Usability of Natural Language Interfaces for querying ontologies

Danica Damljanovic, Milan Agatonovic, Hamish Cunningham
danica@dcs.shef.ac.uk

The University Of Sheffield.

sheffield nlp dfu

Introduction

Natural Language Interfaces (NLIs) to structured data allow users to interact with a system using written or spoken language (e.g., English) to perform tasks that usually require knowledge of a formal query language.

Due to natural language complexity and ambiguity, such interfaces usually support a Controlled Natural Language (CNL): a subset of a natural language that includes certain vocabulary and grammar rules that have to be followed.

On one hand, a CNL provides a simple way to retrieve data without extensive training for the end-user, whilst on the other has less expressiveness than the formal languages typically used for accessing structured data.

Challenging habitability problem

A big challenge when designing a CNL is to consider *habitability*. Habitability indicates how easily, naturally and effectively users can use language to express themselves within the constraints imposed by the system. If users can express everything they need for their tasks, using the constrained system language, then such language is considered habitable [3]. In other words, habitable languages are languages that people can use fluently [1].

As shown in Figure 1, to design habitable CNL system, the system vocabulary has to be aligned to that of the user. This adaptation is possible through application of various usability methods (red circle). Once the user is familiarised with the system vocabulary, the opposite adaptation needs to take place, as the user vocabulary needs to be aligned with that of the system (yellow circle).

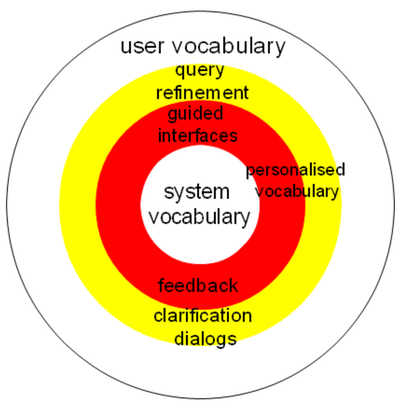

Figure 1: Can the user and the system speak the same language?

Our Approach

By combining various methods such as feedback, query refinement and vocabulary extended from various sources such as WordNet, Watson, LinkedData and user-centric vocabulary, we aim to improve usability of text-based interfaces for querying ontologies and thus:

- eliminate training,
- help the user to learn system-understandable language,
- help the user to express his need more efficiently.

Text-based interfaces for querying ontologies

Text-based interfaces for querying ontologies accept natural language queries as input, generate formal queries behind the scene, execute them against *a knowledge base*, and present results to the user. Minority of them support queries which could be a set of keywords, rather than natural language queries.

Dataset	System	Precision	Recall	Portability
Mooney: geography	PANTO	88.05%	85.86%	0 customisation
	Querix	86.08%	87.11%	0 customisation
	NLP-Reduce	70.7%	76.4%	0 customisation
Mooney: restaurants	PANTO	90.87%	96.64%	0 customisation
	NLP-Reduce	67.7%	69.6%	0 customisation
Mooney: jobs	PANTO	86.12%	89.17%	0 customisation
software engineering ontology	QuestIO	82.14%	71.87%	0 customisation
	AquaLog	86.36%	59.37%	0 customisation
Geographical facts about Germany	ORAKEL	80.60-84.23%	45.15%-53.7%	customised
library data	E-librarian	97%	-	-
biology	CPL	38%	-	-
chemistry	CPL	37.5%	-	-
physics	CPL	19%	-	-

Table 1: Text-based interfaces for querying ontologies

Evaluation conditions:

- *ontology* for querying is of various size and complexity
- *questions* of various complexity i.e. supported CNL varies; higher precision and recall values for:
  - in lab evaluations
  - systems which support simpler forms (e.g. one sentence as a question)
- *evaluation measures* differ e.g. different definitions of precision and recall

Error rate caused by:

- users not familiar with the domain
- knowledge is not in the ontology/knowledge base but the system is not capable of guiding the user to rephrase the question
- feedback messages not helpful i.e. the user can not figure out how to proceed further
- users have assumptions/misconceptsions about the system capabilities and the supported language

FREyA

FREyA (Feedback, Refinement, Extended VocabularY, Aggregation) combines methods such as feedback, query refinement and extended vocabulary, to enable system-user interaction, in order to assist the user formulate the query and express his need more precisely, within the boundaries of system capabilities. FREyA is an optional service which can be used by any CNL system (e.g. ACE [2]) in case when the CNL system is not able to provide the answer immediately.

