УДК 004.912 : 81'33

Oleksandra Yu. Stadnii, Oleh V. Bisikalo

AUTOMATED CLASSIFICATION OF SOCIAL MEDIA USERS ON THE BASIS OF MICROBLOGS' LINGUISTIC ANALYSIS

Vinnytsia National Technical University

Анотація

Було розроблено підсистему лінгвістичного аналізу мікроблогів для автоматизованої класифікації користувачів соціальних мереж. Розробка є складовою частиною загальної інформаційної системи інтелектуального аналізу природно-мовного контенту соціальних мереж. В результаті виконання роботи було розроблено чатбот для месенджера Telegram, що дозволяє застосувати асоціативні та емоційні ознаки тексту мікроблогів для класифікації користувачів мережі.

Ключові слова: чатбот, лінгвістичний аналіз, сентиментальний аналіз, класифікація користувачів, месенджер, соціальна мережа.

Abstract

A subsystem of linguistic analysis of microblogs was developed in order to design automated classification of social networks' users. Development is an integral part of the general information system of intellectual analysis of natural-language content of social networks. As a result of the work, a chat bot for the Telegram messenger was developed, allowing the use the associative and emotional features of microblogs' text to classify network users.

Keywords: chatbot, linguistic analysis, sentiment analysis, classification of the users, messenger, social network.

Introduction

The relevance of research. We live in an age when the amount of information produced by a human is greater than ever and the volume of that data is increasing every day. Though, one can gain significant benefit from this information only when properly processed and analyzed [1].

Now, gigabytes of new data of various kinds are being created every moment around the world: new pictures are being taken, videos are being recorded, hundreds of reviews are being made in online stores, thousands are commenting on Facebook entries, dozens of reviews are being made for movies in online cinemas, stock prices are soaring, then they fall. And much of this raw information is practically useless. To get benefit from it, one has to filter and process them. At a time when technology was not so advanced, these operations had to be done manually. It took hours, days, weeks, sometimes months.

Considering the fact that earlier the information for processing was several times smaller, it is easy to understand that it is simply impossible to process such volumes manually. Therefore, many algorithms have been developed that allow these operations to be done using computer technology. Analysis of such methods in relation to natural language processing will be represented in this paper [2].

Social networks are an essential part of the Internet. They have spread our everyday lives and it is rare to meet a person who does not have an account on Facebook, Instagram, Twitter or the forbidden Vkontakte – and even everyone at the same time. It is through these resources that people quickly receive and disseminate information, communicate. But modern social networks are not just a place to chat with friends online. They are a good playground for marketing, sociological research, dating, interest grouping, information, influence, advocacy, intelligence tool for identifying socially dangerous elements and more. User classification algorithms are already widely used by social networks themselves to support user engagement [3].

The purpose of the study is to improve the classification of social network participants by a specific attribute for a wide range of people based on creating the chatbot that would analyze users' profiles, news and reactions of people in social networks to the important events.

Tasks solved in the work: determining and description of the key approaches of NLP, reviewing existing libraries, developing methods for automated classification of social network users on the base of content and sentimental analysis, development and testing the chatbot as appropriate software.

The object of study is the process of processing information to people classification in the social networks.

The subject of research – methods of automatic classification of participants of social networks.

Research methods: analysis, analogy, modeling, classification and experiment.

The scientific novelty is to automate the construction of methods for classifying social network participants. Unlike existing solutions, the proposed development will work on the basis of associativity (words and concepts will be grouped by semantic similarity), which allows improving the quality of classification of social network participants. The other thing that is unique is that everything is managed via chatbot which makes using the software extremely easy and comfortable.

Practical value. Created models, algorithms and software can be used for corporate, government, or personal purposes, for example, observing a specific sample people in sociological research, finding a target group experience in marketing research, identifying psychologically unstable or potentially dangerous individuals for society, etc.

1 Work with Twitter API

To start work with Twitter API we have to register in Twitter in order to apply for a developer account. Next step is to create an application. Since in the recent year the process of Twitter application registration has become more complicated (the developer has to write quite a long and descriptive text with the reasoning why he needs to access Twitter data), we will use an old application (figure 1) that has been created in 2018 within the confines of writing the bachelor thesis – Twitter allowed old applications not to reregister. The only thing we need from it are keys and tokens, so it suits perfectly. There are four credentials we need to access Twitter: API key, API secret key, access token, access token secret. This data is confidential and only developer should have an access to it.

