

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти  
(повна назва)

Кафедра Програмної інженерії  
(повна назва)

**АТЕСТАЦІЙНА РОБОТА**

**Пояснювальна записка**

другий (магістерський)

(рівень вищої освіти)

Дослідження методів та засобів розпізнавання фейкових новин в соціальних мережах  
(тема)

Виконав: студент 2 курсу, групи ПЗмзд-17-1  
спеціальності 121- Інженерія програмного забезпечення  
(код і повна назва спеціальності)

Освітньо-наукової програми  
Інженерія програмного забезпечення  
(повна назва освітньої програми)

Башмаков Є.Б.

(прізвище, ініціали)

Керівник проф. Єрохін А.Л.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри, проф. \_\_\_\_\_

З.В.Дудар

2019 р.

Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти

Кафедра Програмної інженерії

Рівень вищої освіти другий (магістерський)

Спеціальність 121-Інженерія програмного забезпечення

(код і повна назва)

освітньо-наукова програма Інженерія програмного забезпечення

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

«\_\_\_» \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ  
НА АТЕСТАЦІЙНУ РОБОТУ**

студентові Башмакову Євгену Борисовичу

(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів та засобів розпізнавання фейкових новин в соціальних мережах

затверджена наказом по університету від “\_” \_\_\_\_\_ 2019 р № \_\_\_\_\_

2. Термін подання студентом роботи до екзаменаційної комісії  
26 червня 2019 р.

3. Вихідні дані до роботи автоматичні методи розпізнавання фейкових новин, таблиця порівняння засобів розпізнавання, аналіз методів створення індексу достовірності інформації.

4. Перелік питань, що потрібно опрацювати в роботі мета роботи, аналіз проблемної галузі і постановка задачі, огляд засобів розпізнавання фейкових новин у соцмережах, аналіз автоматичних методів розпізнавання фейкових новин

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) Мета завдання, обґрунтування актуальності роботи, постановка задачі, методи й алгоритми, таблиця порівнянь між засобами та методами, схема автоматичного процесу

6 Консультанти розділів роботи

| Найменування розділу | Консультант<br>(посада, прізвище, ім'я, по батькові) | Позначка консультанта про виконання розділу |      |
|----------------------|--|---|------|
|                      |  | підпис                                      | дата |
| Спецчастина          | проф. Єрохін А.Л.                                    |   |      |

### КАЛЕНДАРНИЙ ПЛАН

| №   | Назва етапів роботи                   | Терміни виконання етапів роботи | Примітка* |
|-----|---------------------------------------|---------------------------------|-----------|
| 1.  | Аналіз предметної галузі              | 19 квітня 2019р.                |           |
| 2.  | Огляд існуючих методів                | 27 квітня 2019р.                |           |
| 3.  | Методи розв'язання задачі             | 10 травня 2019р.                |           |
| 4.  | Підготовка пояснювальної записки      | 25 травня 2019р.                |           |
| 5.  | Спецчастина                           | 26 травня 2019р.                |           |
| 6.  | Підготовка презентації та доповіді    | 28 травня 2019р.                |           |
| 7.  | Попередній захист                     | 30 травня 2019р.                |           |
| 8.  | Нормоконтроль, рецензування           | 02 червня 2019р.                |           |
| 9.  | Занесення диплома в електронний архів | 03 червня 2019р.                |           |
| 10. | Допуск до захисту у зав. Кафедри      | 04 червня 2019р.                |           |

Дата видачі завдання 11 лютого 2019 р.

Студент \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_ проф. Єрохін А.Л.  
(підпис) (посада, прізвище, ініціали)

## РЕФЕРАТ / ABSTRACT

Атестаційна робота магістра містить: 64 с., 16 рис., 2 табл., 10 джерел, 1 додаток.

АНАЛІЗ ПУБЛІКАЦІЙ, ФЕЙК, ПЕРЕВІРКА ФАКТІВ, ДЕЗІНФОРМАЦІЯ, СОЦІАЛЬНІ МЕРЕЖІ, ОЦІНЮВАННЯ.

Метою роботи є виявлення проблемних місць та дослідження методів автоматичного розпізнавання фейкових новин.

У результаті роботи було досліджено суть і походження поняття «фейк», описано процес створення та поширення фейкової інформації у соціальних медіа. Було наведено типові приклади фейкової інформації в загальному медіа-просторі, з'ясовано вплив таких повідомлень на користувачів соціальних мереж.

PUBLICATIONS ANALYSIS, FAKE, FACT-CHECKING MISINFORMATION, SOCIAL NETWORK, SCORING.

