Towards a Thematic Role based Target Identification Model for Question Answering

Rivindu Perera [1] and Udayangi Perera [1]
[1] Department of Computer Science, Informatics Institute of Technology
Colombo 06, Sri Lanka

Corresponding Author Email: pererarivindu@gmail.com

Abstract

Target identification plays a crucial role in web based question answering. But still current approaches are not matured enough to extract the exact target of any given question and therefore leads the system to low precision. To address this gap in the current researches we propose thematic role based methodology to extract the target type of the question. Proposed solution is fully wrapped in the shallow semantic processing of the question rather directing it to the deep parsing. Research employs dative alternation of the question thus providing strict rule based approaches to be implemented to elicit the target with high confidence. Furthermore, the proposed   solution can be extended with semantically rich target types by mapping concepts identified in question to semantic categories. This extensibility exhibits that our new approach is scalable and can be tweaked to achieve high precision level that current methods are incapable to achieve.

Keywords: Question answering, target identification, shallow semantic processing, thematic roles.

1. Introduction

Question answering is the process of extracting the exact answer for a natural language inspired query which usually lies in the Natural Language Processing (NLP) and the Information Retrieval (IR) domains. To extract the answer with high precision, target of the question must be identified in preprocessing stages. Current approaches used in target identification are based on pattern matching approaches and rule based approaches identified through the usage [1] [2]. But drawback noticed in this approach is that such techniques cannot be extended with semantically analyzed structures for target identification.

Due to absence of semantic structures in target identification, question answering process may be subjected to several unseen issues during answer extraction. Among these issues, inability to extract the answer though there is enough information in knowledge base is considered as one of the critical issue to be fixed in future question answering. This issue is placed in even more complex stage when question taxonomies are developed with the use of

learning process which extracts question target types while processing questions formed by users [3]. Furthermore, inaccurate target identification can also lead the question answering systems to formulate incorrect answer patterns when presenting the final answer for the user thus leading them to have low confidence rates.

Therefore, we propose a solution where target identification in question answering is powered by identified thematic roles in questions. We design our heuristic in a way that future researches can also incorporate the method by extending the structure with any thematic role that need to be incorporated.

To evaluate this new paradigm we have used Scholar - question answering system which is designed with the proposed target identification method by this research. This paper will unwrap all steps taken to develop this novel method with an empirical viewpoint of each and every approach we have employed during implementation.

2. Background of the study

2.1 Target identification in question answering

Matthew and Nyberg [4] argue that question answering can be taken in to a level that can challenge human abilities only through a better extraction technique which can get the exact answer for the given query. However, in their research which warps around the OpenEphyra question answering system, shows that passage ranking is not the most important task in question answering. Ganesh et al. [5] also support this concept showing that high quality answer can only be extracted through the proper understanding of the target required by the end user. But Whittaker et al. [6] bring out that factoid question answering cannot be implemented with a preprocessed set of target types which can be selected by the end user rather this research shows the importance of dynamic target type identification in answer extraction can lead question answering systems to be more flexible and useful when such systems are used in open domain question answering.

Table 1 - Syntactic classification of user utterances from [7]

Syntactic from	
Wh-type Question	87.7% (544)
Yes-no Question	9.5% (59)
Imperative (Information request)	2.6% (16)
Declarative (Answer to clarification)	0.2% (1)

Kato et al. [7] show a practical target identification method using 4 different target types which are responsible to generate answers using categorization of answer type. Table 1 below, shows the syntactic classification of user utterances and its distribution found by Kato and his team.

According to these findings it is noted that Wh-type questions are the main type of questions that any particular question answering systems should be able to answer. But this type of a distribution cannot be considered as accurate in all the scenarios that must be handled through a open domain question answering system. Bogdan et al. [8] show that in cross-language

question answering, target of the question cannot be determined by simple rule based approach rather need to be analyzed thoroughly through semantically rich aspects.

2.2 Thematic roles

Pighin et al. [9] introduce a two-steps supervised strategy for the identification and classification of thematic roles. In this approach presented by Pighin and his team, wide variety of themes are considered providing better overview of the recognition of thematic roles and classification in a complex and wide area of natural text. However this research does not employ the verb sense information in classification stage. Therefore, in a question answering system this approach cannot be used with original structure as question answering needs verbs to be defined with high precision considering the sense they provide.

Liu and Soo [10] carried out a research in the area of knowledge acquisition considering thematic role based approach. In this novel method proposed and evaluated by this research, syntactic clues are incorporated to get the exact role to the acquisition phase. But the drawback noticed in this research is that need of extensive syntactic resources to determine the knowledge to be acquired. Therefore when applied to a question answering system this method should be trained with large amount data to make this heuristic available for all sorts of questions.

3. Method

In our approach target identification is entirely based on the thematic role identified which shows the type of the answer to be extracted. This novel paradigm is also inspired from the research carried out by Yang et al. [11] which introduces contextual question answering using relevancy recognition. But to transform this question answering process to a flexible state we also introduce the method that users are given the chance to select the thematic role that they need. However, if such thematic role is absence terms used in the question, its structure and the semantic representation are considered to extract the thematic role.

