



# Дослідження методів емоціонального окрасу тексту з емотіконами

Науковий керівник:  
доцент  
Вечур О.В.

Виконав:  
ст. гр. ПЗСм-16-2  
Пугачов Є.А.



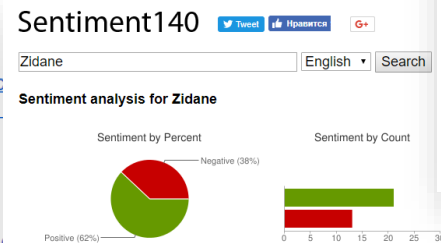
# Мета роботи

---

- ❖ Дослідити методи емоціонального окрасу тексту з емотіконами
- ❖ Визначити метод для аналізу тональності тексту на основі емотіконів
- ❖ Оцінити практичне значення отриманих даних



# Системи з аналізу тональності



**Tweets about: Zidane**

XAVIERN\_AFC: RT @MrFutboy\_: **Zidane** was a manager for 5 minutes and already has a better CV and legacy than Pep loooooool  
Posted: 16 seconds ago

abu3aseed: RT @biolakazeem: **Zidane** had to leave. How can anyone top 3 UCLs without having to spend £500m? He will go down as the greatest manager ever?  
Posted: 17 seconds ago

Nasrafat: RT @ASHAKURHUSSEIN: In resigning as Real Madrid's Head Coach, Zinedine **Zidane** may have taken POLITICAL RESPONSIBILITY for @SergioRamos' dan?  
Posted: 18 seconds ago

The results for this query are:





# Постановка задачі

---

- ❖ Проаналізувати методи які використовуються для аналізу емоційного окрасу тексту з емотіконами
- ❖ Розробити систему для визначення тонального забарвлення тексту
- ❖ Веб-система повина надавати змогу аналізувати повідомлення з соціальної мережі Twitter
- ❖ Твіти повині аналізуватися за декількома методами



# Сентімент аналіз

---

Під словами аналіз тональності слід розуміти область комп'ютерної лінгвістики, що займається вивченням думок і емоцій в текстових документах. Аналіз тональності використовується для знаходжень думок і визначення їх властивостей, відносно вхідного тексту.



# Сентімент аналіз

😊	:)	:-)	:smile:	Посмішка
😄	:D	:-D	:grin:	Широка посмішка
😉	;-)	;-)	:wink:	Підморгування
😞	:(	:-( :(	:sad:	Смук
😮	:o	:-o	:eek:	Здивування
😐	:	:-	:neutral:	Нейтральне
😱	8O	8-O	:shock:	Ботанік
😡	:x	:-x	:mad:	Роздратування
😝	:P	:-P	:razz:	Дражнити
😕	:?	:-?	:???:	В сумніві

Емотікони, емограмма або смайлик - це графічний символ, який використовується для вираження емоції.



# Метод на основі словника

---

При використанні словникового підходу ключову роль відіграє використання тональних словників.

Такі словники, як правило, представляють собою списки слів, які допомагають визначити ставлення автора до деякого об'єкту.

Кожне слово має свою оцінку яка враховується при визначенні кінцевого результату.

//positive	//negative
true,2	loser,3
trust,1	losing,3
trusted,2	loss,3
unbiased,2	lost,3
unequaled,2	lowest,1
unified,1	lugubrious,2
united,1	lunatic,3
unmatched,1	lunatics,3
unstoppable,2	lurk,1
untarnished,2	lurking,1
useful,2	lurks,1
usefulness,2	mad,3
vested,1	maddening,3
vibrant,3	made-up,1
vigilant,3	madly,3
vindicate,2	madness,3



