[1] Resource Description Framework (RDF): https://www.w3.org/RDF/

correct answer a non-trivial task. InK is designed to sidestep the translation dilemma by posting separate queries entities mentioned in the NL question, then assembling relevant connections when matches from natural language vocabulary to DBpedia entities can be found. The triple shown here is one of a group of triples relevant to the “Malaysia Airlines Flight 17”, matches the DBpedia entity “http://dbpedia.org/resource/Malaysia_Airlines_Flight_17”. In the example presented, an analyst would recognize that “casualties” is not the right word and would likely continue the knowledge inquiry process using the term “fatalities”.

Consider another question: ***In which country is Solaize located?*** InK extracts the entity mentions “Solaize” and “country”, identifies candidate interpretations for both entities in the KB, and probes into the KB to return a subgraph with relevant information for the candidate KB entities. The returned subgraph includes the information presented in Table 2.

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

Table 2: The DBpedia RDF triple presenting the information about the location of Solaize.

Note that InK’s response includes the answer to the question, which is France, without a translation of the complete NL question into a query language (e.g., SPARQL, which queries RDF graphs), as is usually the case for semantic parsing (2). We address the problem of knowledge acquisition from RDF graphs that is agnostic to the terminology of the knowledge source through question-guided, generic RDF queries that extract and present relevant information. We are developing a new technology we refer to as knowledge inquiry, InK, which produces responses to NL questions from RDF KBs by taking into account cooperative principles of communication (3). We assume that a literal answer is not always the most cooperative – this is particularly relevant for analysts and others who are engaged in information gathering/foraging. We leave the grounding of each term to the user by providing possible interpretations, and thus we present a system that is not relying on the translation of terms to a particular KB and can be applied to different domains without any manual work in learning the terminology.
The next section presents InK and its architecture, particularly the processes it encompasses, followed by two use cases pointing out the informativeness and relevance of the overall procedure. We then present experiments run on questions sampled from different QA datasets, in order to evaluate InK responses against known answers.

System Overview

InK first extracts entities mentioned in the original question, primarily nominal phrases, and in cases where this is insufficient, also predicating words like verbs and adjectives. The next step is to match the natural language expression (e.g., “Malaysia Airlines Flight 17”) to a knowledge graph entity (e.g., “http://dbpedia.org/resource/Malaysia_Airlines_Flight_17”). In many knowledge graphs, such as DBpedia, it is possible to match a natural language expression to the label attribute of one or more knowledge graph entities. Often, there are multiple matches (ways to ground the language in the knowledge source), meaning the language expression is ambiguous relative to the knowledge source. Through query procedures we refer to as radiating and traversing, InK assembles knowledge about the matched entities consisting of RDF triples. These procedures follow Gricean principles (3) to be relevant, clear, and to balance the tradeoff between being informative and being concise. Instead of trying to reason about the intended meaning behind a question, InK leaves the reasoning to the user (4).