The aim of this work is to identify problem areas and research methods for automatic recognition of fake news.

As a result of the work, the essence and origin of the concept of “fake” was explored, the process of creation and distribution of fake information in social media was described. The typical examples of fake information in the general media space were given, the impact of such messages on users of social networks was clarified.

## ЗМІСТ

|   |    |
|---|----|
| Вступ.....  | 6  |
| 1 Аналіз проблемної галузі та постановка задачі.....                        | 8  |
| 1.1 Проблема поширення фейкових новин та їхній вплив.....                   | 8  |
| 1.2 Аналіз схожих інструментів щодо визначення інноваційності метода.....   | 12 |
| 1.3 Постановка задачі.....  | 21 |
| 2 Аналіз методів розпізнавання фейкових новин .....                         | 22 |
| 2.1 Методи розпізнавання фейкових новин у соціальних мережах .....          | 22 |
| 2.2 Метод сильної евристики .....   | 25 |
| 3 Опис обраного метода .....  | 31 |
| 3.1 Метод розпізнавання фейкових новин на основі зібраних знань.....        | 31 |
| 3.1.1 Ручна перевірка фактів .....  | 31 |
| 3.1.2 Автоматична перевірка фактів.....                                     | 35 |
| 3.1.3 Open Web та збір знань.....   | 37 |
| 3.1.4 Завдання з обробки зібраних знань .....                               | 38 |
| 3.1.5 Побудова бази знань .....   | 40 |
| 3.1.6 Перевірка фактів .....  | 42 |
| 3.2 Метод розпізнавання фейкових новин на основі їхнього розповсюдження. 44 |    |
| 3.2.1 Шаблони розповсюдження фейкових новин.....                            | 47 |
| 3.3 Метод розпізнавання фейкових новин на основі стилю написання .....      | 49 |
| 3.3.1 Аналіз і виявлення обману .....                                       | 50 |
| 3.3.2 Функції мови на основі атрибутів .....                                | 51 |
| 3.3.3 Функції мови на основі структури .....                                | 52 |
| Висновки .....  | 55 |
| Перелік джерел посилання .....  | 56 |
| Додаток А.....  | 57 |

## ВСТУП

Сьогодні важко не погодитись, що фейкові новини є дуже серйозною проблемою. Ми всі хочемо бачити точні і правдиві звіти, і деякі з нас люб'язно ставляться до того, щоб бути навмисно введеними в оману.

Проблема полягає в тому, що не існує абсолютного визначення що являє собою фейкова новина. Недостовірні новини можуть мати різні форми. У самому екстремальному випадку у інтернеті купа історій, що Папа Римський виступає на підтримку Дональда Трампа або чутки, що поширювалися напередодні голосування Брексіт, вказуючи на те, що Європейський Союз мав секретний план імплантувати всіх новонароджених дітей мікрочіпами. Додайте до цього зростання вітчизняної реклами і контент-маркетингу, кожен з яких породив певну кількість комерційно-керованої неправди.

Як користувачі та споживачі, ми можемо бути, до деякої міри, співучасниками стосовно поширенню фейкової інформації тоді коли ми, швидше за все, віримо історіям, які співпадають з нашими особистими соціальними та політичними поглядами. І неправда однієї людини може бути альтернативним фактом іншої. Або ми можемо просто обдурити. Деякі історії, будучи повністю сфабрикованими, врешті були настільки добре підготовлені і представлені, що вони здаються цілком правдивими.

Але як споживачі ми повинні мати право знати, чи має історія, або думка, насправді реальну основу. Отже, потрібні певні засоби для оцінки того, чи базується певний пункт новин на фактах або ж на спекуляціях та неправдах.

Фейк – це різновид дезінформації, навмисне подання неправдивої інформації. Фейк можна сприймати як жарт – навмисне перекручення фактів, аби викликати позитивні емоції, сміх тощо, проте все частіше фейк перетворюється на справжню дезінформацію, яка вводить в оману.

Термін “фейк” має багато значень. В найширшому значенні це будь-яка підробка, яку хтось намагається видати за справжнє. До прикладу, всі відомі знімки НЛО вважаються фейками, проте не завжди це можливо довести. І хоча термін “фейк” в українській мові з’явився зовсім недавно, таке явище існувало й раніше і називалося “качкою”. Щоправда, тоді поширювати “качки” було прерогативою журналістів, зазвичай газетярів (жовтої преси, відповідно). Зараз же запустити таку ж “качку”, тільки у Інтернеті, може практично кожен. Тим фейки небезпечні і важко піддаються перевірці.

