

ANNOTATION

The particular bachelor thesis work contains 135 p., 37 fig., 7 tabl., 2 appendixes and 13 sources.

The theme of this thesis is “Text mining of Twitter for evaluation of soccer teams’ ratings”

This research is aimed to introduce an approach for evaluation of football teams’ popularity ratings in Twitter. The approach is based on Text mining of Twitter with Naïve Bayes classifier, which training samples are formed automatically by an introduced algorithm based on Latent Semantic Analysis.

Achieved results:

- introduced algorithm for automatical formation of training sample for NB classifier, which is based on Latent Semantic Analysis
- implemented system for Text mining of Twitter, which processes data on the fly with introduced algorithm and evaluates popularity ratings of football teams in Twitter

Scientific novelty:

- introduced own approach for automatical formation of training samples based on LSA
- for the first time semantic memory model is used in combination with machine learning algorithm as an utility for sample creation

The results of the particular research can be used as a framework for Twitter Text mining systems and formation of training samples for Text mining tasks.

LSA, NB, TWITTER API, NAÏVE BAYES, DATA MINING, TEXT MINING, TRAINING SAMPLE.