

# Information system for intellectual analysis of natural-language content of social networks

**Developed by:** Y.Kotyk

**Scientific director:** O.Bisikalo



# Part 3. The subsystem of determination of sentimental characteristics of participants of social networks on the basis of linguistic analysis of microblogging

Developed by: Y.Kotyk





# Purpose and research objectives

## Purpose

The purpose of the work is to solve the problem of computer linguistic on the basis of ready-made libraries, algorithms and tools of the NLTK linguistic software package, in particular in carrying out sentimental analysis of natural-language texts on the basis of communication between the members of the sentence.

## Objectives

- 1) Analysis of areas of artificial intelligence in terms of the tasks of computational linguistics;
- 2) Development and testing of the sentimental text analysis software.



## Object and subject of the study and scientific novelty

### Object

The object of study is the process of automated determination of indicator characteristics of profiles of participants in social networks, as well as the process of determining the response of users to certain events.

### Subject

The subject of this study is the methods and means of solving the problem of sentimental analysis of English texts using the NLTK software package.

### Novelty

The proposed approach allows to improve existing methods of sentimental analysis of natural-language texts on the basis of communication between the members of the sentence.



## Development tools, libraries and frameworks



**PyCharm** **PyCharm, Python** (development tools)

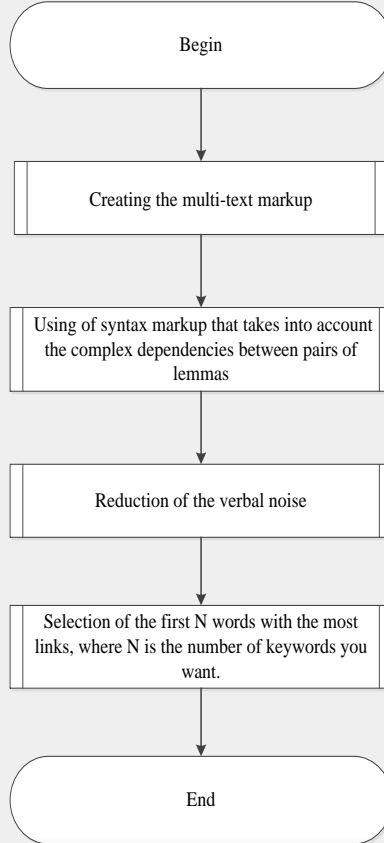


**Twitter API** (source of the data)



**Natural Language Tool Kit** (library for natural language analysis )

# Algorithm of the definition of key words in the text microblogging



phpMyAdmin

Server: MySQL 8.0 Server:3306 - Database: mathling - Table: posts

Recent Favorites

- information\_schema
- mathling
- posts
- users

Showing rows 0 - 24 (274792 total, Query took 0.4892 seconds.) [id\_str: 1000002552073736193... - 1000065309120876544...]

SELECT \* FROM `posts` ORDER BY `id\_str` ASC

1 > >> | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

			id	id_str	created_at	text	user_id
<input type="checkbox"/>				1000002552073736193	Fri May 25 13:16:02 +0000 2018	Flowing into the long weekend like... #MemorialDay ...	38441438
<input type="checkbox"/>				1000003558937382912	Fri May 25 13:20:02 +0000 2018	RT @keisi_keisi: россия как же хочется простого ч...	3412622103
<input type="checkbox"/>				1000005316422008832	Fri May 25 13:27:01 +0000 2018	RT @SciNational: It may seem like a repeat of a 2...	236786367
<input type="checkbox"/>				1000007621091512320	Fri May 25 13:36:10 +0000 2018	City with their unlimited spend can have their rec...	221289012
<input type="checkbox"/>				1000016663977316352	Fri May 25 14:12:06 +0000 2018	Mes sœurs d'engagement pour la cause de la santé d...	29499388
<input type="checkbox"/>				1000017864747823104	Fri May 25 14:16:53 +0000 2018	Good to meet the MG Environmental Trust today on s...	2245759171
<input type="checkbox"/>				1000019861916082183	Fri May 25 14:24:49 +0000 2018	RT @BuzzFeed: "You're damn right it's content" hit...	335534204
<input type="checkbox"/>				1000021066100756480	Fri May 25 14:29:36 +0000 2018	When you're negotiating who's turn it is to crutch...	393028587
<input type="checkbox"/>				1000022597445169152	Fri May 25 14:35:41 +0000 2018	RT @ASCAP: A moment between #ASCAP President @IMPa...	29572254
<input type="checkbox"/>				1000022947120037888	Fri May 25 14:37:04 +0000 2018	RT @cinematic_void: Tonight: HOCUS POCUS in 35mm a...	29572254
<input type="checkbox"/>				1000023067727351809	Fri May 25 14:37:33 +0000 2018	RT @JennyBicks: @CDanielChelliah @GreatestShowman ...	29572254
<input type="checkbox"/>				1000027987675402240	Fri May 25 14:57:06 +0000 2018	RT @natefreiman: There's been a lot written recent...	226769484
<input type="checkbox"/>				1000029081956765696	Fri May 25 15:01:27 +0000 2018	Freaking out on the @UAudio Eden Bass Amp plugin...	29655016
<input type="checkbox"/>				1000029510941773826	Fri May 25 15:03:09 +0000 2018	JP day!!!! congrats lads! https://t.co/G5JoA1rX...	29655016
<input type="checkbox"/>				1000036297149440002	Fri May 25 15:30:07 +0000 2018	We're kicking off #gimblefest on June 16th at 12:3...	195738329
<input type="checkbox"/>				1000045202751479809	Fri May 25 16:05:30 +0000 2018	@MegPrk Love this!	252274930
<input type="checkbox"/>				1000045659213283329	Fri May 25 16:07:19 +0000 2018	@01FirstSecond @verabee Bought this for my girls, ...	252274930
<input type="checkbox"/>				1000046098990252032	Fri May 25 16:09:04 +0000 2018	@bensears Plus Man late night doodle from last nig...	252274930
<input type="checkbox"/>				1000049531151699971	Fri May 25 16:22:42 +0000 2018	Yess!!!! https://t.co/ouibH4Trz	29655016
<input type="checkbox"/>				1000049748982681600	Fri May 25 16:23:34 +0000 2018	SAT WORLDWIDE is excited to announce "We are growi...	120872742
<input type="checkbox"/>				1000056393129431040	Fri May 25 16:49:58 +0000 2018	Last night's Dialogue on women in politics was ama...	32056006
<input type="checkbox"/>				1000064570373877761	Fri May 25 17:22:28 +0000 2018	RT @DollSkinUK: HAPPY BIRTHDAY @punk_plant!! You'...	2460019718
<input type="checkbox"/>				1000064712934150144	Fri May 25 17:23:02 +0000 2018	@DollSkinUK Your car karaoke meme edits are legend...	2460019718
<input type="checkbox"/>				1000064986402697222	Fri May 25 17:24:07 +0000 2018	All I want for my birthday is for @MaxPortnoy to ...	2460019718
<input type="checkbox"/>				1000065309120876544	Fri May 25 17:25:24 +0000 2018	Fan signing in Australia 🇦🇺 gotta love the... http...	85538853