Фейком може бути акаунт у соцмережі, сайт в Інтернеті, одяг (підробки під відомі бренди) чи аксесуари, ліки, продукти в магазині (найчастіше – дорогі алкогольні напої) тощо. Проте найстрашніший і найбільш загрозливий вид фейку – новини в ЗМІ.

Кожен фейк, що потрапляє в ефір телебачення чи радіо, в Інтернет чи на сторінки газети – розрахований на певну аудиторію: на людей, поділених територіально чи то за політичними поглядами. У кожної аудиторії – свої очікування. Комуś треба лише довести те, що він думав раніше, підживити попередні враження.

Час від часу популярні та авторитетні, на перший погляд, ЗМІ поширюють інформацію із неправдивих чи неперевіраних джерел або й самостійно створюють та поширюють свої неправдиві новини. На думку деяких журналістів, такі дії видань та інформаційних агенцій покликані підвищити відвідуваність їхніх сайтів, “накрутити” рейтинги, відволікти увагу масового читача від важливіших новин та вплинути на репутацію політичних діячів [1].

Як можна зазначити, дуже багато часу витрачається на ручну роботу, коли більшу частину цих дій можна автоматизувати. Зі збільшенням кількості статей, що публікуються у соцмережах, та в інших соціальних медіа, це стає реальною проблемою стосовно оперативного реагування та модерування такої інформації.

# 1 АНАЛІЗ ПРОБЛЕМНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАДАЧІ

## 1.1 Проблема поширення фейкових новин та їхній вплив

Сьогодні разом з швидким ростом створюваних користувачами публікацій та новин у соціальних мережах, постає проблема перевірки автентичності фактів у таких новинах перед їх потраплянням до інших користувачів.

Проблематику впливу на масову свідомість людей розглядали Г. Тард, Г. Лебон, Е. Фром, К. Юнг, У. Макдаугол, Ш. Сігеле та інші. Натомість А. Чернова, Б. Грушин, В. Лисенко, Г. Ділігенський, Г. Почепцов, Г. Ортега-і-Гассет, Д. Ольшанський, Е. Фром, І. Панарин, М. Лібікі, М. Маклюен, С. Кара-Мурза, С. Московічі, С. Сьомін, П. Лайнбарджер, М. Сенченко, Я. Варивода та інші досліджували вплив на масову свідомість крізь призму інформаційних війн.

Окремо варто зазначити, що у Росії технологію ведення інформаційних війн досліджували ще в 90-ті роки. Цю тему розробляли такі науковці: В. Березин, Г. Грачев, И. Дзялошинский, С. Зелинский, В. Крисько, А. Манойло, А. Назаретян, С. Расторгуєв.

Натомість в Україні проблемами інформаційної війни активно почали займатися від 2013 року. Найактивніше у цій сфері працює Г. Почепцов. У найновішій його праці “Сучасні інформаційні війни” не лише висвітлено суть та походження інформаційної війни, а й детально досліджено зміни в інформаційному просторі в процесі інформаційного протистояння.

Щодо конкретно фейкової інформації у ЗМІ, то наукових досліджень цієї теми практично немає. Пізнавальними у цій сфері є публіцистичні матеріали, розміщені на сайтах “Медіакритики”, “Mediasapiens”, “Детектор медіа”, “StopFake”, журналу “Ї” та ін.

Найбільш доступні канали для поширення фейків — інтернет-платформи (Reddit, 4chan тощо), соцмережі (Facebook, Twitter та інші) та месенджери (Whatsapp,

Facebook Messenger). Непереверену інформацію, поширену через ці канали, можуть підхопити як “класичні” ЗМІ, так і інші сайти.

Базуючись на цих поняттях та концептах, можна виділити три базові характеристики: автентичність (правда чи ні), намір (поганий чи ні) та чи є ця інформація новинами або ж ні. У таблиці 1 надані ці деталі.

Таблиця 1 – Порівняння між типами інформації стосовно фейкових новин.

| Тип інформації           | Автентичність | Намір              | Новини   |
|--------------------------|---------------|--------------------|----------|
| Зловмисно фейкові новини | Ні            | Поганий            | Так      |
| Фейкові новини           | Ні            | Невідомо           | Так      |
| Сатиричні новини         | Невідомо      | Не поганий         | Так      |
| Дезінформація            | Ні            | Поганий / Невідомо | Невідомо |
| Чутки                    | Невідомо      | Невідомо           | Невідомо |

Для того, щоб якась інформація потрапила в Інтернет, не треба бути журналістом чи власником якогось медіа. Варто лише опублікувати цю інформацію у соцмережі, на Твіттері, у блозі тощо та позначити правильним тегом, щоб пост дістався до цільової аудиторії. Зробити це може будь-хто. І саме тому в Інтернет потрапляє неймовірно багато інформації, яка є неправдою.

