Robust Rule-based Approach in Arabic Processing

Abstract

A parsing system is a key element of many computer applications such as Information Retrieval, Knowledge Extraction and automatic translation. This paper presents a robust large-scale parser system for parsing Arabic sentences. From a practical point of view, the system is able to analyze real-world sentences thanks to a wide coverage of its linguistic knowledge that is realized within the DIINAR-MBC European project. The parser is designed for robustness against difficult input that cannot be parsed correctly according to the standard grammar rules in the system, whether it is an extra-grammatical, ill-formed or unexpected input. Most systems use algorithmic approaches to robustness where parsing programs are extended to include heuristics to handle defect cases. This study adopts another solution based on a robust grammar-based approach for parsing. It consists of introducing robust rules in the grammar itself and relaxing constraints if necessary. The parser has been evaluated against real-world sentences and the results were very encouraging. The parser provides 95% coverage.

References

Information Technology, vol. 6, no. 2, pp. 191-195.
- Blache, P and Azulay, D. O., 2002, ”Parsing ill-formed inputs with constraint graphs”, Lecture notes in computer science ISSN 0302-9743, Computational linguistics and intelligent text processing: Mexico City, 17-23 February.
- Ditters, E, 1992, A formal approach to arabic syntax: the noun phrase and the verb phrase, Phd, Nijmegen University, Holland.
- Ousersighni, R. 2002, La conception et la réalisation d'un système d'analyse morphosyntaxique pour l'arabe : utilisation pour la détection et le diagnostic des fautes. PHD, Lyon2 University.
- Ousersighni, R. and Ghenima, M. 2009, Un système d'analyse morphologique à large couverture de l'arabe, actes de la 2ème Conférence internationale Systèmes d'information & Intelligence Economique (www.sie.fr), IHE édition pp. 559-572, 12-14, Hammamet, Tunisie.

Index Terms

Computer Science
Text Processing
Keywords
Morphological analysis  Lexicon  Parsing  Formal grammar  Arabic language