

# Textual Entailment for Modern Standard Arabic

Maytham Alabbas

Department of Computer Science, College of Computer Science and Information Technology

University of Basrah, Basrah, Iraq

E-mail: ma@uobasrah.edu.iq, http://faculty.uobasrah.edu.iq/faculty/534

## Thesis summary

**Keywords:** recognizing textual entailment, extended tree edit distance, tree edit distance, arabic textual entailment

**Received:** June 12, 2021

*This paper summarizes the Doctoral Thesis that examines various techniques to recognizing Arabic textual entailment, deciding whether one fragment of text entails another, where there is an exceptional level of structural and lexical ambiguities. As far as we know, the current work is the first study to apply this task for Arabic. For this purpose, we firstly describe a semi-automatic method for constructing a first Arabic textual entailment dataset. Then, we have investigated various system combination techniques for improving tagging and parsing depending on having accurate linguistic analyses. Finally, we have improved the standard tree edit distance (TED) algorithm. This extended version of TED, ETED, calculates the distance between two trees by applying operations on subtrees and single nodes. The current work also uses the artificial bee colony (ABC) algorithm to automatically guess the edit operations cost for both subtrees and single nodes and to decide thresholds. The current findings were encouraging for Arabic and English RTE-2 test sets. It should be noted most of the methodologies presented here could be utilized in research projects on poorly resources languages.*

*Povzetek: Predstavljena je doktorska disertacija za obdelavo arabskih besedil.*

## 1 Introduction

One of the essential tasks for natural language systems is to decide whether one natural text snippet entails another. Nowadays, textual entailment (TE) is considered as one of the most popular generic tasks in this regard. TE can be described as a relation between two natural sentences in which one sentence's truth, the entailing expression (T), compels the truth of the other, what is entailed (H). For instance, 'The president was assassinated.' entails 'The president is dead.', whereas the reverse does not hold.

TE definition contrasts with the standard entailment definition, i.e., T entails H if H is true whenever T is. The TE recognition task is in some ways easier than the classical entailment task. It has led to different techniques that diverge from the tradition of translating from natural language into logical forms and using standard approaches of theorem proving to determine the relationships between these logical forms.

## 2 Methodology

The current work [1,2] aims to see how well existing approaches for recognizing textual entailment (RTE) work when utilized to the Arabic language; and to provide suggestions for improvements that deal with the particular issues posed by the language. We used the TE architecture system that is illustrated in Figure 1. At each stage, we aimed to take advantage of variations on the standard machinery to assist us in overcoming the additional challenges posed by written Arabic.

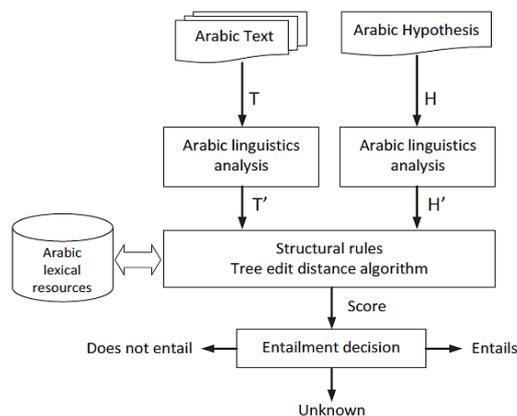


Figure 1: General diagram of current system [1].

### 2.1 Arabic linguistic analysis

Such a system depends on the presence of accurate linguistic analyses. It is notoriously difficult to obtain such analyses for the Arabic language. Concerning these problems, we looked into solutions that used system combination strategies to improve tagging and parsing to overcome these issues [3]. These strategies outperform any of the contributing tools by a significant margin [4]. A tagger and a parser are implemented as preprocessing tools to represent each sentence as dependency trees. We use the method described by [5] of merging the three taggers (MADA, AMIRA, and a maximum-likelihood tagger) based on their confidence levels, using the built-in

tokenizer from MADA to preprocess the text. In [6], we show that the combination strategy achieves 99.5% accuracy for the ‘Bies’ tagset. We then use a combination of three parsers (MSTParser plus two MALTParser algorithms) as described by [7]. This gives around 85% for labeled accuracy, which is the best Penn Arabic treebank (PATB) result we have seen. We apply these combinations in all our series of experiments.

