A Natural Language Interface for Querying General and Individual Knowledge

Yael Amsterdamer, Anna Kukliansky, Tova Milo
Many real-life scenarios require the joint analysis of **general knowledge**, which includes facts about the world, with **individual knowledge**, which relates to the opinions or habits of individuals.
A setting with mixed individual and general knowledge raises unique challenges: In particular, to support the distinct query constructs associated with these two types of knowledge, the NL question must be partitioned and translated using different means; yet eventually all the translated parts should be seamlessly combined to a well-formed query.
EXAMPLE1:

Consider the following scenario, based on a real question in a travel-related forum: a group of travelers, who reserved a room in Forest Hotel, Buffalo, NY, wishes to know:
EXAMPLE1:

“What are the most interesting places near Forest Hotel, Buffalo, we should visit in the fall?”
Any suggestions on fun, non-wine or golf related things to do in Carmel?
In recent work, we introduced crowd mining as a novel approach for answering user questions about a mix of individual and general knowledge:

In particular, we have implemented the **OASSIS platform** which supports a declarative query language, OASSIS-QL, enabling users to specify their information needs. Queries are then evaluated using both standard knowledge bases (ontologies) and the crowd of web users.
We therefore develop a principled approach for the translation of NL questions that mix general and individual knowledge, into formal queries.
Challenges:

1. The general and individual parts of an NL question may be “mixed”, and therefore must be distinguished from each other to allow separate evaluation.
2. The parts of the NL question that deal with individual knowledge may need to be translated into formal query constructs without putting them first to a structured knowledge base (as they may relate to the not yet-collected knowledge of the crowd).
Challenges:

2. The parts of the NL question that deal with individual knowledge may need to be translated into formal query constructs without putting them first to a structured knowledge base (as they may relate to the not-yet-collected knowledge of the crowd).
3. The translated general and individual query parts need to be seamlessly integrated into a well-formed query. This must be done carefully, since the two types of query parts are translated by separate means.
Suggested solution:

- The modular design of a translation framework
- Development of new modules
Short Overview of the process

“What are the most interesting places near Forest Hotel, Buffalo, we should visit in the fall?”
Output of N-L parsing tool
The Implementation of IX DETECTOR (a naïve approach…)

- **Opinion mining** (also called sentiment analysis), namely, the identification of opinions and subjective statements in NL text.

**Thus,** one may attempt to reuse such opinion-detection tools as-is for IX detection. However, these tools do not account for identifying information needs involving the habits of people.
The Implementation of IX DETECTOR
(an another naïve approach…)

- **Ontology matching**
  declaring every expression that cannot be matched to the ontology as individual. However:

  - **False negatives**: general-purpose ontologies may contain (parts of) IXs in their concept taxonomy (e.g., emotion-related concepts), leading to their mistaken identification as general.
  - **False positives**: due to an incomplete ontology, an expression that cannot be matched is wrongly identified as an IX.
The Implementation of IX DETECTOR (Finally…)

**Patternbased IX Detection:**

In the NL processing literature, there are two main approaches for detecting certain types of knowledge:

- Using training sentences and employing machine learning techniques
- Using predefined structural patterns and vocabularies.
The Implementation of IX DETECTOR
(Finally…)

- **Patternbased IX Detection:**

  The approach we take here is pattern-based, where the patterns partly use vocabularies of opinion mining, which can be constructed by machine learning techniques.
The Implementation of IX DETECTOR
(Finally…)

**Lexical individuality:**

A term (or phrase) in the IX has a lexical meaning which conveys individuality
The Implementation of IX DETECTOR
(Finally…)

**Participant individuality:**

The IX refers to an event that includes a participant which indicates individuality.
The Implementation of IX DETECTOR
(Finally…)

**Syntactic individuality:**

A syntactic structure may indicate individuality.
The *NL$_2$CM* System

(Natural Language interface to Crowd Mining).

Users feed questions into *NL$_2$CM* in a real usage scenario, and analyzed the resulting queries.

Using *NL$_2$CM* to translate questions from Yahoo! Answers, the question-and-answer platform.
The \( NL_2CM \) System
The $NL_2CM$ System – Resolving Missing Query Parameters

some parameters necessary for the translation may not be explicitly specified by the NL question.

when possible, a better solution is asking the users targeted questions about the missing parameters in their queries.
The $NL_2CM$ System – Resolving Missing Query Parameters

- Completing missing LIMIT
- Completing missing THRESHOLD values
- Deciding which variables to project out.
CONCLUSION

- We have studied the problem of translating NL questions that involve general and individual knowledge into formal queries, thereby making crowd mining platforms accessible to the public.

- We have described the novel modules of a translation framework, which detect, decompose, translate and recompose the individual query parts with the general ones.
Future

- Replace or wrap Stanford Parser to allow interactive parsing refinement, and assess the effect on the output query quality versus the user effort incurred.

- IX Detector, future research may study machine learning techniques for automatically suggesting detection patterns.

- Analyzing NL answers collected from the crowd, forums or social platforms.