↑  Check all With selected:

Console > >> | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

comments\_keywords.zip 5 923 файла

- output\_without\_stop\_words\_frequency[100306472365326340\_text](2019-12-06\_20-07-52).txt
- output\_without\_stop\_words\_frequency[10031786885135168\_text](2019-12-06\_20-08-20).txt
- output\_without\_stop\_words\_frequency[10036661219\_text](2019-12-06\_20-08-58).txt
- output\_without\_stop\_words\_frequency[100970454898057216\_text](2019-12-06\_20-09-32).txt
- output\_without\_stop\_words\_frequency[101036155994251266\_text](2019-12-06\_20-10-04).txt
- output\_without\_stop\_words\_frequency[101272445423583232\_text](2019-12-06\_20-10-37).txt
- output\_without\_stop\_words\_frequency[101354579337285632\_text](2019-12-06\_20-11-14).txt
- output\_without\_stop\_words\_frequency[1013719339053056\_text](2019-12-06\_20-11-46).txt
- output\_without\_stop\_words\_frequency[101372727486726144\_text](2019-12-06\_20-12-19).txt
- output\_without\_stop\_words\_frequency[101416893755293696\_text](2019-12-06\_20-12-51).txt
- output\_without\_stop\_words\_frequency[101438751158435841\_text](2019-12-06\_20-13-27).txt
- output\_without\_stop\_words\_frequency[101460319041372160\_text](2019-12-06\_20-13-57).txt
- output\_without\_stop\_words\_frequency[101746502032031744\_text](2019-12-06\_20-14-29).txt
- output\_without\_stop\_words\_frequency[101813227926728705\_text](2019-12-06\_20-15-00).txt
- output\_without\_stop\_words\_frequency[101867799185588224\_text](2019-12-06\_20-15-31).txt
- output\_without\_stop\_words\_frequency[10198835536994304\_text](2019-12-06\_20-15-54).txt
- output\_without\_stop\_words\_frequency[101989412962963456\_text](2019-12-06\_20-16-25).txt
- output\_without\_stop\_words\_frequency[102035381729509377\_text](2019-12-06\_20-16-56).txt
- output\_without\_stop\_words\_frequency[102089289185230848\_text](2019-12-06\_20-17-26).txt
- output\_without\_stop\_words\_frequency[10238415774\_text](2019-12-06\_20-17-57).txt
- output\_without\_stop\_words\_frequency[102467219417600001\_text](2019-12-06\_20-18-27).txt

output\_without\_stop\_words\_frequency[116617091083280384\_text](2019-12-06\_21-26-44)...

```
jone 6
procter 4
borgman 4
panelist 3
wouter 2
mcperson 2
rob 1
digitalscholarship 1
tara 1
christine 1
clifford 1
tatum 1
phil 1
pochoda 1
steve 1
paul 1
```





## Conclusion

1

It was made a detailed analysis of existing frameworks that allow to solve the problem of computational linguistics.

2

The advantages of the developed solution were substantiated and fully described on the example of problems that can be solved with its help.

3

Developed and successfully tested an algorithm based on the tools software package NLTK which analyzes text and counts keywords between the members of the sentences