Pragmatics Aware Querying in Heterogeneous Knowledge Graphs

Amar Viswanathan
Tetherless World Constellation
Rensselaer Polytechnic Institute
110 Eighth Street, Troy, NY USA 12180

Abstract
With the increase in structured Linked Data sources, Knowledge Graphs (KG) are becoming indispensable in areas like Question Answering, Web Search and Data Analytics. Given the richness present in the KG schema definitions, more expressive user querying is possible. However, accessing KGs is still challenging, because the rich schema is not exposed in a more intuitive way to the naive user. This results in user intent not being translated into a precise query satisfying all the user needs. My thesis focuses on making KGs accessible for querying by using Gricean notions of Cooperative Answering. More specifically, using Query Reformulations, Data Awareness, and Pragmatic Context, KGs can be made more responsive to user requirements and provide quality results in context.

Motivation
Knowledge Graphs KG today, are changing the way information is stored, accessed and utilized. Given their rich schema definitions, user querying is becoming quite expressive. This is also changing the search landscape. For example, user queries like Movies directed by Martin Scorcese featuring Jose Pesci and Robert De Niro or Michelin star rated restaurants in New York City are easily answered by resources like DBpedia (Auer et al. 2007). Thus, KGs are able to handle more expressive precise querying. This makes it more ideal for a variety of search tasks. However, while the research world is focused on adding increased expressiveness, an often overlooked part is the complexity of data access. With an increase in schema definitions, the naive user finds it harder to find information suitable to his needs in the KG space. Concepts can interact in a myriad of different ways. This means that even if the user has a precise need, translating that to a specific query becomes difficult with the increase in concept size and concept interaction space. Also given the long association with databases, querying the data from a table point of view is a strong impediment to utilize the richness of the Knowledge Graph. We contribute the following as potential reasons:

- **Ever Expanding Schema**: Heterogeneous Knowledge Graphs have all the required information, but their ever expanding schema, which is not consolidated, leads to mis-understanding of schema semantics and inhibits precise querying (Dolog et al. 2009).
- **Incomplete and Inconsistent data**: Even if the schema is complete sometimes the data is not always complete. In addition given the reliance on automatic extraction techniques a lot of the extracted facts can be redundant, irrelevant or errors (Pujara et al. 2013). This leads to inaccurate results.
- **Complex Querying**: Querying data with SPARQL requires the user to build the exact triple patterns for his requirement. However, such requirements aren’t always mapped exactly, and there is a huge gap between the user intent and what has been formulated as a query. Users are very often able to provide only broad queries, whereas to build a precise query they need more schema familiarity. (Dolog et al. 2009) With the increase in schema size, the complexity of query patterns increases as well. This induces more imprecise querying even when the context of the queries are somewhat clear in the mind of the user.
- **Unadaptable to User needs**: Current interfaces to the Knowledge Graph are also very unadaptable to the user and make no effort in aiding the user. They fall short when query correctness has to be assessed and also fail to provide better means to adapt the query to changing schemas. In addition current interfaces also don’t provide mechanisms to differentiate between various interpretations of the same query. This usability is one of the major impediments towards adoption of Knowledge Graph in various domains.

In my thesis, I would want to address the challenge of finding Relevant Information in a KG given an imprecise query. While understanding the user’s precise intent is an open problem, we take the Gricean approach of co-operative answering (Grice 1970), where our system adapts to the user by taking in his query and providing alternate possible interpretations-specific and or -generic, along with the hypothesis. This gives the user a set of queries, in addition to his original query; and, these query suggestions augment the user’s understanding of the system. Thus, aiding the formulation of a more specific query suited to his needs.
Targeted Research Contributions

We propose a novel data-aware Pragmatic Query Reformulation system, which provides a user with a set of reformulated queries that take into account data availability and query context. However, as the schema and concept space gets richer and denser, existing techniques require careful considerations in terms of scaling. In our system, we consider the availability of data as a heuristic parameter in determining the reformulation. We look at a query as a combination of concepts, some of which may be precise and some of which may be imprecise. In brief, these would be the contributions:

- Better Data Awareness
- Pragmatics Aware Query Reformulation
- Providing a faceted discourse oriented user interface for better hypothesis understanding.

Background and State-of-the-art

The foundations of my approach lie in the application of Grice’s principles (Grice 1970). Grice proposed that, talk exchanges do not normally consist of a succession of disconnected remarks, but they are, to some degree, cooperative efforts and the participants recognize in them, a common set of goals. To be able to achieve this in the current context a system should be able to talk to the user and provide contextually relevant information or additional similar relevant queries. So we rely on Query Relaxations and Query Reformulations, which are a part of Cooperative Answering. Generally Reformulations for RDF graphs are focused on Relaxations or Generalizations (Hurtado, Poulovassilis, and Wood 2008) aimed at pushing more relevant content to the users. Such relaxations are either deductive relaxations or use RDFS semantics i.e. type hierarchy or property hierarchy (Poulovassilis and Wood 2010) to relax triple patterns to generate more data. While such systems work on the concept and property level, they do not consider the implications of data availability and user query context. In my proposal thus, the focus is on approximate data awareness, which is used to reformulate queries and thereby produce only reformulations that are pragmatic in the context of the query.

Current Progress

Our first approach was to devise a novel Pragmatic and Data Aware Query Reformulation Algorithm. This is a work in progress and we are targeting the 25th International World Wide Web Conference 1. We summarize the results of this reformulation with an example query $q_1 \in \{\text{Find all nations who are involved in attacks}\}$, which looks like:

$$ q_1 \{\text{entity event role}\} := \text{entity role event} \\quad \text{entity rdf:type \text{individual}} \\quad \text{event rdf:type \text{attack}} $$

Table 1 shows the results of the reformulation for query $q_1$ using techniques developed by our Algorithm. This example is queried on a sample Knowledge Graph KG that is extracted from 75,000 documents, which are in the ACE ‘05 2 schema. In $q_1$, the given query is matched against \{\text{ENTITY, ROLE, EVENT}\} from the schema. The Schema built from the documents has a total of 132 classes and 233 Logical axioms, along with 37 Object Properties and 10 Data Properties. In addition to these results we have built a discourse enabled faceted user interface that eases the interaction with the KG.

Future Directions

In the immediate future I would like to investigate how Query Reformulation fits into the larger scheme of Relevance and Pragmatic Context. Also how does a diverse schema with more expressiveness change the reformulation space. How Query Intent and Relevance augments KG is another open question, which I would like to partially address. In addition utilizing the faceted interface to provide better discourse between the user and the KG would be key to providing adaptive KGs. From there, I look at my thesis as moving towards the field of Computational Pragmatics on Heterogeneous Knowledge Graphs.

References


---

1 http://www2016.ca/

2 http://www.itl.nist.gov/iad/mig//tests/ace/ace05/doc/