# Метод наївного Баєса

---

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

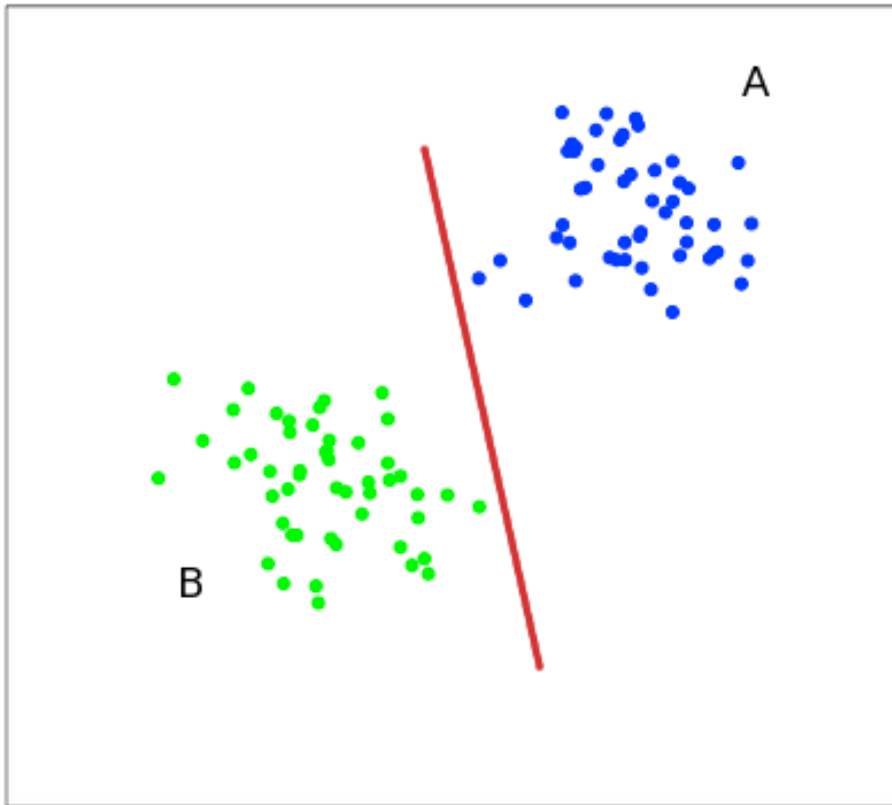
$$\log \frac{D_C}{D} + \text{foreach}(\text{word}) \left\{ \log \frac{W_{ic} + 1}{|V| + L_C} \right\} \quad (2)$$

де  $D_C$  – кількість документів в навчальній вибірці, що належать класу  $C$ ,  $D$  – загальна кількість документів в навчальній вибірці,  $|V|$  – кількість унікальних слів у всіх документах навчальної вибірки,  $L_C$  – сумарна кількість слів у документах класу  $C$  в навчальній вибірці,  $W_{ic}$  – скільки разів  $i$ -е слово зустрічалося в документах класу  $C$  в навчальній вибірці.





# SVM метод



Метод опорних векторів (Support Vector Mashine, SVM).

Визначає приналежність об'єктів до класів за допомогою кордонів областей.

Один з найбільш популярних методів навчання по прецедентам.

Може використовуватися для розпізнавання образів або створення спам фільтрів.

Даний метод відноситься до бінарних класифікаторів, хоча може використовуватися для задач мультікласифікації.



# Метод на основі емотіконів

```
{
  "name": "slightly_frowning_face",
  "emoji": "😞",
  "polarity": -1
},
{
  "name": "slightly_smiling_face",
  "emoji": "😊",
  "polarity": 1
},
{
  "name": "smile",
  "emoji": "😄",
  "polarity": 2
},
{
  "name": "smile_cat",
  "emoji": "😸",
  "polarity": 2
},
},
```

Емотікони не залежать від мови і не підпорядковуються граматичним правилам, будучи поняттям наднаціональним, одним з основних призначень якого – вираз невербальної інформації, емоцій.

Один емотікон може мати декілька варіантів відображення.



# Головна сторінка веб-системи

## Emoticon Meaning

Search tweet

star wars

Tweet Url

1002898877597577217

Message

It was a good day

Just smile

☒ Analysys only by smile

Dictionary

☒ Use dictionary method

SVM

☒ Use SVM method

Naive Bayes

☒ Use Naive Bayes method

Start



# Сторінка результатів

Emoticon Meaning

## AnalyzeTweet

[Search](#) [Tweet](#) [Text](#)

**TweetId**  
1002898877597577217

**TweetAuthor**  
pugachoveu

**TweetText**  
I was very happy to study at this university, after its graduation, I will have new oport

**SmileResult**

**DictionaryResult**

**SvmResult**

**BayesResult**

[Back](#)

