УДК 681.12

О. О. Шинкаренко

РОЗРОБКА ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ АНАЛІЗУ ТЕКСТУ НА ВМІСТ ОБРАЗЛИВИХ ВИСЛОВЛЮВАНЬ

Вінницький національний технічний університет;

Анотація

Запропоновано систему із веб-інтерфейсом для аналізу тексту на наявність образливих, неприйнятних висловів, нецензурної лексики за декількома критеріями, яка допоможе пришвидшити процес модерації коментарів, повідомлень користувачів у соціальних мережах, форумах, комп'ютерних іграх тощо. Також дана система допоможе акцентувати увагу робітників-модераторів на розгляд більш спірних випадків вживання образливих висловлювань.

Ключові слова: веб-інтерфейс, образливі висловлювання, нецензурна лексика, модерація.

Abstract

Suggested a system with a web interface for analyzing text for the presence of offensive, unacceptable utterances, obscene language on several criteria, which will help speed up the process of moderation of comments, user messages on social networks, forums, computer games, etc. This system will also help to focus the attention of moderators on the consideration of more controversial cases of usage of abusive language.

Keywords: web interface, offensive language, obscene language, moderation.

Вступ

Сьогодні більшість онлайн-ресурсів, таких як платформи розповсюдження новин, платформи відповідей на запитання та обміну досвідом, соціальні мережі, форуми, а також онлайн-ігри включають в себе простір для спілкування, висловлення думок. З ціллю збереження поважного ставлення до співрозмовника та плідного обговорення тем, компанії наймають контент-модераторів, які слідкують за тим, аби правила обговорення платформи виконувались. У разі невиконання таких правил, модератори видаляють дописи користувачів та, за необхідності, блокують і самого користувача. Відповідно, модератори повинні правильно класифікувати та визначити причину видалення коментаря та повідомити про неї автора. Така робота є рутинною та вимагає великих часових затрат на прийняття рішень, а зі збільшенням кількості коментарів для розгляду виникає потреба у наймі більшої кількості працівників[1, 2]. Процес можливо пришвидшити та спростити, якщо відсортовувати коментарі по ступеню ймовірності порушення правил спілкування платформи автоматизовано та пропонувати на розгляд модераторів у пріоритеті більш спірні випадки порушення правил.

На ринку існують розроблені рішення даної проблеми. Першим прикладом є система BattlEye, що спеціалізується на модерації спілкування в комп'ютерних іграх та слідкує за використанням нелегального програмного забезпечення з ціллю отримання переваги у грі. Система працює у реальному часі та є автоматизованою. Головним недоліком даної системи є автоматичне блокування підозрілих дій та сумнівних коментарів, тому для оскарження такого рішення необхідно писати звернення у службу підтримки, де кожен з випадків розглядається окремо. Іншим прикладом є Ethical AI, що спеціалізується на створенні штучного інтелекту для модерації під потреби замовника. Головним недоліком даної системи є необхідність розробки нового рішення під кожен індивідуальний випадок

Тому метою даної роботи є створення системи із веб-інтерфейсом для автоматизованого аналізу тексту на елементи, що порушують правила платформи, та пріоритетного пропонування на розгляд працівникам-модераторам на розгляд спірних, неоднозначних випадків порушення. Задачею роботи є створення програмного інтерфейсу з використанням існуючих перевірених нейронних мереж із можливістю огляду, блокування відповідного допису, що подається на вхід системі.

Результати дослідження

Враховуючи потреби ринку, була поставлена задача створення web-інтерфейсу системи аналізу коментарів на вміст образливих висловлювань за декількома критеріями ідентифікації, з можливістю

видалити проаналізований коментар, а також за необхідності написати повідомлення або заблокувати користувача, який залишив даний коментар.

Для виконання поставленої задачі розробки системи аналізу тексту на вміст образливих висловлювань, було використано мову програмування JavaScript, бібліотеку React та фреймворк Express.js. JavaScript є єдиною мовою програмування для створення веб-додатків, а також може виконуватись і на сервері за допомогою використання платформи Node.js, фреймворком для роботи з якою є Express.js. Бібліотека React забезпечує оновлення інтерфейсу без перезавантаження веб-сторінки, тобто дозволяє створити односторінковий додаток(SPA). Таким чином, забезпечується швидкість створення системи за рахунок використання єдиної мови програмування як для веб-інтерфейсу, так і для програмного інтерфейсу на сервері.

Висновки

В результаті роботи було створено інтелектуальну систему аналізу тексту на вміст образливих висловлювань. Встановлено, що запропонована система дозволяє пришвидшити процес модерації вмісту коментарів та дописів, а також зменшує кількість дописів, для яких існує необхідність ручної модерації працівниками, так як основна робота аналізу проводиться автоматизовано.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. J. Risch, R. Ruff, R. Krestel. Offensive Language Detection Explained, 2020. — 137c.

2. R. Pradhan, A. Chaturvedi, A. Tripathi, D.K. Sharma. A Review on Offensive Language Detection. URL: https://www.researchgate.net/publication/338355806_A_Review_on_Offensive_Language_Detection.

Шинкаренко Олег Олександрович — студент групи 1КН-17б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: oshynkarenko1503@gmail.com

Науковий керівник: ***Сілагін Олексій Віталійович*** — канд. техн. наук, доцент кафедри КН, Вінницький національний технічний університет, м. Вінниця

Shynkarenko Oleh O. — Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email : oshynkarenko1503@gmail.com

Supervisor: ***Silagin Oleksii V.*** — Cand. Sc. (Eng), Professor of Computer Science, Vinnytsia National Technical University, Vinnytsia