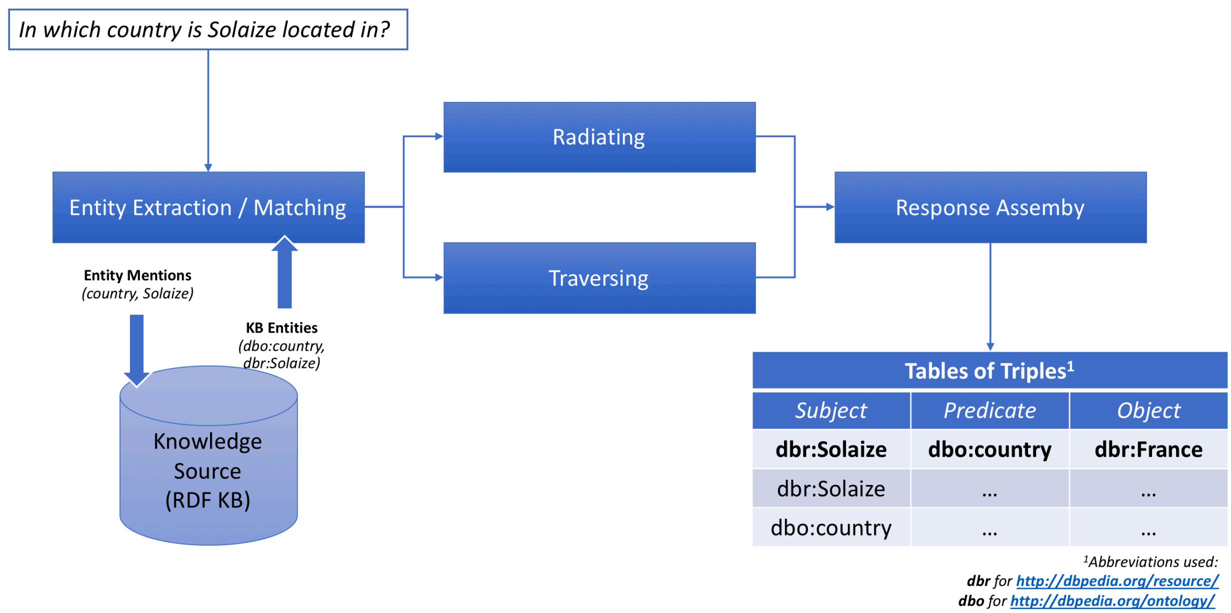

Figure 1: The InK pipeline.

Figure 1 shows the pipeline of our system. The user poses a NL question to the system, which is then parsed, and entity mentions are extracted. We focus on entity mentions of the type: named entity, noun phrase, proper noun, noun-noun compound and noun. Each extracted entity mention is then matched to one or multiple KB entities, in a procedure which as of now relies mainly on string matching the NL word or phrase to candidate KB entities, using the data property “http://www.w3.org/1999/02/22-rdf-syntax-ns#label” that matches them with a human readable label. Given the candidate matches, the radiating and traversing procedures consist of SPARQL queries to gather the most relevant, succinct information that is connected to the matched entities.

Radiating assembles a subgraph of triples that include each candidate KB entity in the subject or object position, indicating what relevant information is available concerning the user’s inquiry. Traversing searches for sequences of triples (paths) that connect the candidate KB entities, and ranks the paths. Short paths in the form of adjacent triples in the KB are extracted, where one candidate entity is part of the first of the sequence of triples forming the path, and another shows up in the last triple of the sequence. The paths are ranked by the informativeness of the predicates in each triple, taking into account that shortest paths are more likely to give out the most pertinent information, without breaking the maxim of manner. As informativeness, we calculate the negative loglikelihood of the occurrence of the predicate in the KB (5). Frequently, there is more than one path that connects two entity mentions; since one 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 response. For instance, country is matched with the following KB entities in DBpedia: “http://dbpedia.org/ontology/Country”, “http://dbpedia.org/ontology/collectionSize”, “http://dbpedia.org/property/country”, “http://dbpedia.org/ontology/country”, but only the last entity occurs in a path of the shortest length. Figure 2 shows example questions from the QALD datasets (6) and the corresponding traversing results.

QALD-7: *In which country is Solaize located?*	
Node	Path Length = 1
S	`http://dbpedia.org/resource/Solaize`
P	`http://dbpedia.org/ontology/country`
O	`http://dbpedia.org/resource/France`
QALD-7: *Which languages are spoken in Romania?*	
Node	Path Length = 2
S1	`http://dbpedia.org/resource/Romania`
P1	`http://dbpedia.org/property/languages`
O1, S2	`http://dbpedia.org/resource/Armenian\_language`
P2	`http://www.w3.org/1999/02/22-rdf-syntax-ns\#type`
O2	`http://dbpedia.org/ontology/Language`
S1	`http://dbpedia.org/resource/Romania`
P1	`http://dbpedia.org/property/languages`
O1, S2	`http://dbpedia.org/resource/Yiddish`
P2	`http://www.w3.org/1999/02/22-rdf-syntax-ns\#type`
O2	`http://dbpedia.org/ontology/Language`

Figure 2: Example questions and InK traversing results.