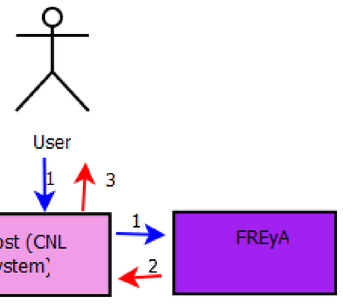

Figure 2: FREyA: logical diagram

Examples

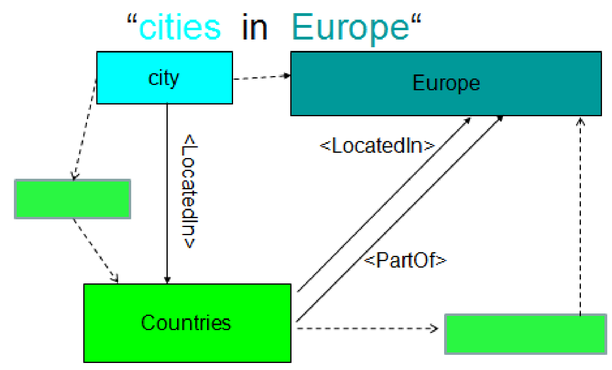

Figure 3: Ontology structure: a big challenge to finding a correct answer

The user's query (1): *which cities are located in Europe?*
OR *cities in europe*
FREyA (2):
[main subject] Cities – locatedIn – Countries – locatedIn – Europe, 0.85
[main subject] Cities – locatedIn – Countries – partOf – Europe, 0.9
CNL (3): *which cities are located in countries that are part of Europe?*

The user's query (1): *What is the date of birth of Johnny Depp?*
FREyA (2): *[main subject] date of birth - the actor – hasName – Johnny Depp?, 1.0*
CNL (3) When was the actor Johnny Depp born?

The user's query (1): *When was the actor who played the main role in Whats eating Gilbert Grape born?*
FREyA (2):
[main subject] date of birth – the actor – hasMainRole - What's eating Gilbert Grape, 1.0
[main subject] date of birth – the actor – Johnny Depp, 1.0
CNL (3): When was the actor Johnny Depp born?

FREyA: detailed view

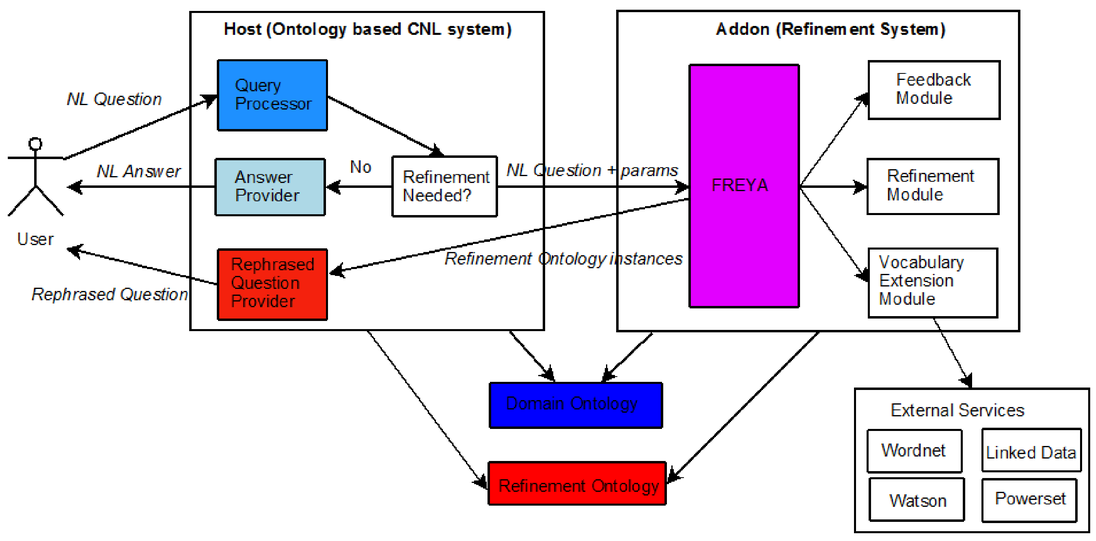

Figure 4: FREyA: component diagram

Conclusion

Usability of NLIs can be improved by implementing the user-system interaction based on feedback, ontology-based query refinement and extending vocabulary from various sources such as WordNet, LinkedData, and others. This interaction will:

- eliminate training,
- help the user to easily familiarise himself with the system capabilities, and
- express his need more precisely, in a way which is understandable by the system.

References

[1] Samuel S. Epstein. Transportable natural language processing through simplicity—the pre system. *ACM Trans. Inf. Syst.*, 3(2):107–120, 1985.

[2] Norbert E. Fuchs, Kaarel Kaljurand, and Tobias Kuhn. Attempto Controlled English for Knowledge Representation. In Cristina Baroglio, Piero A. Bonatti, Jan Małuszyński, Massimo Marchiori, Axel Polleres, and Sebastian Schaffert, editors, *Reasoning Web, Fourth International Summer School 2008*, number 5224 in Lecture Notes in Computer Science, pages 104–124. Springer, 2008.

[3] William Ogden and Philip Bernick. Using Natural Language Interfaces. In M. Helander, editor, *Handbook of Human-Computer Interaction*. Elsevier Science Publishers B.V. (North-Holland), 1996.