

## DocSORTER AUTOMATIC TEXT CLASSIFICATION SYSTEM



### Areas of Application

The system is designed for detection and analysis of signs of thematic or emotional class of texts. It is aimed at news agencies and analytic centers processing information from open sources

### Specification

The system detects the text elements corresponding to certain concepts. This enables to detect the thematic elements containing in the text and the text subject. The system is based on reference texts stored in topic directories, so it can distinguish various topics such as physics, mathematics, politics, sports, etc.

If there is a set of emotionally colored reference texts, such as movie reviews, the algorithm can be used for sentiment analysis of the texts

Requirements: Windows XP operating system or higher

### Stage of Development. Suggestions for Commercialization

IRL3, TRL5  
Customization to types of texts and requirements of end users, upon request

### Advantages

As compared with other systems, DocSorter is designed to work with English, Ukrainian, and Russian. Due to the use of semantic analysis, in particular, UkrWordNet database, the system is aimed at semantic similarity of text elements instead of text vocabulary

### IPR Protection

IPR1, IPR3

### Contact Information

*Volodymyr Y. Taranukha*, International Research and Training Center for Information Technologies and Systems of the NAS of Ukraine and the Ministry of Education and Science of Ukraine; +38 044 502 63 19, e-mail: dep165@irtc.org.ua