QGASP: a Framework for Question Generation Based on Different Levels of Linguistic Information

Hugo Rodrigues
INESC-ID
IST - Universidade de Lisboa
Lisboa, Portugal
Carnegie Mellon University
Pittsburgh, PA, USA
hpr@l2f.inesc-id.pt

Luísa Coheur
INESC-ID
IST - Universidade de Lisboa
Lisboa, Portugal
luisa.coheur@inesc-id.pt

Eric Nyberg
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA, USA
ehn@cs.cmu.edu

Abstract

We introduce QGASP, a system that performs question generation by using lexical, syntactic and semantic information. QGASP uses this information both to learn patterns and to generate questions. In this paper, we briefly describe its architecture.

1 Introduction

As in the TheMentor system (Curto et al., 2012), QGASP (**Q**uestion **G**ener**a**tion with **S**emantic **P**atterns) creates patterns based on a set of seeds. However, contrary to TheMentor that relies on lexicon-syntactic patterns, QGASP tries to take advantage of semantic information. The use of semantic information is not new (see, for instance, Mannem et al. (2010)), but to the best of our knowledge QGASP is the first system that relies on the lexical, syntactic and semantic information in both the Pattern Acquisition (PA) and the Question Generation (QG) steps.

2 QGASP overview

Figure 1 illustrates QGASP architecture.

2.1 Pattern Acquisition

Our seeds are triples constituted by a question, its answer (optional), and a snippet that could answer that question. The question and the snippet from each seed are processed by the Stanford syntactic and dependency parsers (de Marneffe et al., 2006), and MatePlus Semantic Role Labeler (SRL) (Roth and Woodsend, 2014). A pattern is a bidirectional mapping between subtrees of the question and the correspondent snippet.

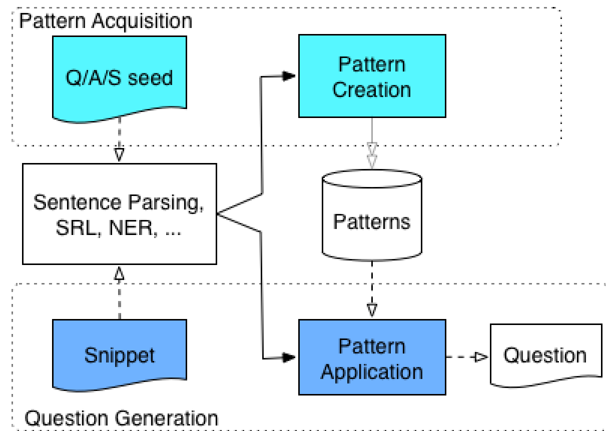

Figure 1: QGASP overview

2.2 Question Generation

Given a sentence, QGASP starts by parsing it, exactly as before; then it matches the previously learned patterns with the obtained structures.

2.3 The Matching Step

The matching step is the same, both in the PA and QG stage. Considering that a loose matching strategy will result in many patterns and questions, thus introducing noise, whereas a too restrict approach will end up in too specific patterns and low variability of questions, QGASP allows the matches to be done at lexical, syntactic and semantic level. First, it compares both subtrees by checking if their structure is the same, that is, if the subtrees' labels are syntactically equivalent and the number of children is the same (as suggested by Wang and Neumann

Proceedings of The 9th International Natural Language Generation conference, pages 242–243, Edinburgh, UK, September 5-8 2016.

(2007)). Then, QGASP checks, for each token pair, if they match. For the lexical match, lemmas are obtained from WordNet. The semantic match is based on the SRL predicted verb and a verb dictionary. This dictionary is the mapping between PropBank (Palmer et al., 2005), VerbNet (Kipper et al., 2000) and FrameNet (Baker et al., 2003), gathered from SemLink[1]. If two verbs belong to the same set in any of the resources, they are considered to match. It is also considered a semantic match if two non-verb tokens belong to the same synset, from WordNet (Miller, 1995), or if two Named Entitiess (NEs) have the same type, according to Stanford Named Entity Recognition (NER).

3 Evaluation

We tested QGASP on the Engarte corpus[2]. We used Engarte's 32 revised triples labeled as true. These triples were then used both for PA and QG, and tested in a leave one out approach (that is, if a pattern is learned from a specific sentence during the PA step, that pattern is not applied to that same sentence during the QG phase).

In the PA step we obtained 23 Semantic patterns. The generated questions with those patterns were manually evaluated by two annotators according to a simplification of Curto et al. (2012) guidelines: plausible, with exception of minor edits such as verb agreement (`y`), plausible needing context (`c`), and implausible (`n`). There are 201 questions generated, from which 92% are considered plausible of any sort – a total of 184, from which only 32 were labeled as plausible needing context. The Cohen's Kappa agreement was calculated on a subset of 115 random questions. The obtained value was 0.67, considered as a substantial agreement.

4 Conclusions and Future Work

This paper briefly describes QGASP, a framework for question generation. Although several points can be improved in QGASP, it is possible to demonstrate how seeds are learned, and how semantic features can improve the QG process.

[1] `http://verbs.colorado.edu/semlink`

[2] `http://nlp.uned.es/clef-qa/repository/ave.php`

Acknowledgement

This work was partially supported by national funds through FCT (Fundação para a Ciência e a Tecnologia) under project UID/CEC/50021/2013 and through the project LAW-TRAIN with reference H2020-EU.3.7.653587. Hugo Rodrigues is supported by a PhD fellowship from FCT (SFRH/BD/51916/2012).

References

Collin F. Baker, Charles J. Fillmore, and Beau Cronin. 2003. The structure of the FrameNet database. 16(3):281–296.

Sérgio Curto, Ana Cristina Mendes, and Luísa Coheur. 2012. Question generation based on lexico-syntactic patterns learned from the web. *Dialogue & Discourse*, 3(2):147–175, March.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the International Conference on Language, Resources and Evaluation (LREC)*, pages 449–454.

Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 691–696. AAAI Press.

Prashanth Mannem, Rashmi Prasad, and Aravind Joshi. 2010. Question generation from paragraphs at upenn: Qgstec system description. In *Proceedings of QG2010: The Third Workshop on Question Generation*, pages 84–91.

George A. Miller. 1995. Wordnet: a lexical database for english. *Commun. ACM*, 38:39–41, November.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106, March.

Michael Roth and Kristian Woodsend. 2014. Composition of word representations improves semantic role labelling. In *Empirical Methods for Natural Language Processing*, pages 407–413.

Rui Wang and Günter Neumann. 2007. Recognizing textual entailment using sentence similarity based on dependency tree skeletons. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, RTE '07, pages 36–41, Stroudsburg, PA, USA. Association for Computational Linguistics.