Так, часто люди помиляються. Проте, все частіше неправдиву інформацію пишуть з певною метою. Відповідаючи на запитання “для чого поширювати брехню?”, можна виділили декілька типів фейків залежності від їхнього завдання, які:

- сіють паніку серед людей;
- розпалюють міжнаціональну (расову, релігійну тощо) ворожнечу;
- поширюють хибні думки для того, щоб заплутати людей, відволікти від правди;
- маніпулюють свідомістю;
- рекламують когось або щось;

- приносять прибуток ЗМІ, що його поширює (“жовта преса”);
- плямують чиюсь репутацію;
- мають розважальний характер.

Неправдивою інформацією засмічують інформаційне поле за допомогою коментарів (найчастіше з тих самих, фейкових, акаунтів), хештегів (особливо в години, коли цей хештег є найпопулярніший, і люди шукають інформацію за темою хештега, спамери починають свою інформаційну атаку).

Часто творці фейкових новин чи іншої фейкової інформації імітують цитування авторитетного джерела.

Відтак у мережі поширюють скріншоти з начебто цитатами відомих людей. Інформація виглядає дуже правдоподібною, проте правдивою вона не є. В таких випадках треба завжди перевіряти, чи є ця інформація на офіційному сайті чи на офіційній сторінці людини, про яку пишуть.

Проте поширеною є ситуація, коли неправдиву інформацію поширюють офіційні ЗМІ. Засоби масової комунікації стали головним інструментом для розповсюдження повідомлень, які впливають на суспільну свідомість.

А. Моль так пише про ЗМІ: “Вони фактично контролюють всю нашу культуру, пропустивши її через свої фільтри, виділяють окремі елементи із загальної маси культурних явищ і надають їм особисту вагомість, підвищують цінність однієї ідеї, знецінюючи іншу, поляризують таким чином все поле культури.”

Те, що не потрапило до каналів масової комунікації, сьогодні майже не впливає на розвиток суспільства. Отже, сучасна людина не може оминати впливу ЗМІ [2].

Н. Леонов справедливо зазначає: “Інформація, як ніколи, стала інструментом влади. Коли стало відомо про сприйняття людською психікою навіювання, інформація у вигляді пропаганди і агітації стала головним важелем управління людьми. Вона поступово замінила собою грубу силу, насилля, яке довгий час вважалось єдиним знаряддям управління.”

Покійний нині президент США Річард Ніксон, виступаючи одного разу у Раді національної безпеки з питання бюджетних витрат, сказав, що він вважає один долар, вкладений в інформацію і пропаганду, більш цінним, ніж десять доларів, вкладених на створення систем озброєння, бо остання навряд чи буде колись використана, тоді як інформація діє щохвилино і повсюди [3].

Для того, щоб з'ясувати, наскільки читачі/глядачі/слухачі довіряють ЗМІ та чи вміють розпізнавати неправдиві повідомлення у ЗМІ, було проведено соціологічне дослідження анкетуванням.

Вибіркова сукупність – 100 респондентів. Анкети були роздані людям віком від 18 років, це такі особи які мають стале ставлення до певних ЗМІ, та такі, що орієнтуються в політичній ситуації, мають підстави довіряти чи не довіряти засобам масової інформації. Проте було опитано людей в широкому віковому діапазоні – молодь, люди середнього та похилого віку.

Як показало опитування, велика частина респондентів читає новини лише в Інтернеті (41%), а Інтернет, як відомо – це місце, де поширення фейків є наймасштабнішим та найбільш неконтрольованим явищем порівняно з іншими видами ЗМІ. 28% опитуваних дивляться новини по телевізору, 12% – слухають по радіо, 9% – читають газети (найстарша категорія опитуваних загалом), а 10% опитуваних вибрали свій варіант відповіді, – зазвичай це була комбінація Інтернету з іншим видом ЗМІ.

Мало хто з опитуваних перевіряє інформацію, яку він чи вона отримали з одного ЗМІ, на інших. Хоча, якщо задуматись, це робота не для читачів/глядачів/слухачів, а для самих журналістів.

