

РЕФЕРАТ

Актуальність. З розповсюдженням Інтернету та соціальних медіа зараз доступна кількість новин, статей та іншого тексту онлайн. Цей величезний обсяг інформації постав під загрозу правдивість новин, які поширюються.

Підроблені новини - це будь-яка форма помилкової інформації чи контенту, що поширюється в мережі Інтернет, для впливу на погляд людей на певну подію чи інформацію. Виявлення фальшивих новин у цифровому світі є важливим завданням у подоланні широкого розповсюдження чуток та упереджень. Багато досліджень було проведено для виявлення елементів дезінформації для англійської мови, проте українська та російська мови не має досліджень у цій галузі. Такі компанії, як Facebook, Twitter та Google, стикаються з проблемою вирішення цієї проблеми, щоб забезпечити платформу, де люди можна довіряти вмісту стрічки новин. Вплив фейкових новин було таким глибоко вкорінене в суспільстві, що це навіть вплинуло на вибори в США 2016 року. Також багато неправдивої інформації поширюється протягом війни України в зоні АТО, що призводить до дестабілізації населення, поширення неправильних думок, відображення фейкової картини перебігу подій.

Отже, необхідною задачею є створення інструменту перевірки текстової інформації на наявність елементів дезінформації для інформаційної безпеки та аналізу новин, які поширюються для дестабілізації та обману населення.

Метою дослідження є полегшення виявлення елементів дезінформації за рахунок створення методу та алгоритму для перевірки потоку текстових даних на наявність елементів дезінформації у вигляді лінгвістичних конструкцій та оборотів, які вказують на неправдивість представленої інформації.

Для досягнення поставленої мети необхідно виконати наступні завдання:

- виконати аналіз існуючих алгоритмів та методів комп'ютерної лінгвістики та машинного навчання для класифікації текстових потоків даних та виявлення елементів дезінформації;
- розробити алгоритм первинної обробки тексту для збільшення точності визначення елементів дезінформації;
- розробити метод виявлення елементів дезінформації в текстових потоках даних;
- виконати програмну реалізацію розробленого методу виявлення елементів дезінформації в текстових потоках даних;
- провести аналіз отриманих результатів для оцінки якості;
- провести дослідження ефективності алгоритму.

Предметом дослідження є методи виявлення елементів дезінформації в текстових потоках даних.

Методами дослідження є методи комп'ютерної лінгвістики та машинного навчання для виявлення елементів дезінформації.

Наукова новизна отриманих результатів: науковою новизною є розробка методу виявлення елементів дезінформації в потоках даних з підтримкою обробки текстів української та російської мови.

Публікації: основні положення роботи доповідались і обговорювались на IV всеукраїнській науково-практичній конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2020), а також подано до друку на XVI міжнародній науковій конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMCI'2020)» результати магістерської дисертації докладались на наукових конференціях.

Ключові слова: МАШИННЕ НАВЧАННЯ, КЛАСИФІКАЦІЯ, ДЕЗІНФОРМАЦІЯ, ПОТОКИ ТЕКСТОВИХ ДАНИХ