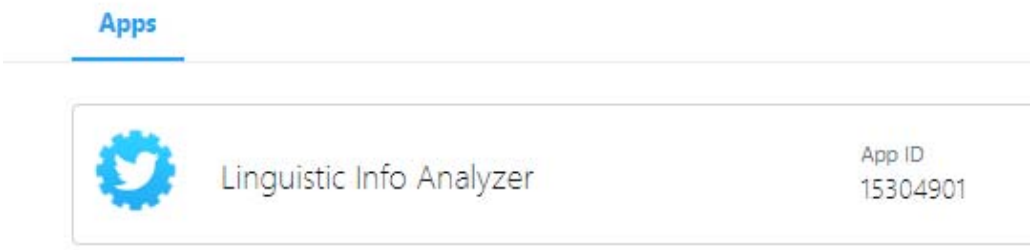

Figure 3.1 – Our Application in Twitter

For more convenient work with Twitter API we will use a special Python library – Tweepy, it works well with the tweets, longer than 140 characters and is brief and reliable.

2 Interaction with Telegram Bot API

As we mentioned before, Telegram has a bot that is a boss of all other bots. Its name is BotFather. BotFather is a very handy tool that simplifies and standardizes the creation of apps of this type (figure 2). To create a bot, we will send him command /newbot. After that bots offers us to write a name and username for our bot (figure 3). The main conditions – it must end with the word “bot” so the users will always understand that they are chatting with it, not with the human-being.

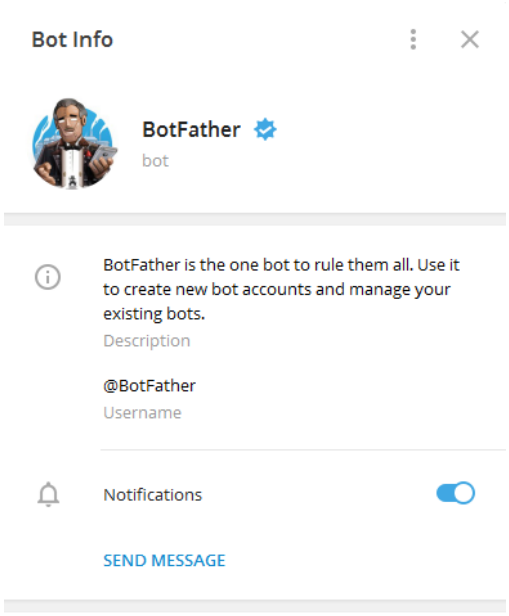

Figure 2 – BotFather in Telegram

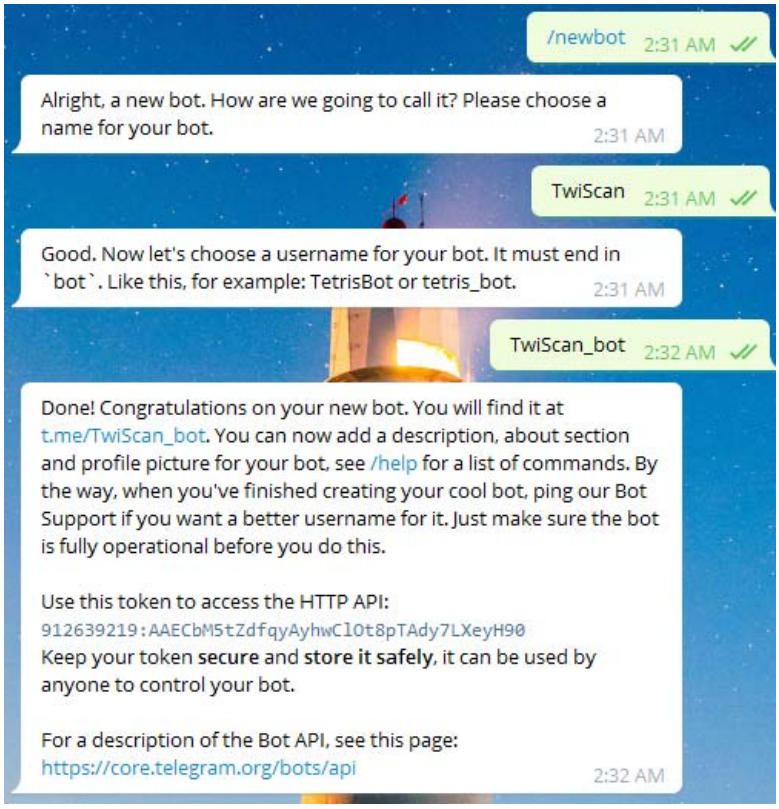

Figure 3 – The process of bot creation

After bot registration we get access token that is secret and may is used for bot control. We also added a description – users will see it in the bot info window (figure 4) and about information that will be visible while

