

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

Міністерство освіти та науки України

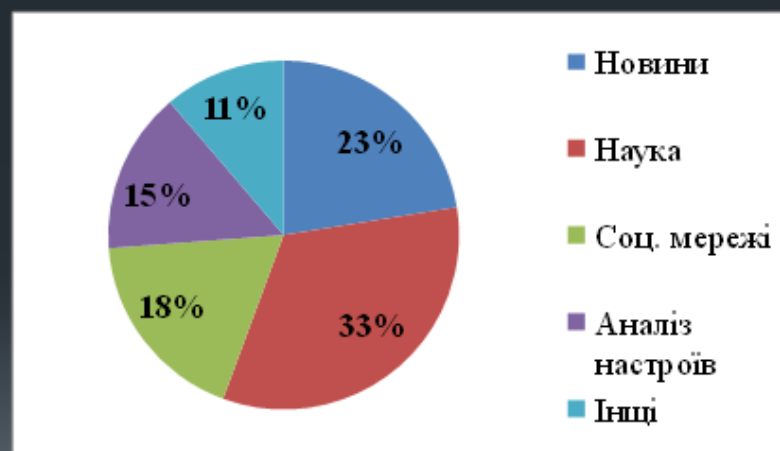
Харківський національний університет радіоелектроніки
Кафедра «Електронних обчислювальних машин»

Кваліфікаційна робота
на тему:

«Концептуальна модель двохступеневої класифікації тексту»

Виконав: ст. гр. КСМм-21-1 Настенко О.С.
Керівник доц. Барковська О.Ю.

Актуальність кваліфікаційної роботи



Огляд методів, що застосовуються при класифікації

3

Методи	Точність	Повнота	Обчислювальна складність алгоритму	Швидкість класифікації
Метод k-найближчих сусідів	78% - 86%	86% - 91%	висока	низька
Метод опорних векторів	63% - 90%	83% - 87%	низька	низька
Наївна байєсівська класифікація	40% - 83%	80% - 90%	низька	низька
«Випадковий ліс»	57% - 78%	75% - 82%	висока	висока
Згорткові нейронні мережі	83,98%	70% - 85%	висока	висока

Мета кваліфікаційної роботи

Метою роботи є створення концептуальної моделі двохступеневої класифікації тексту на прикладі стандартизованих навчальних та тестових наборів текстових даних.

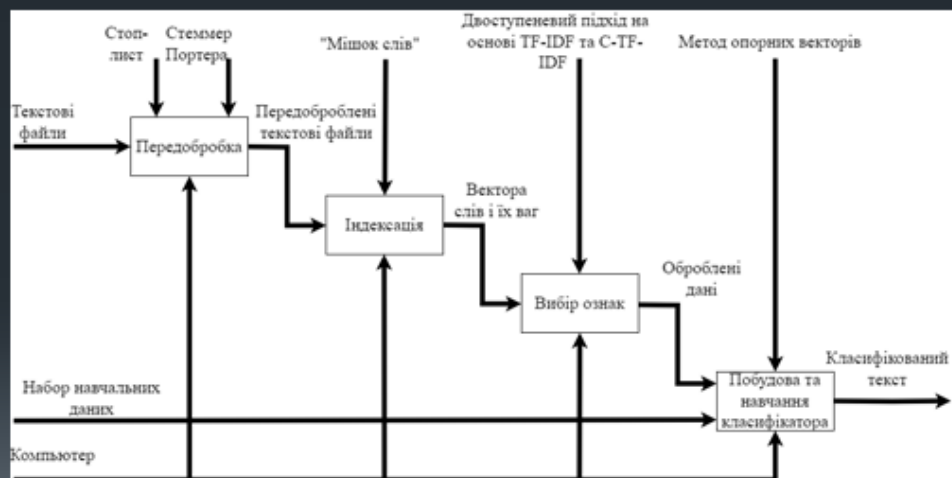
Задачі кваліфікаційної роботи

5

- огляд існуючих методів класифікації текстових даних;
- аналіз методів попередньої обробки та підготовки вхідних текстових даних;
- розробка концептуальної моделі двохступеневої класифікації тексту;
- виконання досліджень впливу скорочення вектору частот появи у тексті ключових слів на точність класифікації тексту. Дослідити вплив на текстових колекціях різного розміру та тематики;
- аналіз отриманих результатів.