Use Cases

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

In this case the extracted entity mentions are: *Malaysia Airlines, Malaysia Airlines Flight 17, the casualties from Malaysia Airlines Flight 17, casualties*. We are using DBpedia (7) as the KB against which we run our experiments. In DBpedia, there is an entity that matches “casualties” but it does not connect to any of the candidate entities for “Malaysia Airlines Flight 17”. The knowledge engineering decisions for each knowledge base are somewhat arbitrary, and it is unrealistic to try and predict them. As a consequence, our approach is based on grounding terms extracted from the question independently, then assembling knowledge about them, rather than trying to translate the question into a single query specific to a given knowledge source. In this case, traversing returns no results and the user has the opportunity to look into DBpedia entities concerning the “Malaysia Airlines Flight 17” and “casualties”. Table 3 presents the radiating response for the DBpedia entity matching "Malaysia Airlines Flight 17", showing 20 triples in which the matched entity is either the subject or object of the triple.

http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/lat1Dir	N
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/lat1Min	18
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/lon1Dir	E
http://dbpedia.org/resource/Malaysia_Airlines_Flight_17	http://dbpedia.org/property/lon1Min	45
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

Table 3: Radiating results for "Malaysia Airlines Flight 17".

We can see that in the InK's radiating results, the number of casualties appears in a triple where the predicate is "http://dbpedia.org/property/fatalities". Because a list of 20 or more triples, as in Table 3, is insufficiently succinct, our future work will develop methods to aggregate triples that share common information (e.g. the three entities connected via the relation "deathCause" to "Malaysia Airlines Flight 17") into their common characteristic (here: belonging to the class "Person"), making the information presented in the response easier to digest.

A strength of our approach is that InK does not need to learn how to translate "casualties" to "http://dbpedia.org/property/fatalities", which is the usual direction taken in semantic parsing. Instead, InK queries the knowledge source about each entity independently, then searches for connections to expose relevant knowledge that can easily be found in the knowledge source. An InK summary (table of triples) serves multiple purposes; it is very common for users to not have a clear view of what they would like to inquire information about. By summarizing what is available to the user we can help build on his/her knowledge so another more specific question can then be formed. What is more, InK provides the user with more knowledge than initially requested, in the example question some related terms for "flight MH17" are "eastern Ukraine", "Buk missile", "downed flight".

Question 2: Who is the leader of the Islamic State of Iran?
For this question, the user makes an incorrect assumption, which can be either of the following; the user is interested in finding information about the leader of Iran, which has not an Islamic State, but its long name is "Islamic Republic of Iran"; or the user searches information about the leader of the Islamic State of Iraq. InK overcomes this problem since it cannot find a candidate entity in the KB for "Islamic State of Iran", so it changes the direction of the search to the most specific entity mentions, "Islamic State" and "Iran". In the first case, the results point to "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" and "Islamic State of Iraq and the Levant – Sinai Province". By shifting the focus of the search, InK returns a response which is both informative and relevant; traversing results include all the aforementioned Islamic States connected through sequences of triples to current and previous leaders. For the second entity, InK returns information about Iran, including triples that connect the KB entity to "Islamic Republic of Iran" through the relation "http://dbpedia.org/ontology/longName", pointing the user to the erroneous assumption made. The name of the leader is also included in the traversing results.
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

Computing the relevance of an answer is hard to achieve automatically. In order to evaluate whether InK's responses are relevant to a question, we test whether the response includes in the aggregated triples any URIs that match a known answer. Our experiments were against available QA datasets, where the answers are included, and the KB used is available for us to use and validate our system on. For this purpose, we chose the Question Answering over Linked Data (QALD) (6) datasets. QALD is an open challenge with a 7-year history. Data provided to researchers that took part in the challenge for the years 2011, 2012 and 2017 appeared to be quite useful for testing InK against DBpedia 2016-04. We accumulated 240 questions from these datasets, which fulfill the requirements of having gold standard answers in the DBpedia version that we employed and also the type of question would require an answer in the form of DBpedia resource. InK's performance was evaluated using recall as a metric, computed for each question Q following the formula below:

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

Table 4 shows the percentage of questions that have perfect recall, recall more than 0.5 and those that showed non-zero recall, for the cases of radiating, traversing, as well as the final response from InK (taking into account both procedures), while Table 5 shows InK's recall calculated for all the 240 questions.