3.1 Thematic role identification

In the target identification process the first task is to identify the thematic role to be identified which later transformed in to a target type. In our approach several different thematic roles are incorporated and some of them are shown in Table 2 blow.

Table 2 - Thematic roles

Thematic role	Applicability
Agent	To get the agent role of a question. This may incorporate any object type if specified object is involved in the act playing the role of agent.
Instrument	If the question is related with an event, instruments used in the event are classified under this role
Goal	Goal thematic role can show any type of a objective such as a location, event or some other result which is carried out to invoke a different type of an event
Patient	Object type of a event is categorized under this thematic role

Once the thematic role is identified it is associated with the specified question to support the answer extraction process.

3.2 Thematic role assignment and metadata processing

Identified thematic role will be assigned to the specified question showcasing the answer type required to be extracted. But with the thematic role several other metadata can also be attached to the question to make the answer extraction process more accurate and fast. If thematic role required represent any type of supported named entity them the named entity type will also be attached to the question. For an example for a question like "who is the founder of Google" will be assigned with the "agent" thematic role. But in question processing it can be identified that this agent type is actually mapped to a "person" named entity type. Therefore, rather assigning the generic theme of agent as a metadata representation "person" named entity will also be attached to the question to support answer extraction by reducing the search space.

3.3 Answer extraction

When the thematic role is assigned to a question, answer extraction process can be stated focusing answers which represent the type required by the thematic role and which are compatible with the named entity type specified. After the extraction process, confidence level can be assigned to the extracted answer by analyzing the compatibility that answer carries with thematic role and metadata associated with the question being processed.

4. Discussion

In this paper we illustrated an approach to determine the target type of a question by analyzing the thematic role of the question to be processed. As thematic roles are based on the semantic representation of the natural text this approach can be extended to support several semantic processing tasks. Furthermore, in several stages we have employed rule based approaches to process the question as probabilistic approaches cannot be applied with semantic representation with high accuracy.

To evaluate this novel heuristic we have used the question answering system- Scholar which uses the same strategy to identify the target. During evaluation we achieved excellent accuracy which inspires us to develop this model as a independent library to incorporate with other question answering systems. In future our focus is entirely placed on the implementation of this heuristic as a library and to apply several other semantic processing methodologies to increase the accuracy level of this novel paradigm.

References

[1] A. Shtok, G. Dror, Y. Maarek, and I. Szpektor, "Learning from the past: answering new questions with past answers," presented at the Proceedings of the 21st international conference on World Wide Web, Lyon, France, 2012.

[2] R. Higashinaka and H. Isozaki, "Automatically Acquiring Causal Expression Patterns from Relation-annotated Corpora to Improve Question Answering for why-Questions," vol. 7, pp. 1-29, 2008.

[3] S. Hartrumpf, "Adapting a semantic question answering system to the web," presented at the Proceedings of the Workshop on Multilingual Question Answering, 2006.

[4] M. W. Bilotti and E. Nyberg, "Improving text retrieval precision and answer accuracy in question answering systems," presented at the Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering, Manchester, UK, 2008.

[5] G. Ramakrishnan, A. Jadhav, A. Joshi, S. Chakrabarti, and P. Bhattacharyya, "Question Answering via Bayesian inference on lexical relations," presented at the Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering - Volume 12, Sapporo, Japan, 2003.

[6] E. W. D. Whittaker, J. Hamonic, D. Yang, T. Klingberg, and S. Furui, "Monolingual web-based factoid question answering in Chinese, Swedish, English and Japanese," presented at the Proceedings of the Workshop on Multilingual Question Answering, 2006.

[7] T. Kato, F. Masui, J. i. Fukumoto, and N. Kando, "WoZ simulation of interactive question answering," presented at the Proceedings of the Interactive Question Answering Workshop at HLT-NAACL 2006, New York City, NY, 2006.

[8] B. Sacaleanu, G, and n. Neumann, "Cross-cutting aspects of cross-language question answering systems," presented at the Proceedings of the Workshop on Multilingual Question Answering, 2006.

[9] D. Pighin, A. Moschitti, and R. Basili, "RTV: tree kernels for thematic role classification," presented at the Proceedings of the 4th International Workshop on Semantic Evaluations, Prague, Czech Republic, 2007.

[10] R.L. Liu and V.W. Soo, "An empirical study on thematic knowledge acquisition based on syntactic clues and heuristics," presented at the Proceedings of the 31st annual meeting on Association for Computational Linguistics, Columbus, Ohio, 1993

[11] F. Yang, J. Feng, and G. D. Fabbrizio, "A data driven approach to relevancy recognition for contextual question answering," presented at the Proceedings of the Interactive Question Answering Workshop at HLT-NAACL 2006, New York City, NY, 2006.