## 2.2 Arabic TE dataset

To evaluate our Arabic TE system, an appropriate dataset is required. As far as we know, no Arabic datasets are available for the TE task; therefore, we have had to create one. We have utilized one of the approaches applied for collecting the T-H pairs in the RTE tasks, with a slight alteration. We developed in [8] a semi-automatic approach for producing a first Arabic textual entailment dataset relying on an improved version of the ‘headline-lead paragraph’ technique. We outlined the challenges that come with depending on volunteer inter-annotators to make the judgment and developed a regime to address some of these issues. There are 600 pairings in the preliminary testing dataset, each with a binary annotation of ‘yes’ or ‘no’ (a 50-50 split). This dataset is similar in size to the RTE-2 dataset but with typically longer sentences.

## 2.3 Tree matching

We investigate various systems for the task of Arabic TE, starting with basic and reliable but approximate systems and proceeding to more advanced systems. There are two primary groups of these systems [1]: surface string similarity systems (bag-of-words system and Levenshtein distance systems) and syntactic similarity systems (tree edit distance systems, our extended version of TED with subtree operations systems (ETED) [9,10,11], hybrid ETED with optimization algorithms such as ABC algorithm). Six systems out of 10 are reimplementations of existing methods that have been implemented for other languages. These serve as baselines and indicate that when applied to Arabic, the findings are comparable to those achieved with English. While four systems cover our contributions, each representing a distinct version of our system.

## 2.4 Entailment decision

This part is responsible for making the final entailment decision depending on the final score. To evaluate if this score should lead to a certain judgment, one threshold (entails/not-entail tests) or two thresholds (entails/unknown/not-entail tests) are utilized [9].

## 3 Conclusion

The current findings were extremely encouraging on the Arabic test set, notably the F-score improvement. The fact that some of these findings were replicated for the RTE2 test set, where we did not have any control over the dependency trees parser, gives some evidence for the

current approach’s robustness [1]. In both circumstances, we anticipate that having a more accurate parser (our Arabic parser achieves approximately 84% accuracy on PATB, whereas MINIPAR is estimated to reach around 80% accuracy on the Suzanne tested corpus) would improve the performance of both versions of TED.

## References

- [1] Alabbas, M. (2013). Textual Entailment for Modern Standard Arabic. PhD Thesis, The University of Manchester, Manchester, UK.
- [2] Alabbas, M. (2011). ArbTE: Arabic textual entailment. RANLP Student Research Workshop 2011, Hissar, Bulgaria, pp. 48–53.
- [3] Alabbas, M. and Ramsay, A. (2012). Combining black-box taggers and parsers for modern standard Arabic. In Proceedings FedCSIS-2012, IEEE, Wrocław, Poland, pp. 19–26.
- [4] Alabbas, M. and Ramsay, A. (2014). Combining strategies for tagging and parsing Arabic. In Proceedings of the EMNLP 2014 Workshop on ANLP 2014, pp. 73–77, doi:10.3115/v1/W14-3609.
- [5] Alabbas, M. and Ramsay, A. (2012). Improved POS-tagging for Arabic by combining diverse taggers. In Proceedings of AIAI, volume 381, Springer Berlin, pp. 107–116, doi: 10.1007/978-3-642-33409-2\_12.
- [6] Alabbas, M. and Ramsay, A. (2014). Improved Parsing for Arabic by Combining Diverse Dependency Parsers. LTC 2011, Revised Selected Papers, Lecture Notes in Computer Science, Springer, Vol. 8387, pp. 43–54, doi:10.1007/978-3-319-08958-4\_4.
- [7] Alabbas, M. and Ramsay, A. (2011). Evaluation of combining data-driven dependency parsers for Arabic. In Proceedings of LTC 2011, Poznań, Poland, pp. 546–550.
- [8] Alabbas, M. (2013). A dataset for Arabic textual entailment. RANLP Student Research Workshop 2013, Hissar, Bulgaria, pp. 7–13.
- [9] Alabbas, M. and Ramsay, A. (2013). Natural language inference for Arabic using extended tree edit distance with subtrees. Journal of Artificial Intelligence Research, 48:1-22, doi:10.1613/jair.3892.
- [10] Alabbas, M. and Ramsay, A. (2013). Optimising tree edit distance with subtrees for textual entailment. In Proceedings of RANLP2013, Hissar, Bulgaria, pp. 9–17.
- [11] Alabbas, M. and Ramsay, A. (2012). Dependency tree matching with extended tree edit distance with subtrees for textual entailment. In Proceedings of FedCSIS-2012, IEEE, Wrocław, Poland, pp. 11–18.