



Ідентифікація емоційного окрасу повідомлень в соціальних мережах

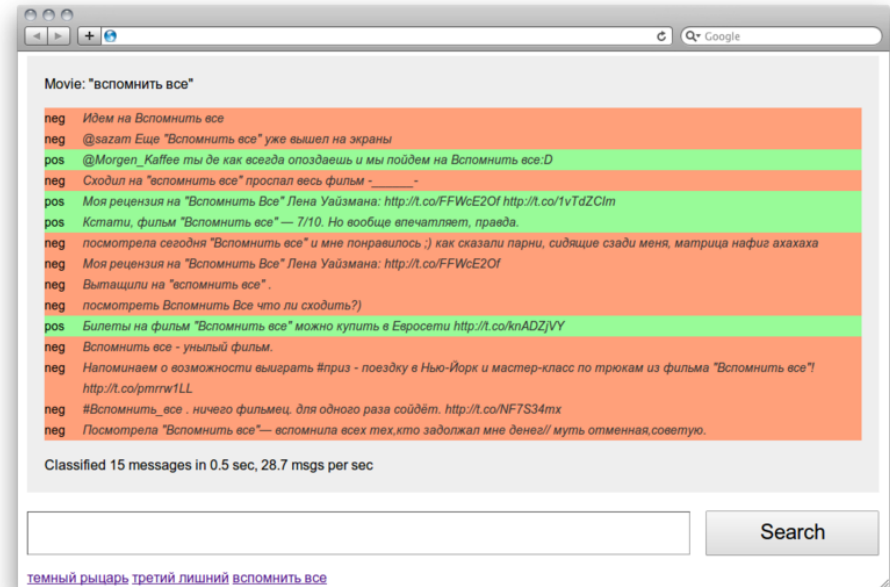
Виконав:
Сокирко Н.О., КА-44
Дипломний керівник:
Асистент Кухарев С.О.

Предмет

2

Предметом моєї роботи є аналіз емоційної тональності текстів за допомогою методу, правил та словників.

Цей метод заснований на пошуку емотивної лексики (лексичної тональності) в тексті по заздалегідь складеним тональним словникам і правилам із застосуванням лінгвістичного аналізу.



Об'єкт та мета

3

- Об'єктом роботи виступають публікації (так звані «пости») в соціальній мережі Твіттер.
- Метою дослідження є навчитися ідентифікувати емоційний окрас публікацій в соціальній мережі Твіттер.



BradTheLadLong

@BradTheLadLong

Follow

I apologize to those who love Twitter but the abuse I got on here was disgusting and that is why action needed to be taken.

[#SaveTwitter](#)

9:07 PM - 10 Aug 2016

43 49



Donald J. Trump

@realDonaldTrump



Even Crazy Jim Acosta of Fake News CNN agrees: "Trump World and WH sources dancing in end zone: Trump wins again...Schumer and Dems caved...gambled and lost." Thank you for your honesty Jim!

6:31 AM - Jan 23, 2018

12,766 12,188 52,627



Постановка задачі

- Провести огляд існуючих методів автоматичного аналізу емоційного забарвлення текстів.
- Провести дослідження особливостей текстових повідомлень в соціальній мережі Твіттер в контексті розробки методів аналізу їх емоційного забарвлення.
- Розробити програму, яка реалізує метод автоматичного аналізу емоційно забарвлених повідомлень в соціальній мережі Твіттер.

Огляд предметної області

5

У контексті завдання аналізу емоційного забарвлення тексту, основними особливостями повідомлень в Твіттер є:

- Малий розмір повідомлення - 140 символів.
- Наявність сленгу, скорочень і граматичних помилок.
- Наявність спеціальних символів.
- Використання посилань на інших користувачів і на зовнішні ресурси.



WWE @WWE

8 mins

Who will win "#Trending Superstar of the Year" at the #SlammyAwards? **Zack Ryder**, Daniel Bryan, Cody Rhodes or Dolph Ziggler? #MakeThemTrend!



