عنوان مقاله:

A Semantic Approach to Person Profile Extraction from Farsi Documents

محل انتشار:

فصلنامه سیستم های اطلاعاتی و مخابرات, دوره 4, شماره 4 (سال: 1395)

تعداد صفحات اصل مقاله: 12

نویسندگان:

Hojjat Emami - *Department of Information and Communication Technology (ICT), Malek-Ashtar University of Technology, Tehran, Iran*

Hossein Shirazi - *Department of Information and Communication Technology (ICT), Malek-Ashtar University of Technology, Tehran, Iran*

Ahmad Abdollahzadeh Barforoush - *Computer Engineering Department, Amir Kabir University of Technology, Tehran, ,Iran*

خلاصه مقاله:

Entity profiling (EP) as an important task of Web mining and information extraction (IE) is the process of extracting entities in question and their related information from given text resources. From computational viewpoint, the Farsilanguage is one of the less-studied and less-resourced languages, and suffers from the lack of high quality language processing tools. This problem emphasizes the necessity of developing Farsi text processing systems. As an element of EPresearch, we present a semantic approach to extract profile of person entities from Farsi Web documents. Our approach includes three major components: (i) pre-processing, (ii) semantic analysis and (iii) attribute extraction. First, our system takes as input the raw text, and annotates the text using existing pre-processing tools. In semantic analysis stage, weanalyze the pre-processed text syntactically and semantically and enrich the local processed information with semantic information obtained from a distant knowledge base. We then use a semantic rule-based approach to extract the relatedinformation of the persons in question. We show the effectiveness of our approach by testing it on a small Farsi corpus. The experimental results are encouraging and show that the proposed method outperforms baseline methods

کلمات کلیدی:

Web Mining; Information Extraction; Person Profiling; Farsi Language