

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

(інститут)
інформаційних технологій
(факультет)
Кафедра системного аналізу і управління
(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеню бакалавра
(бакалавра, спеціаліста, магістра)

студента Пономарьова Олександра Валентиновича
(ПІБ)

академічної групи 124-16-1
(шифр)

спеціальності 124 Системний аналіз
(код і назва спеціальності)

спеціалізації _____
за освітньо-професійною програмою Системний аналіз

(офіційна назва)

на тему « Дослідження впливу різноспрямованих новин на емоційний фон користувачів соціальної мережі»

(назва за наказом ректора)

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	<i>к.ф-м.н., доц. Коряшкіна Л.С.</i>			
розділів:				
<i>Інформаційно-аналітичний розділ</i>	<i>к.ф-м.н., доц. Коряшкіна Л.С.</i>			
<i>Спеціальний розділ</i>	<i>к.ф-м.н., доц. Коряшкіна Л.С.</i>			

Рецензент	<i>к.ф-м.н., проф. Ус С.А.</i>			
------------------	--------------------------------	--	--	--

Нормоконтролер	<i>к.т.н, асист. Малієнко А.В.</i>			
-----------------------	--	--	--	--

Дніпро
2020

Реферат

Пояснювальна записка: 50 сторінок, 6 таблиць, 8 рисунків, 5 додатків, 13 джерел.

Об'єкт дослідження: системи обробки та аналізу текстової інформації, Internet-сайт «КиноПоиск».

Предмет дослідження: методи класифікації повідомлень, які дозволяють аналізувати емоційне забарвлення текстів.

Мета дослідження: розробка методу класифікації текстових повідомлень, який дозволив би визначати тональність тексту й використовувався для фільтрації та виявлення важливої (з певної точки зору) інформації в соціальних мережах.

Методи дослідження та апаратура: методи машинного навчання, наївний Байєсовський класифікатор; мова програмування Python.

В *інформаційно-теоретичному розділі* наведена постановка задачі семантичного аналізу тексту; наведено стислий огляд існуючих досліджень, методів і програмних продуктів в області класифікації текстової інформації; аналізу тональності тексту описано основну ідею обраного методу, наведені характеристики сайту, дані з якого класифікуються, визначені проблеми і труднощі, що виникають на практиці під час розв'язання поставленої задачі, а також описані головні переваги обраної мови програмування.

У *спеціальному розділі* розроблений алгоритм автоматизованого збору інформації з сайту, підготовки даних для створення власного набору даних для навчання, описаний алгоритм Байєсовського класифікатора і програмна його реалізація. Наведені результати, що були отримані під час розв'язання поставленої задачі.

Практична цінність отриманих у роботі результатів полягає в тому, що запропонована методика збору інформації та розроблений алгоритм класифікації можуть бути застосовані для розв'язання широкого кола задач контент-аналізу, наприклад, для виявлення в соціальних мережах повідомлень

з інформацією про протиправні дії або ставлення населення до деякого об'єкта дослідження.

Ключові слова: АНАЛІЗ ДАНИХ, РОЗПІЗНАВАННЯ ТЕКСТУ, КОНТЕНТ-АНАЛІЗ, ТЕХТ МІНІНГ, АВТОМАТИЗАЦІЯ, ОБРОБКА ДАНИХ.