Abstract

Persian (Farsi) language named entity recognition is a challenging, difficult, yet important task in natural language processing. This paper presents an approach based on a Local Filters model to recognize Persian (Farsi) language named entities. It uses multiple dictionaries, which are freely available on the Web. A dictionary is a collection of phrases that describe named entities. The framework is composed of two stages: (1) detection of named entity candidates using dictionaries for lookups and (2) filtering of false positives based. Dictionary lookups are performed using an efficient prefix-tree data structure. Our dictionary-based recognizer performs on Persian (Farsi) language with up to 88.95% precision, 79.65% recall, and an 82.73% F1 score using ASEM.


- http://ece.ut.ac.ir/dbrg/bijankhan/Bijankhan corpus was created in DBRG Lab. at University of Tehran ECE department.


- Micha Marcinczuk, Maciej Piasecki, Study on Named Entity Recognition for Polish
Index Terms

Computer Science
Text Processing

Keywords

Persian Natural language processing  Named Entity Recognition  Local Filters