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%

Table 4: Percentage of the 240 questions with Recall = 1.0, Recall > 0.5 and Recall > 0.0 for Radiating, Traversing and InK.

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

Table 5: Recall for InK over all the QA datasets used.

Conclusion

We presented an inquiry system that parses natural language questions, grounding them in knowledge graph concepts through a procedure that is agnostic to the semantics of the knowledge graph. Questions are not translated into a query language, but key concepts are extracted and matched to the knowledge graph, and two processes accumulate important triples that these concepts take part in. The final response is formulated by focusing on relevance and informativeness. An evaluation of the recall of the system's responses showcases that InK usually returns information containing the answer or a significant part of the answer. Future work will incorporate the presented procedures in a more user-friendly graphical interface, making the visualization of the subgraph returned easier to evaluate and draw conclusions. Information theoretic metrics will be introduced to rank the most valuable information higher. Expendability in other languages will be identified; since the system runs agnostically on the knowledge source specified by the human analyst, making it available in multiple languages seems within reach, making information shareable amongst coalition forces including soldiers from multiple nations.

References

1. *The Cost Structure of Sensemaking.* **Daniel M. Russel, Mark J. Stefik, Peter Pirolli, Stuart K. Card.** Amsterdam, The Netherlands : s.n., 1993. CHI '93 Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. pp. 269-276.

2. *Semantic Parsing on Freebase from Question-Answer Pairs.* **Percy Liang, Jonathan Berant, Andrew Chou, Roy Frostig.** 2013. Proceedings of EMNLP. pp. 1533-1544.

3. **Grice, Herbert Paul.** Logic and Conversation. [book auth.] Peter and Morgan, Jerry L. Cole. *Syntax and Semantics, Vol. 3, Speech Acts.* New York : Academic Press, 1975.

4. *Testing a Knowledge Inquiry System on Question Answering Tasks.* **Kyriaki D. Zafeiroudi, Leah Eckman, Rebecca J. Passonneau.** Monterey, California, United States of America : s.n., 2018. Proceedings of the Joint proceedings of the 4th Workshop on Semantic Deep Learning (SemDeep-4) and NLIWoD4: Natural Language Interfaces for the Web of Data (NLIWOD-4) and 9th Question Answering over Linked Data challenge (QALD-9) 2018 (SemDeep-4 and NLIWO. pp. 17--32. 1613-0073.

5. *Using information content to evaluate semantic similarity in a taxonomy. In: - Volume 1. pp. 448{453. IJCAI'95, ,.* **Resnik, Philip.** San Francisco, CA : Morgan Kaufmann Publishers Inc., 1995. Proceedings of the 14th International Joint Conference on Articial Intelligence. Vol. 1, pp. 448-453.

6. *7th Open Challenge on Question Answering over Linked Data (QALD-7).* **Usbeck, Ricardo and Ngomo, Axel-Cyrille Ngonga and Haarmann, Bastian and Krithara, Anastasia and Röder, Michael and Napolitano, Giulio.** s.l. : Springer International Publishing, 2017. pp. 59-69.

7. *A Comparative Survey of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO.* **Rettinger, Michael Färber and Basil Ell and Carsten Menne and Achim.** 2015, Semantic Web - Interoperability, Usability, Applicability (Semantic Web Journal), Vol. 1, pp. 1-26.

---

[i] This work was supported in part by the Army Research Laboratory and accomplished under Cooperative Agreement Number W911NF-09-2-0053.

[ii] Resource Description Framework (RDF): https://www.w3.org/RDF/