

An Automatic Indexer for Persian Language

Masoud Tashakori
Tashakori@noavar.com

M. R. Meybodi
Meybodi@ce.aku.ac.ir

Software Systems Laboratory
Computer Engineering & Information Technology Department
Amirkabir University of Technology
Tehran Iran

Abstract: In this paper an automatic indexer for Persian language has been developed and tested on a corpus of 450 abstracts of Persian texts in the domain of computer science. Experiments were performed on this collection using 32 queries. Also we study the distribution of words in Persian language and a list of Persian stop words is provided. The proposed indexer has been evaluated using two parameters recall and precision. Also in developing the automatic indexer the Persian stemmer has been used and the effectiveness of the indexer during stem finding has been studied.