Marius Cioca, Lucian-Ionel Cioca,
Cosmin Cioranu, Daniela Gifu
Romania

Extracting Features from the On-Line News for Making Templates Used in the Process of Educating the Next Generation of Politicians

Abstract

This paper presents an application for extracting features used in the process of education of politicians based on Open Source technologies, for extracting features from the public press; it enables the user to analyze data from text files and from the Internet. The analyzed data are from the field of politics, a topic which is both a current one and a cyclic process, occurring every four years for the parliamentary elections and every five years for the presidential elections. A specific political vocabulary was compiled for feature identification and analysis and for the interpretation of results obtained.

Keywords: *text classification, analysis, open source, political domain*

Introduction

Taking into account that the field of computers and information technology, as well as that of data volumes have grown exponentially, the use and development of new methods and technologies for revealing the information "hidden" in data have become imperative because such information cannot be detected by means of human analysis capacity (Hopcroft, J., Motwani, R., Ullman, J., 1979).

Communication is both a means and purpose to man, but to agents it is (still?) a means, and thus the starting point should be the object of communication, namely information (Barbat, 2002, p. 195).