Sergey Bartunov @sbos

14 Oct

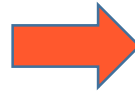
Для исследования нужна помощь в составлении тестовой выборки. RT please! j.mp/qeSTmQ
↻ Retweeted by コーヒー豆

Огляд предметної області

6

Аналіз тональності тексту включає в себе такі основні завдання:

- Визначення **полярності речення.**



$$P = \frac{\sum_{i=1}^n x_n}{n},$$

- Витяг аспектів з емоційно забарвленого тексту.

- n - кількість слів;
- x - вага слів.

Методи визначення тональності тексту

7

Існують 3 методи:

1) Векторний аналіз.

2) Пошук емотивної лексики (лексичної тональності) в тексті по раніше складеним тональним словникам з застосуванням лінгвістичного аналізу.

3) Змішаний метод (комбінація першого і другого підходів).

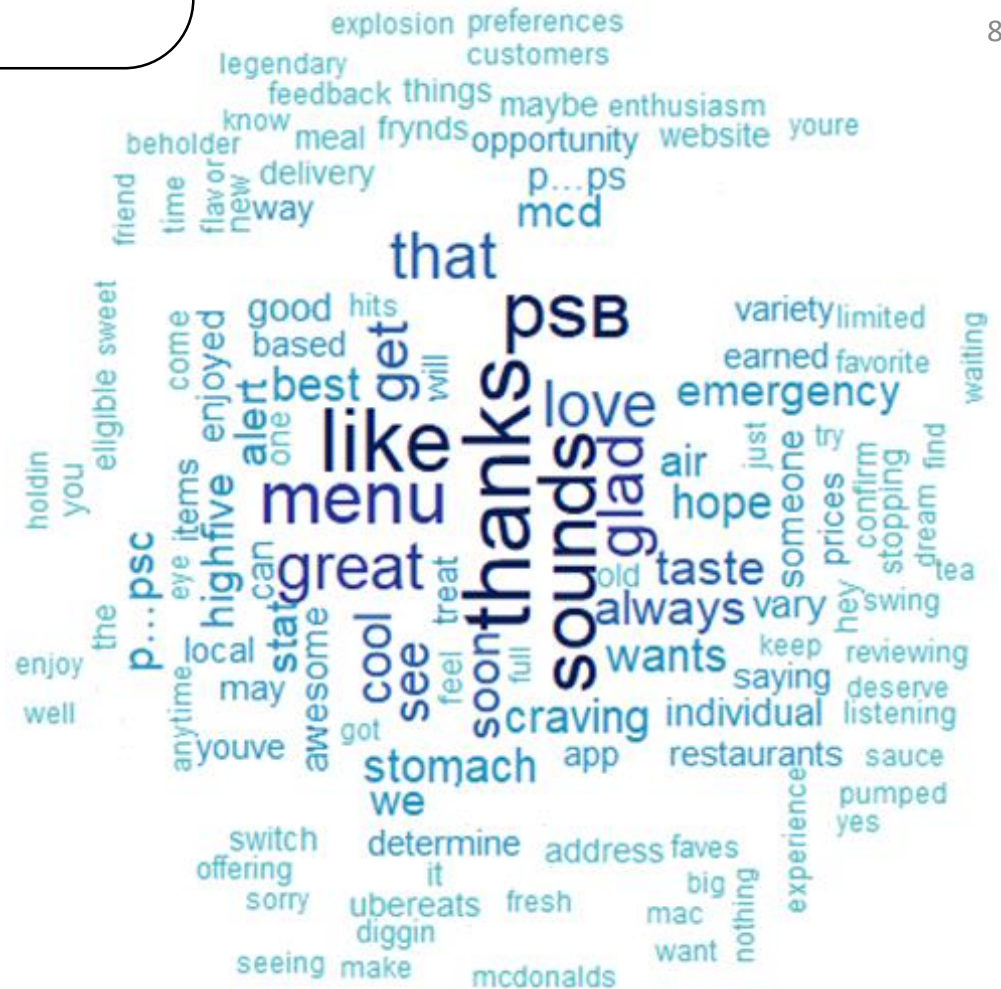
Word Cloud

$$s_i = \left[\frac{f_{max}(t_i - t_{min})}{t_{max} - t_{min}} \right]$$

для $t_i > t_{min}$; інакше $s_i = 1$

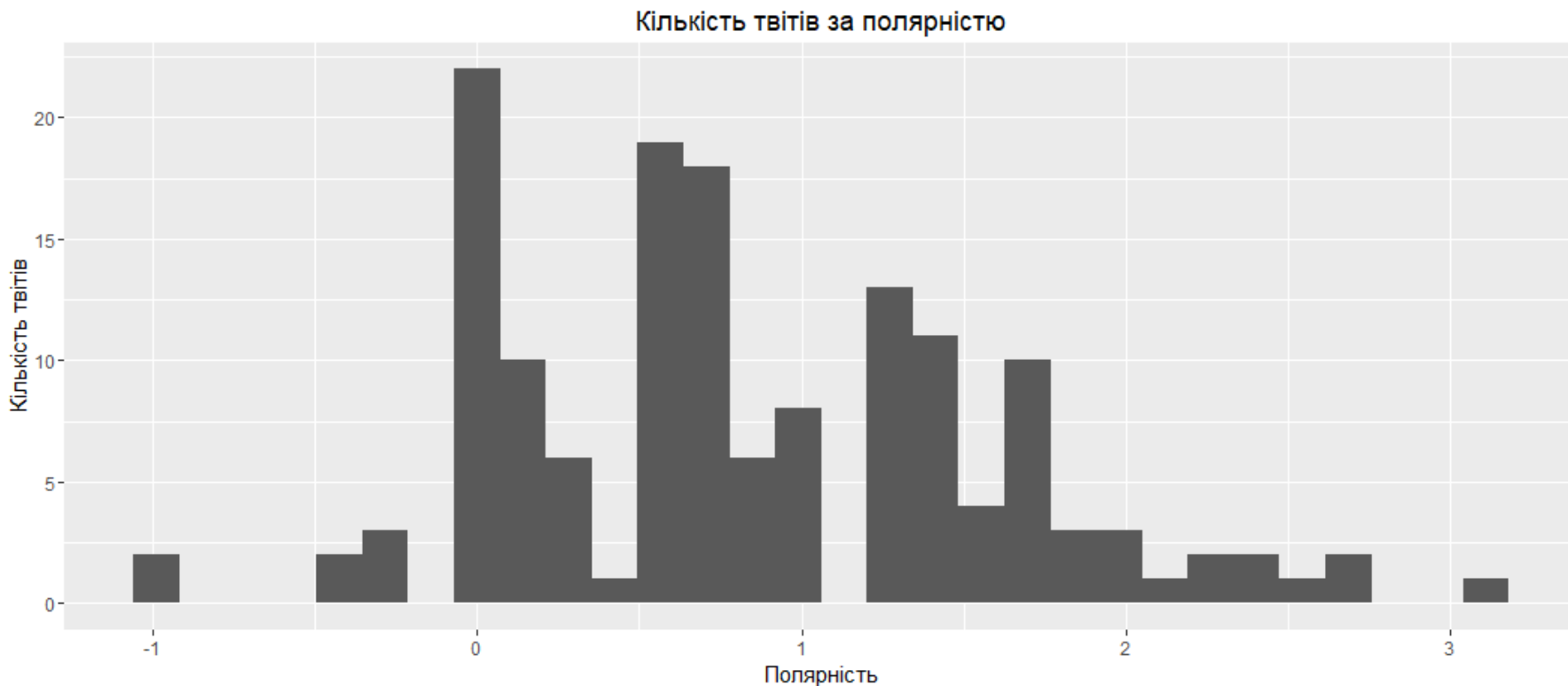
- s_i : розмір шрифту
- f_{max} : максимальний розмір шрифту
- t_i : вага тегу
- t_{min} : мінімальна вага
- t_{max} : максимальна вага

Для прикладу беремо
обліковий запис
“@McDonalds”.



Полярність

Для визначення полярності було прийнято взяти за основу словник розроблений в літературній лабораторії Небраски під керівництвом Метью Л. Джокерса. Перш за все оцінимо в якій полярності лежить більшість публікацій. На нашому прикладі бачимо, що більшість публікацій лежить за значенням 1, тобто позитивні.



