Computing Solutions to Multifacited Natural Language Questions over Semantic Knowledge Bases

Ernest Kirstein

January 4, 2015

Query languages like SPARQL help expert users answer complex questions about semantic data knowledge bases. Unfortunately, query languages can only be used by expert users who require extensive training. A natural language interface could help non-expert users access the semantic web without expert training. The end goal would be to allow non-expert users to achieve the same level of sophistication as an expert query language user.

Work by Kaufmann and Bernstein [7] indicates that users have a clear preference for even limited natural language interfaces when compared to keyword or query language interfaces. Other recent works [3, 5, 6] have shown that simple natural language questions can be translated based on known sentence structure using (among other things) NER, named-entity recognition.

These systems are limited - they can only handle simple, direct questions. A recent publication by Sharef et al. [9] outlines obstacles in developing full natural language interfaces for the semantic web. That paper notes a particular difficulty with parsing multifaceted questions - questions with multiple variable, constraints, or operations. This is the gap which my thesis will attempt to bridge.

The foundations of both compiler design and natural language processing have significant overlap [2,8]. However, in practice, there has not been much synergy between the two disciplines [1, 4, 8]. I believe cooperation between these astranged fields is necessary to move forward in either.

My approach will be to build a "natural language compiler" of sorts. That is, to approach the problem of converting natural language questions into SPARQL queries just as one might convert Java source code into byte code. The problems are, or course, an order of magnitude appart in com-

plexity. So I will only aim to parse a limited range of natural language questions. But in that range of questions, I hope to show that multifacited questions can be handeled on a limited basis.

References

[1] Alphred V. Aho, Ravi Sethi, and Jeffrey D. Ullman. *Compilers: Principles, Techniques, and Tools*. Addison-Wesley, Pearson Education, Inc., 1986.

[2] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124, 1956.

[3] A. Damljanovic, M. Agatonovic, and H. Cunningham. Freya: An interactive way of querying linked data using natural language. *FREyA: An Interactive Way of Querying Linked Data Using Natural Language*, 7117:125–138, 2011.

[4] Alice E. Fischer and Frances S. Grodzinsky. *The Anatomy of Programming Languages*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993.

[5] B. Galitsky. Natural language question answering system. *Adelaide, Australia: Advanced Knowledge International*, 2013.

[6] M. Gao, J. Liu, N. Zhong, F. Chen, and C. Liu. Semantic mapping from natural language questions to owl queries. *Computational Intelligence*, 27(2):280–314, 2011.

[7] E. Kaufmann and A. Bernstein. Evaluating the usability of natural language query languages and interfaces to semantic web knowledge bases. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4):377–393, 2010.

[8] Stefano Crespi Reghizzi. *Formal Languages and Compilation*. Springer-Verlag London Limited, Italy, 2009.

[9] N. Sharef, S. Noah, and M. Murad. Issues and challenges in semantic question answering through natural language interface. *Journal of Next Generation Information Technology (JNIT)*, 4(7):50–60, 2013.