Emoticon Meaning

## AnalyzeTweet

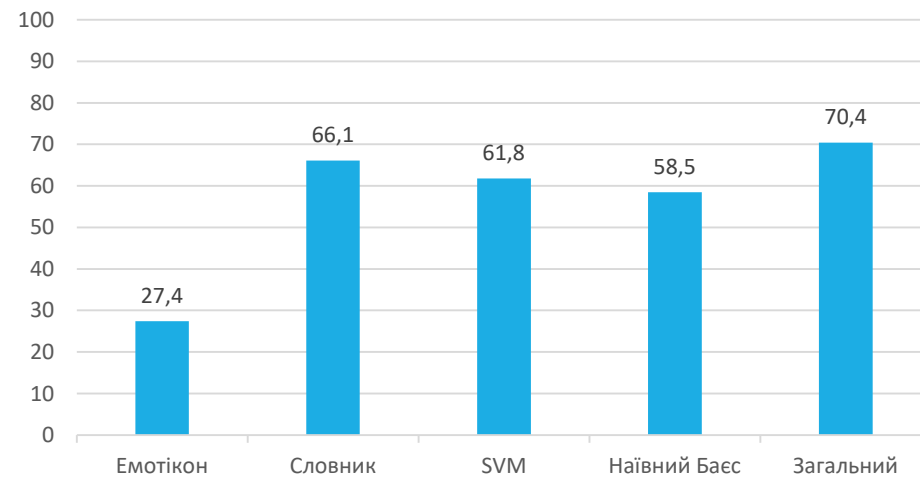
[Search](#) [Tweet](#) [Text](#)

Tweet Author	Text	Smile	Distionary	SVM	Naive Bayes
thepodcouple	RT @NerdyThingsPod: 🌟 NEW EPISODE 🌟 This week we talk Pokémon news, #TWD, AND OUR MAAAAIN TOPIC is Solo: A Star Wars Story!! We go talk what...				
BrothersBinge	RT @kiersmclean: REVIEW! #SoloAStarWarsStory is a blast from beginning to end! @RealRonHoward and everyone involved did a great job. #StarW...				
one_abud	RT @heroichollywood: 'Star Wars': Ron Howard Retweets 'Solo' Praise That Bashes 'Last Jedi' https://t.co/J5EkmMcebm https://t.co/BOPx51P6J5				
jedifarfy	"Here's this detail that everyone noticed 10 years ago when this book/movie/song was released and I just noticed!" = 10k upvotes. Fandoms on reddit, or why I've had to unsub from things like Harry Potter and Star Wars. :)				
ManuelDuarte24	Won't even dare to check the comments on the latest post on the Star Wars page.				
T1meTraveler01	RT @Lost789Boy: Ways the world will implode due to an overdose of haterade: 1) The Fountainhead directed by Zack Snyder nominated & earnin...				
Warmustbeend123	RT @heroichollywood: 'Star Wars': Ron Howard Retweets 'Solo' Praise That Bashes 'Last Jedi' https://t.co/J5EkmMcebm https://t.co/BOPx51P6J5				