we just start our work the bot (figure 5). Telegram does not limit the developers very much, but kindly asks them to implement the standard command – /start and /help (as you may have noticed, all the commands in Telegram start with the slash). We implemented them in our bot as well (figure 6). It is done with the purpose to make bot usage intuitive and convenient.

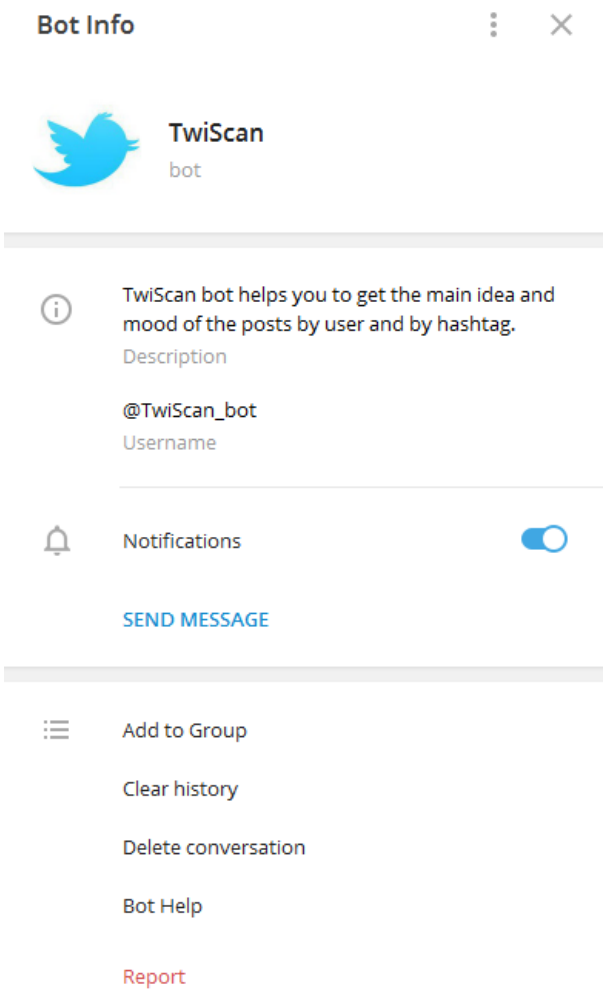

Figure 4 – Bot Info

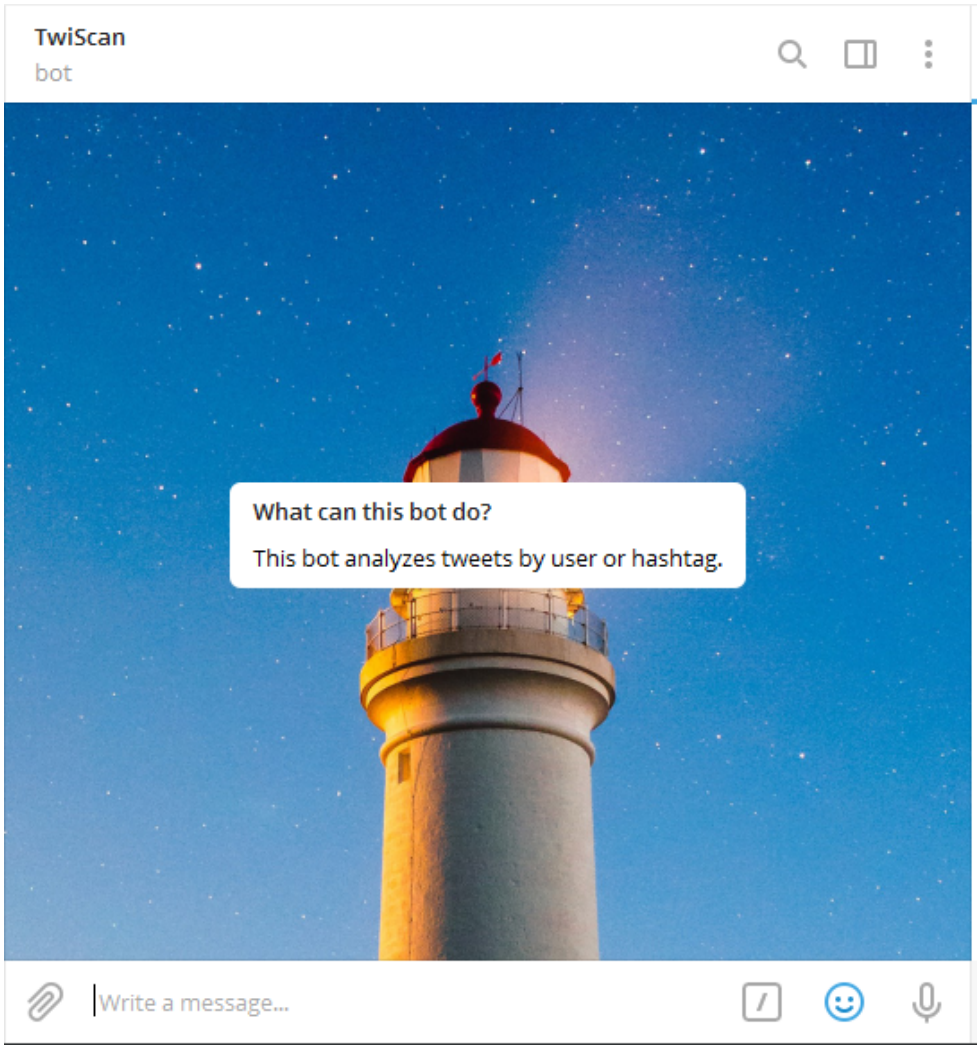

Figure 5 – Start of the interaction with bot

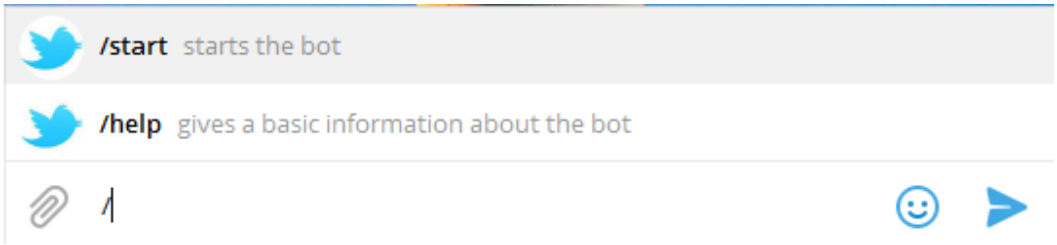

Figure 6 – Standard commands implemented in our bot

3 Application Functionality Overview

When we first open the bot, it gives us a brief explanation about its functions, as it is shown on the figure 6. To start bot user has to call /start command. After that bot asks how would user like to search the posts for

analysis. There are two options: by hashtag or some keyword or by username. These options are represented as keyboard buttons (figure 7).

After choosing a single option user has to input a number of tweets that must be analyzed. Bot represents a result of keyboard analysis (the quantity of keywords is limited to 10, because usually 10 keywords are enough to get the meaning of idea of text). It also shows us sentimental analysis of tweets. The keyboard and sentimental analysis is done on the basis of TextBlob tools. The diagram is built with the use of matplotlib – a library for data visualizing. In case if the search is done by username, we also get general information about his profile. It is important to remember that Twitter has daily limits, so it is impossible to use app if the limit has been reached. The result of application performance is shown on figures 8-11.