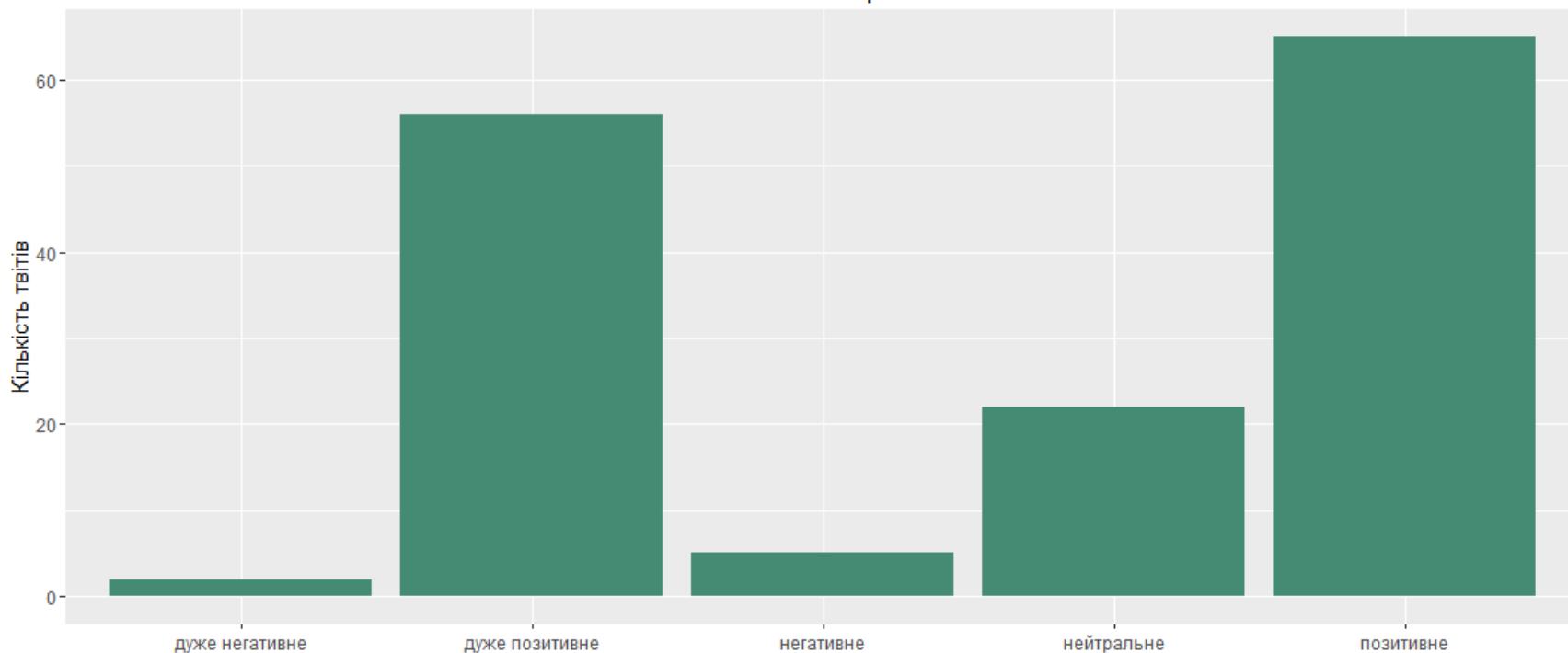
Полярність

Нехай x – оцінка полярності публікації.

$$f(x) = \begin{cases} \text{дуже позитивне,} & x > 1 \\ \text{позитивне,} & x \in (0; 1] \\ \text{нейтральне,} & x = 0 \\ \text{негативне,} & x \in [-1; 0) \\ \text{дуже негативне,} & x < -1 \end{cases}$$