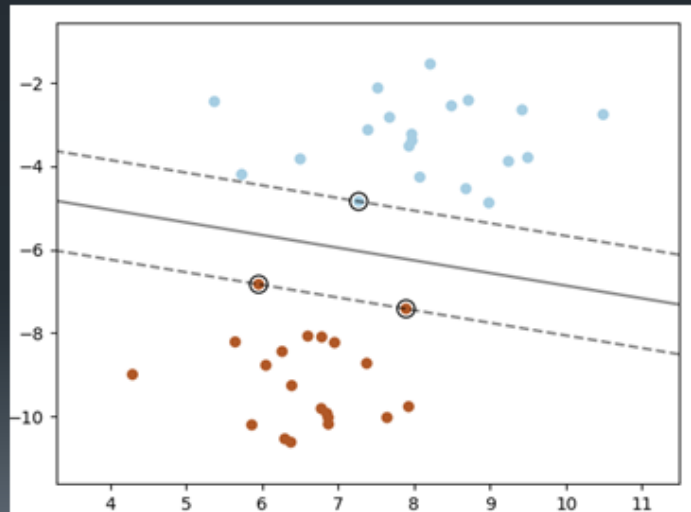
Рішення поставленої задачі. Концептуальна модель

6



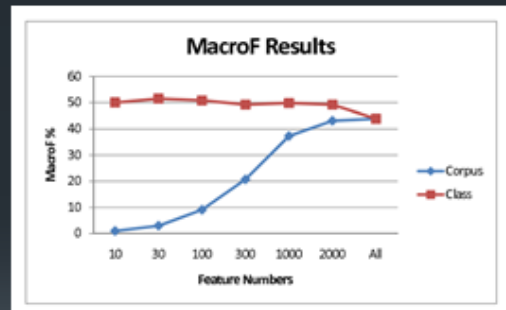
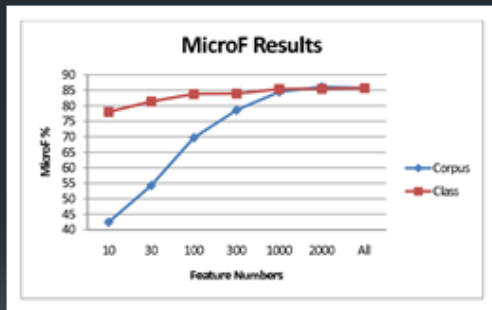
Рішення поставленої задачі. Метод опорних векрів

7



Аналіз отриманих результатів

8



Аналіз отриманих результатів

9

Метод, Параметр	Reuters		NSF		MiniNg20	
	MicroF	MacroF	MicroF	MacroF	MicroF	MacroF
AWPK, 13, 2000	86.40	53.95	66.06	50.11	57.43	55.66
AWPK, 13, 4000	86.70	53.98	66.10	50.12	57.43	55.66
AW	85.58	43.83	64.46	46.11	46.42	43.44
AWP, 13	85.84	44.85	64.58	46.49	53.62	51.02
AWK, 2000	85.58	52.03	65.19	49.31	54.04	52.10
AWK, 4000	85.84	52.10	65.71	49.35	55.25	53.73

Висновки

10

- В процесі роботи над кваліфікаційною роботою були досліджені методи класифікації текстів, а також можливі модифікації стандартних алгоритмів
- Проаналізовано зміни якості роботи при зміні конфігурацій алгоритмів.
- Описано концептуальну модель двоступеневої класифікації тексту та виконано програмну реалізацію даної моделі.

Апробація

Стаття у фаховому виданні:

Барковська О. Ю., Хомич В. М., Настенко О. С.
 Дослідження методів обробки та аналізу тексту при організації електронних сховищ інформаційних об'єктів. Сучасний стан наукових досліджень та технологій в промисловості. 2022. № 1 (19). С. 5-12. DOI: <https://doi.org/10.30837/ITSSI.2022.19.005>

Тези доповіді:

Настенко О. С., Барковська О. Ю. Обґрунтування вибору ознак для класифікації тексту. Проблеми інформатизації. 2022. № 2. С. 68. DOI: <https://doi.org/10.32620/PI.22.t2>

