

Міністерство освіти і науки України

Національний університет харчових технологій

89

**Міжнародна наукова
конференція молодих учених,
аспірантів і студентів**

**"Наукові здобутки молоді –
вирішенню проблем
харчування людства у ХХІ
столітті"**

3-7 квітня 2023 р.

Частина 2

Київ НУХТ 2023

89 International scientific conference of young scientist and students "Youth scientific achievements to the 21st century nutrition problem solution", April, 3-7, 2023. Book of abstract. Part 2. NUFT, Kyiv.

The publication contains materials of 89 International scientific conference of young scientists and students "Youth scientific achievements to the 21st century Nutrition problem solution".

It was considered the problems of improving existing and creating new energy and resource saving technologies for food production based on modern physical and chemical methods, the use of unconventional raw materials, modern technological and energy saving equipment, improve of efficiency of the enterprises, and also the students research work results for improve quality training of future professionals of the food industry.

The publication is intended for young scientists and researchers who are engaged in definite problems in the food science and industry.

© NUFT, 2023

Матеріали 89 Міжнародної наукової конференції молодих учених, аспірантів і студентів "Наукові здобутки молоді – вирішенню проблем харчування людства у ХХІ столітті", 3-7 квітня 2023 р. – К.: НУХТ, 2023 р. – Ч.2. – 359 с.

Видання містить матеріали 89 Міжнародної наукової конференції молодих учених, аспірантів і студентів "Наукові здобутки молоді – вирішенню проблем харчування людства у ХХІ столітті".

Розглянуто проблеми удосконалення існуючих та створення нових енерго-та ресурсощадних технологій для виробництва харчових продуктів на основі сучасних фізико-хімічних методів, використання нетрадиційної сировини, новітнього технологічного та енергозберігаючого обладнання, підвищення ефективності діяльності підприємств, а також результати науково-дослідних робіт студентів з метою підвищення якості підготовки майбутніх фахівців харчової промисловості.

Розраховано на молодих науковців і дослідників, які займаються означеними проблемами у харчовій науці та промисловості.

© НУХТ, 2023

15. Проектування програмного засобу для автоматичного визначення мови ворожнечі

Катерина Дьогтяр, Микола Костіков

Київський національний університет ім. Тараса Шевченка, м. Київ, Україна

Вступ. У сучасному світі величезна кількість людей щодня спілкується в соціальних мережах і месенджерах. Тож актуальною є проблема виявлення мови ворожнечі (англ. *hate speech*), яка може міститись у текстових повідомленнях, публікаціях і коментарях в інтернеті.

Матеріали і методи. Джерелами аналізу в дослідженні є інтернет-публікації, застосовуються методи роботи з API, створення БД та комп'ютерної лінгвістики.

Результати. Метою дослідження є створення програмного засобу, який би міг автоматично визначати мову ворожнечі в текстах українською мовою.

Першим етапом на цьому шляху є розроблення та наповнення корпусу текстів із мережі інтернет. Як джерело можна взяти соціальні мережі або месенджери. Зокрема в ході дослідження було вирішено зібрати публікації з одного з Telegram-каналів. Це можна реалізувати як вручну, так і автоматично з використанням Telegram API [1].

Другим етапом є паспортизація отриманого корпусу текстів та збереження отриманого результату як бази даних для подальшого зручного програмного опрацювання.

На третьому етапі має бути створено власне програмний засіб, який визначає в поданих текстах обценну лексику. Слід зазначити, що, на перший погляд, така лексика має свідчити про наявність мови ворожнечі в повідомленні, однак насправді це не завжди так. Як показує аналіз інтернет-повідомлень, обценна лексика може вживатися з різною метою: для створення гумористичного ефекту; як вставні слова та вризи; для підсилення висловлення або навіть для вираження позитивних емоцій.

Отже, наступним і ключовим етапом роботи має бути фільтрування отриманих повідомлень із обценною лексикою, та відсіювання тих, де мови ворожнечі немає.

Одним із можливих підходів до цього є створення автоматизованої системи, яка лише робить вибірку текстів із обценною лексикою та передає її на розгляд людині-модератору. В такому разі саме ця особа робитиме остаточний висновок щодо мови ворожнечі та позначатиме проаналізовані повідомлення відповідним чином.

Іншим підходом може бути спроба реалізації автоматичного тонального аналізу (сентимент-аналізу). Для цього можна використати методи машинного навчання, зокрема бібліотеку NLTK для мови програмування Python [2]. Як тренувальний набір у нашому випадку буде взято повідомлення, розмічені вручну за показником наявності або відсутності мови ворожнечі (як у першому підході). На тестовому ж наборі наша програма вже самостійно визначатиме, де мова ворожнечі є, а де — ні.

Висновки. Перспективою подальшого дослідження є програмна реалізація й тестування проєктованого засобу.

Література

1. Костіков М. П. (2022) Використання Telegram API для збору та опрацювання текстової інформації, *Матер. 88 міжнар. наук. конф. молодих учених, аспірантів і студ. «Наукові здобутки молоді — вирішенню проблем харчування людства у XXI столітті»*, квітень-травень 2022 р., К.: НУХТ, с. 296.
2. Bird S., Klein E., Loper E. (2023) *Natural Language Processing with Python* [online]. URL: <https://www.nltk.org/book>.