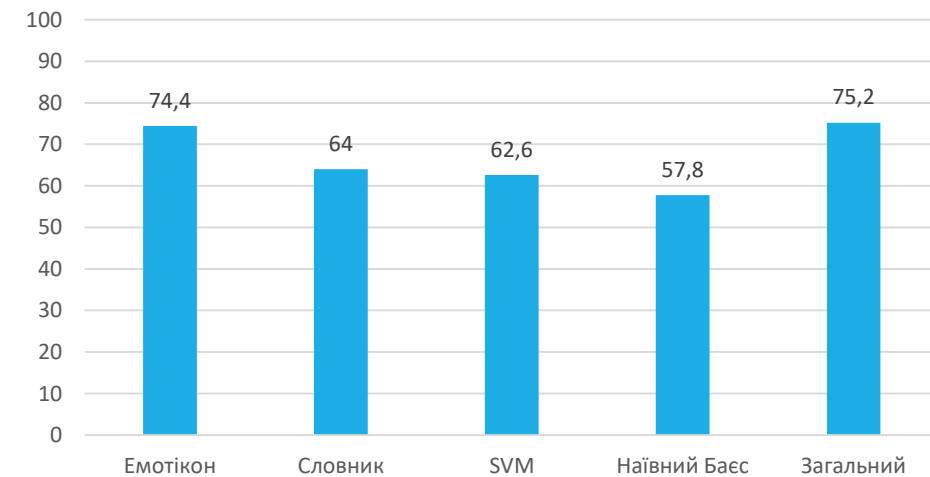


# Результати

Класифікація текстів

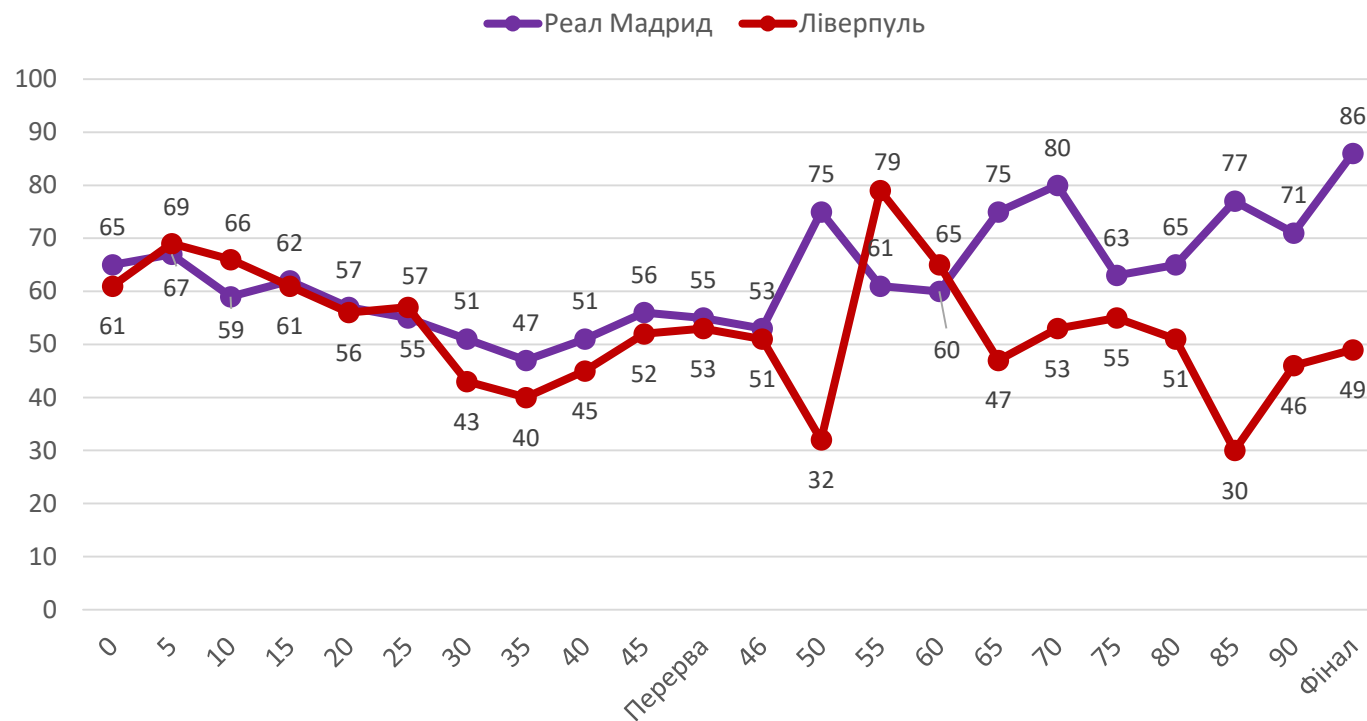


Класифікація текстів з емотіконами

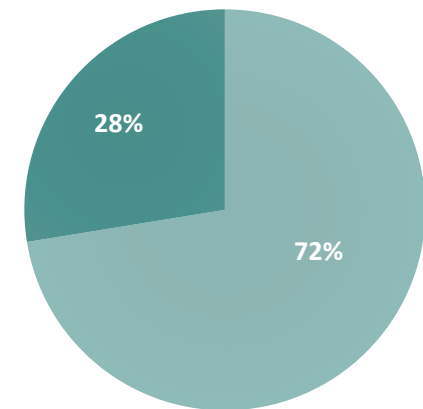




# Прикладне використання



Відгук на фільм  
“Solo: A Star Wars Story”



■ Позитивні ■ Негативні



# Висновки

---

- ❖ Досліджені методи які використовуються для аналізу емоційного окрасу тексту.
- ❖ Спроекована і розроблена система для визначення тонального забарвлення тексту. Веб-система надає можливість аналізувати повідомлення з соціальної мережі Twitter.
- ❖ Знайдено практичне використання отриманих результатів роботи
- ❖ Шляхи до удосконалення:
  - ❖ Розширення інформаційної бази, яка використовується для аналізу текстів
  - ❖ Використання для аналізу біграм та уніграм
  - ❖ Додання інтеграції з іншими соціальними мережами