10

Твіти за полярністю

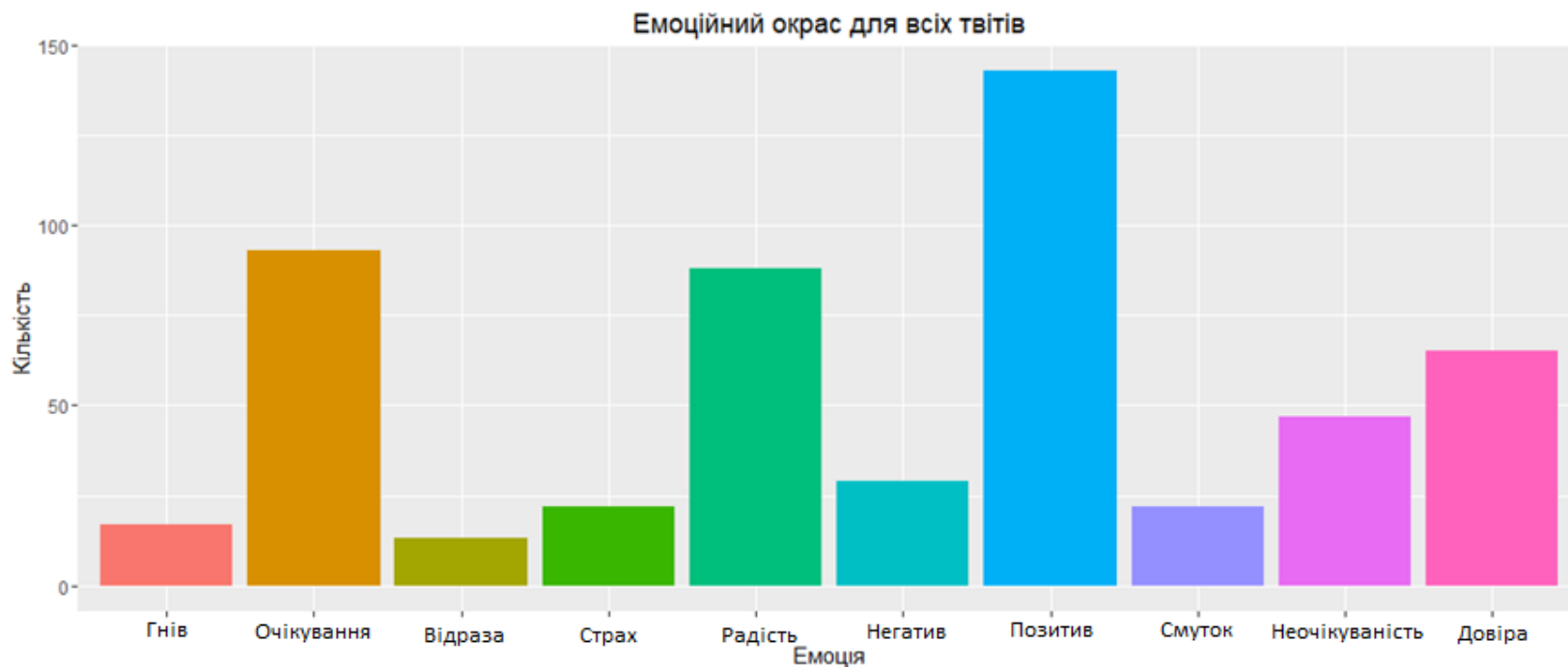


Емоційний окрас (Тональність)

11

Основні емоції до яких намагається звернутися “@McDonalds” є:

- Очікування
- Радість
- Довіра
- Неочікуваність



Кластерна дендрограма

Нехай задана навчальна вибірка: $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$.