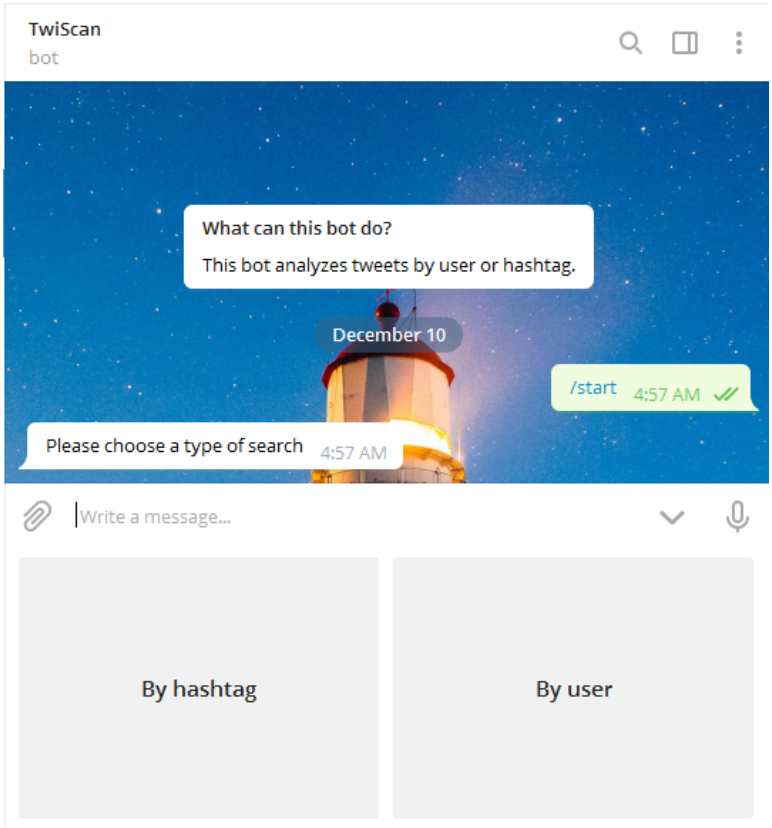

Figure 7 – Choosing a type of search

Figure 8 – Keywords analysis of tweets that contain hashtag “hongkong”

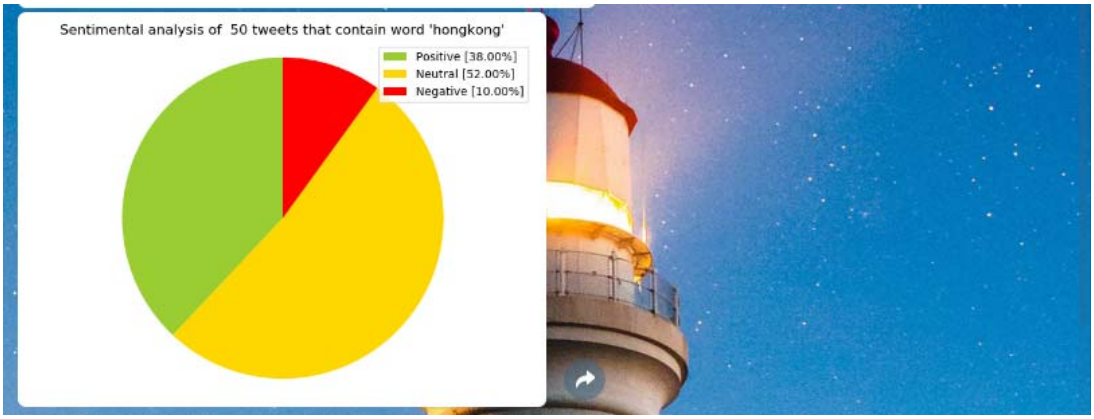

Figure 9 – Sentimental analysis of tweets that contain hashtag “hongkong”

