

عنوان مقاله:

RePersian - A Fast Relation Extraction Tool in Persian

محل انتشار:

فصلنامه بین المللی وب پژوهی، دوره 2، شماره 2 (سال: 1398)

تعداد صفحات اصل مقاله: 8

نویسندگان:

Raana Saheb-Nassagh - IUST

Majid Asgari - Department of Computer Engineering Iran University of Science and Technology

Behrouz Minaei-Bidgoli - Associate Professor and director of research in the Computer Engineering Department at
Iran University of Science and Technology

خلاصه مقاله:

The task of extracting semantic relations from raw data is called relation extraction. One of the most important fields in open information extraction is the automatically extraction of relations in any domain, especially in web mining. There are many works and approaches for relation extraction in English and other languages. Some of these approaches are based on parsing trees. Dependency parsing in the Persian language is difficult and time-consuming, since Persian is a low resource language and has also a dependency grammar and lexical structure, which affects also the speed of relations extraction in Persian. In this paper we will introduce a fast relation extraction method in Persian called RePersian. RePersian is dependent on part-of-speech (POS) tags of a sentence and special relation patterns, which are extracted by analyzing sentence structures in Persian. For finding relation patterns, RePersian searches through POS-tags that are given in regular expression forms. By matching the correct POS pattern to a relation pattern, RePersian extracts the semantic relations in a sentence. We appraise RePersian in two different scenarios on the Dadegan Persian dependency tree dataset. RePersian had on average the precisions 78.05%, 80.4% and 54.85% in finding the first argument on a relation, the second argument and the right relation between them.

کلمات کلیدی:

Relation Extraction, Persian language, Regex, POS Tag

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1035893>