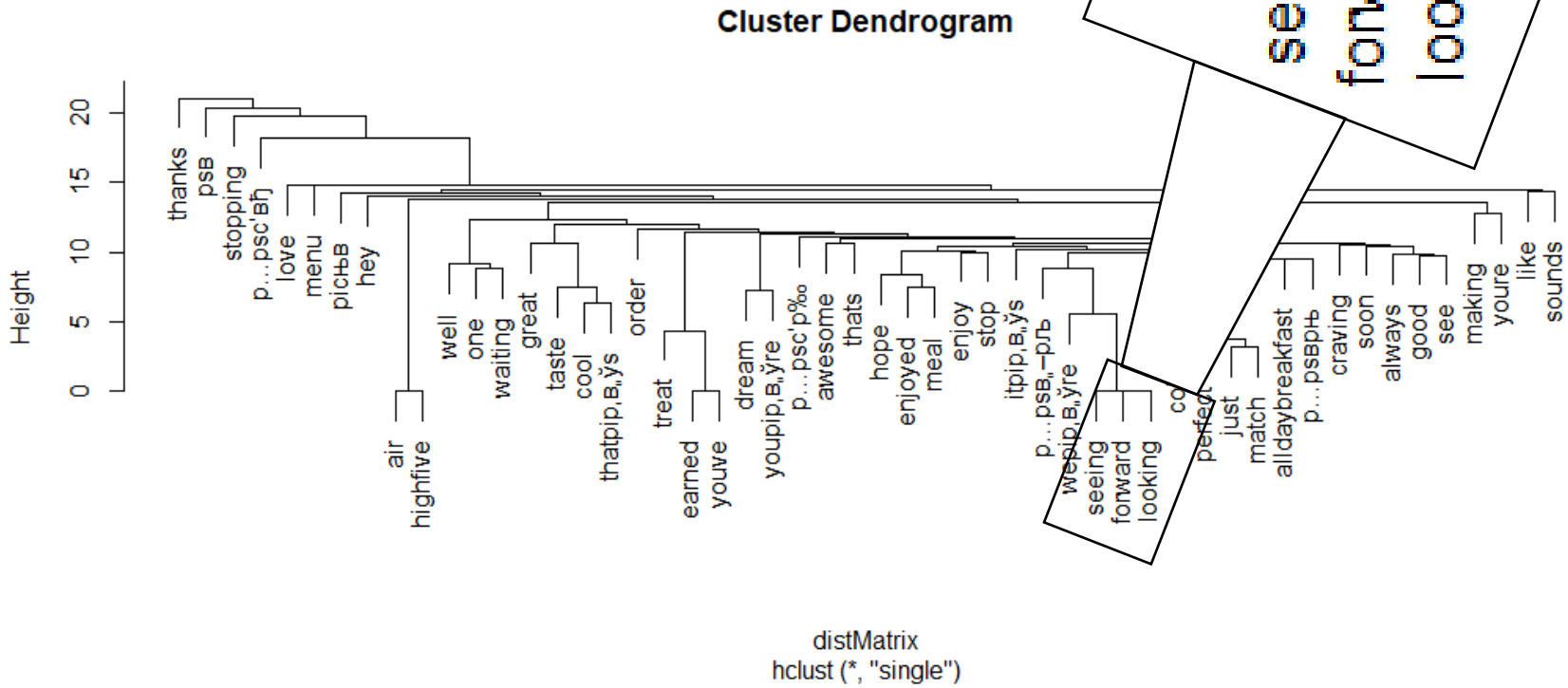
Нехай на множині об'єктів задана функція відстані $p(x, x')$. Для довільного об'єкту u розмістимо об'єкти навчальної вибірки x_i в порядку зростання u :

$$p(u, x_{1;u}) \leq p(u, x_{2;u}) \leq \dots \leq p(u, x_{m;u}),$$

де $x_{i;u}$ означає той об'єкт навчальної вибірки, який являється i -им сусідом об'єкта u . Аналогічне позначення введемо для відповіді на i -ому сусіді: $y_{i;u}$. Таким чином, будь-який об'єкт u породжує свою перенумерацію вибірки.

$$a(u) = \operatorname{argmax}_{y \in Y} \sum_{i=1}^m [x_{i;u} = y] w(i, u),$$

Кластерна дендрограма



Висновки

В рамках дипломної роботи були вирішені наступні завдання:

- Проведено огляд існуючих методів автоматичного аналізу емоційного забарвлення текстів.
- Проведено дослідження текстових особливостей повідомлень в соціальній мережі Twitter в контексті використання методів аналізу їх емоційного забарвлення.
- Розроблено програму, яка ідентифікує емоційного забарвлення повідомлень в соціальній мережі Твіттер.

Подальші кроки

15

- Додати можливість ідентифікації емоційного окрасу тексту написаного кирилицею.
- Зробити можливим використання в інших соціальних мережах.



ДЯКУЮ ЗА УВАГУ