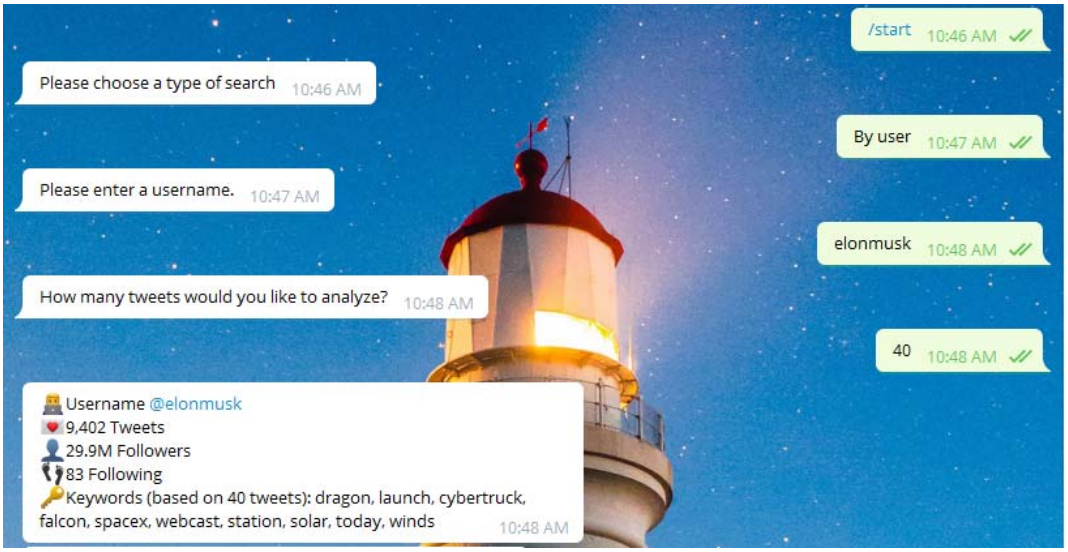

Figure 10 – Keywords analysis of tweets of user with the username “elonmusk”

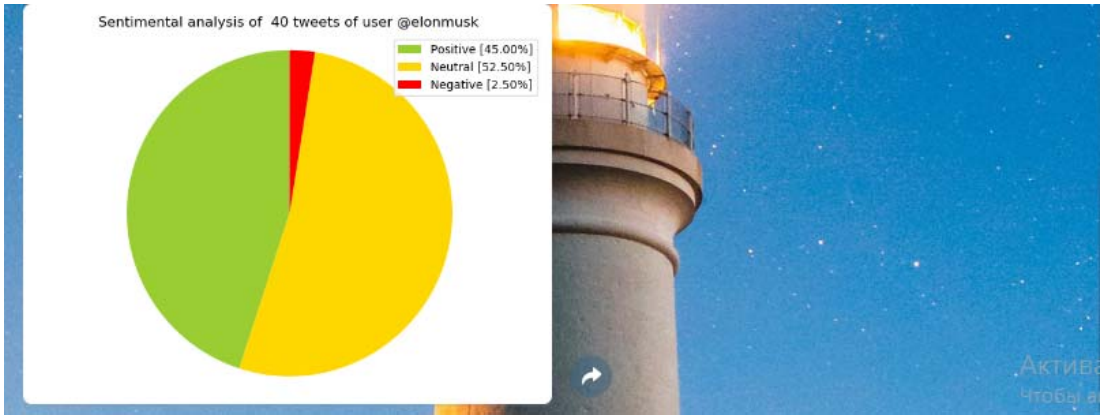

Figure 11 – Sentimental analysis of tweets of user with the username “elonmusk”

Conclusions

The chatbot application for semantic and sentimental linguistic analysis of the social media participants’ publications been developed. We were decided to create a chatbot in Telegram, since it is powerful and perspective messenger that has one of the best bot API and was a pioneer as the chatbot development platform. Twitter was chosen as a target social media since it is the greatest network for opinion exchange. Developed tool determines mood of the people.

REFERENCES

1. Manning C. Foundations of Statistical Natural Language Processing / C. Manning, H. Schütze – The MIT Press, 1 edition, 1999 – 620 p.
2. Text categorization [Electronic Source]. – Mode of Access: http://www.scholarpedia.org/article/Text_categorization.
3. Oleg V. Bisikalo ; Sergei M. Dovgalets ; Paweł Pijarski and Anna I. Lisovenko "Development of dialog system powered by textual educational content", *Proc. SPIE* 10031, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2016, 100314E (September 28, 2016); doi:10.1117/12.2248863.

Стадній Олександра Юріївна – студентка групи 3АКІТ-18м, Вінницький національний технічний університет, м. Вінниця, e-mail: alix.stadniy@gmail.com

Науковий керівник: ***Бісікало Олег Володимирович*** – д-р техн. наук, декан факультету КСА, Вінницький національний технічний університет, м. Вінниця, e-mail: obisikalo@vntu.edu.ua

Stadnii Oleksandra Yuriyivna – student of the Faculty of Automation, Electronics and Computer Control Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: alix.stadniy@gmail.com

Supervisor: ***Bisikalo Oleg V.*** – Dr.Sc. (Eng.), Professor, Dean of the Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia, e-mail: obisikalo@vntu.edu.ua