ANNOTATION

This work is dedicated to the solution of controlling information flow from Internet media in a way of collecting and aggregating news by their content. Such a resource will give the ability to get to know all actual world events in a few minutes.

To achieve this goal was held an analysis of existing ways to build news collecting and aggregating module, and algorithms based on these ways, also found their drawbacks.

Work consists of 4 sections:

Section 1 is dedicated to topicality analysis of this work, overviewing of existing ways and solutions for solving task. The most popular data scraping and measuring text similarity algorithms are described over here. Also formulation of the problem is included.

Section 2 describes mathematical foundations of algorithms, choosen for realization of software, what will solve current problem. Included information about measuring text similarity in a general way and in the context of news aggregation problem. Existing ways for solving this task are described here. System quality criterias are mentioned too.

Section 3 describes software architecrure in detail. Also in this section the reasoning about choosen platform are stated. User case analysis is included here. This section contains user manual for software too.

Section 4 includes costing analysis of created software.