

ОБҐРУНТУВАННЯ ДОЦІЛЬНОСТІ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ АВТОМАТИЧНОГО ФІЛЬТРУВАННЯ ОБРАЗЛИВИХ ПОВІДОМЛЕНЬ У РОЗРОБЦІ ВЕБ-ЧАТІВ

Вінницький національний технічний університет

Анотація

Впровадження технологій автоматичного фільтрування образливого контенту в чат-платформи має великий потенціал для покращення онлайн-комунікації та забезпечення безпеки користувачів. Сучасні онлайн-чати часто стикаються з проблемою неприємних повідомлень, таких як образи, агресія або ненормативна лексика, що може створювати негативну атмосферу та впливати на досвід користувачів. Для вирішення цієї проблеми пропонується інтеграція автоматичної системи фільтрації образливого контенту в чат-платформи, яка дозволить виявляти і блокувати небажані повідомлення в реальному часі. Це забезпечить безпечніше та комфортніше середовище для користувачів, підвищить якість комунікації та зменшить кількість випадків кібербулінгу.

Ключові слова: автоматичне фільтрування образливого контенту, чат-платформи, безпека онлайн-комунікації, інтерактивність, кібербулінг.

Abstract

The implementation of automatic filtering technology for offensive content in chat platforms has great potential to improve online communication and ensure user safety. Modern online chats often face the problem of unpleasant messages, such as insults, aggression, or profanity, which can create a negative atmosphere and affect the user experience. To address this issue, the integration of an automatic offensive content filtering system in chat platforms is proposed, which will allow identifying and blocking unwanted messages in real-time. This will provide a safer and more comfortable environment for users, improve communication quality, and reduce instances of cyberbullying.

Keywords: voice recognition, Snake game, voice control, accessibility of the game, interactivity of the game.

Вступ

Забезпечення безпеки користувачів у чат-платформах є важливим аспектом для створення здорового онлайн-середовища. У світлі зростаючої популярності онлайн-комунікацій, проблема агресивного контенту, такого як образи, агресивні висловлювання та інші негативні повідомлення, стає все більш актуальною. Такий контент може призводити до конфліктів, неприємних ситуацій і навіть кібербулінгу, що негативно впливає на досвід користувачів. Тому інтеграція технології автоматичного фільтрування образливих повідомлень є важливим кроком для покращення якості онлайн-комунікацій [1].

Технології автоматичного фільтрування дозволяють здійснювати перевірку повідомлень у реальному часі, що значно знижує навантаження на модераторів і підвищує ефективність боротьби з образливим контентом [2]. Це дає змогу виявляти образи, агресивну мову та інші небажані елементи, автоматично блокуючи їх і забезпечуючи безпечне середовище для користувачів. Впровадження таких технологій сприяє не лише покращенню безпеки, але й зменшенню соціальних напруг в онлайн-спільнотах.

Враховуючи вищесказане, стає цілком очевидно, що впровадження автоматичних фільтрів образливого контенту є необхідним етапом розвитку сучасних чат-систем, що сприятиме більш комфортному та безпечному спілкуванню для всіх користувачів.

Результати дослідження

На сьогоднішній день існує досить багато популярних месенджерів, але найбільше з них використовуються Telegram, WhatsApp, Discord, Slack та інші, які здобули величезну аудиторію завдяки своїй доступності та різноманіттю функцій. Проте, незважаючи на їхню популярність, ці платформи мають певні недоліки, зокрема відсутність ефективної системи автоматичного фільтрування так званого токсичного контенту та недостатній рівень безпеки у деяких випадках [3 –

7]. Водночас, впровадження технології автоматичного фільтрування образливих повідомлень може стати дуже ефективним рішенням для подолання цих проблем.

У табл. 1 представлена порівняльна характеристика цих месенджерів.

Таблиця 1 – Порівняльна характеристика деяких найпопулярніших месенджерів

Характеристика	Telegram	WhatsApp	Discord	Slack
Мультимедійність	+	+	+	+
Реєстрація та авторизація	+	+	+	+
Пошукова система	+	+	+	+
Типи чатів	Приватні, групові канали	Приватні, групові	Приватні, групові, голосові	Приватні, групові
Використання сучасних веб-технологій у дизайні	+	–	+	+
Шифрування повідомлень	End-to-End	End-to-End	TLS	TLS
Автоматична фільтрація токсичного контенту	–	–	–	–

На основі аналізу деяких характеристик представлених месенджерів можна зробити висновок, що Telegram, WhatsApp, Discord та Slack мають певні недоліки:

- відсутність вбудованих інструментів для автоматичного виявлення та блокування токсичних повідомлень, що може негативно впливати на досвід користувачів;
- хоча WhatsApp та Telegram підтримує End-to-End шифрування для всіх чатів, у Slack та Discord, наприклад, використовуються тільки TLS-з'єднання, що є менш безпечним порівняно з End-to-End шифруванням;
- у WhatsApp спостерігається слабка підтримка сучасних веб-технологій, що знижує адаптивність до нових потреб користувачів.

Враховуючи вищезазначене, можна зробити висновок, що для покращення досвіду користувачів та підвищення рівня безпеки у месенджерах, доцільно впроваджувати автоматичну фільтрацію образливого контенту.

Висновки

За результатами аналізу предметної області та сучасного стану розвитку інформаційних технологій було визначено, що розробка веб-чату з автоматичним фільтруванням контенту є досить важливим та актуальним завданням. Запропонований чат поєднує сильні сторони існуючих платформ-аналогів та доповнюється системою розпізнавання та блокування токсичних повідомлень. Інтуїтивно зрозумілий інтерфейс та гнучкі налаштування модерації забезпечують комфортне спілкування, підвищують рівень безпеки та надають веб-чату значні переваги порівняно з іншими рішеннями.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Кулиняк Р. В., Каштелян І. В. Алгоритм визначення «токсичності» повідомлення з використанням засобів NLP. III Науково-практична конференція молодих вчених і студентів «Інтелектуальні комп'ютерні системи та мережі». 26 листопада 2020, Тернопіль, Україна. Тернопіль: ЗУНУ, 2020. – С. 23
2. Бондарчук В. Ю., Арсенюк І. Р. Обробка та аналіз природної мови методами машинного навчання. // Тези доповідей XLIX науково-технічної конференції факультету інформаційних технологій та комп'ютерної інженерії. – Вінниця: ВНТУ, 2020. Електронний ресурс (режим доступу) <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2020/paper/view/9161/7764>
3. Telegram. URL: <https://web.telegram.org/a/> (дата звернення 10.03.2025).
4. WhatsApp. URL: <https://www.whatsapp.com/?lang=uk> (дата звернення 10.03.2025).
5. Discord. URL: <https://discord.com/> (дата звернення 10.03.2025).

6. Slack. URL: <https://slack.com/> (дата звернення 10.03.2025).
7. Собко О. В., Бармак О. В. Виявлення кіберзалякувань в інформаційному середовищі засобами машинного навчання. Інформаційна, функційна і кібербезпека СКІФіК2024 : матеріали четвертої науково-технічної конференції, 29-30 листопада 2024. Харків, 2024. С. 96 – 97.

Гнідий Костянтин Вікторович – студент групи ЗКН–21б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: gnkostia3@gmail.com .

Арсенюк Ігор Ростиславович – доцент кафедри комп’ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: air@vntu.edu.ua.

Hnidy Kostiantyn Viktorovich – student of group ЗКС–21b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: gnkostia3@gmail.com.

Arseniuk Ihor Rostyslavovych – Associate Professor, Department of Computer Science, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, Ukraine, e-mail: air@vntu.edu.ua.