Тим не менше, щоб не повестись на фейк, треба вміти відфільтрувати інформацію та перевіряти її, особливо, якщо йдеться про дуже важливі сенсаційні факти. Перевіряють інформацію лише 13%, 41% – перевіряють іноді, всі інші (46%) – не перевіряють узагалі.

Сам термін “фейк” знають лише 36% опитуваних, 23% про нього чули, проте

значення його не розуміють, 41% про цей термін не чули взагалі.

Отож, як показало опитування, читачі погано розуміють проблему фейків у ЗМІ, а отже, їм потрібно навчитися розпізнавати неправдиві новини і вміти вберегти себе від дезінформування.

У свою чергу, Рада Європи пропонує відмовитися від терміну “фейкові новини” й натомість використовувати поняття “інформаційний розлад”. Це питання було розглянуто тому, що термін “фейкові новини” неточно характеризує складні явища забруднення інформаційного простору.

Також було зазначено ту ж особливість, про яку раніше казали у CNN, що фейковими новинами політики почали називати всі медіа, з позицією яких вони не згодні.

Було запропоновано виділити три типи неправдивих повідомлень:

- навмисно створені й поширені з метою заподіяти шкоду;
- створені без наміру заподіяти шкоду;
- створені на основі реальних подій чи інформації й поширені для того, щоб заподіяти шкоду.

Важливо аналізувати три різні фази: створення повідомлення, перетворення його на медійний продукт і поширення. А комунікацію сприймати не як передачу інформації, натомість визнати її фундаментальну роль у представленні спільних переконань.

## 1.2 Аналіз схожих інструментів щодо визначення інноваційності метода

Насамперед, для того, аби вберегти себе від неправдивої інформації та щоб не стати жертвою фейкової інформації, користувачам варто дотримуватись таких правил:

- читати більше, ніж назву. Якщо проглянути новину до кінця, можна виявити, що назва ніяк не пов'язана з текстом новини або ж у ній немає жодних даних, які підтверджують її правдивість;
- оцінити надійність джерела повідомлення – незнайомі сайти, заповнені оголошеннями й назвами, написаними капслоком, повинні викликати скепсис. Перевіряти через пошукові системи тексти, які були опубліковані раніше (це дасть уявлення про те, чи є джерело надійним), а також URL сторінок, які виглядають підозріло, аби переконатися в тому, що це не фальшивка, що маскується під надійне джерело;
- звертати увагу на дату публікації й час. Поширений елемент фейкових новин – видавати старі публікації за нові. Якщо в новині є посилання на інші джерела, проглянути їх, звертаючи увагу на те, коли вони були опубліковані та як у них описується подія, особливо коли йдеться про час;
- хто автор? Ознайомитися з попередніми публікаціями автора – його бекграунд допоможе зрозуміти, чи дійсно автор є журналістом, чи не писав він фейкових статей до цього;
- ознайомитися з джерелами, на які посилається новина. Відсутність посилань або джерел, з яких були взяті ті чи інші заяви, є очевидним сигналом того, що пост, скоріше за все, є фейком. Будь-які заяви або цитати посадовців мають бути підкріплені надійними джерелами;
- звертати увагу на сумнівні цитати і фотографії. Скептично ставитися до шокуючих або підозрілих цитат і опублікованих фотографій з місця подій;
- не варто поширювати новини лише тому, що поділяєте точку зору автора або ж приймаєте одну із сторін конфлікту;
- з'ясувати, що інші видання пишуть з цього приводу. Якщо жодне інше надійне джерело не повідомляє про дані події, скоріше за все, новина є фейковою;

- замислитися, перш ніж робити перепублікацію. Новинні сайти, які поширюють дезінформацію, працюють з розрахунком на те, що читачі поширюватимуть їхні новини [4].

Враховуючи вищезазначене, зростає важливість підготовки аналітиків, інформаційних працівників до сучасних тенденцій у висвітленні новин, підвищення їхнього фахового рівня й готовності до глибокого аналізу контенту при підготовці інформаційно-аналітичних матеріалів на базі інформації із соціальних медіа.

Набуває також актуальності питання пошуку інформаційно-аналітичними структурами бібліотек нових форм і методів виявлення фейкової інформації та розробки способів мінімізації її впливу.

Аналітик повинен володіти навичками грамотного пошуку інформації, аналізу джерел, уміти критично відбирати й оцінювати її, перевіряти факти, адже мета бібліотеки – дати не лише максимально повну, а й ретельно перевірену інформацію.

У Сполучених Штатах Америки з'явилася Firstdraft News – коаліція провідних світових медіа, які об'єдналися заради кращої перевірки фактів. Серед найбільш відомих учасників CNN, Washington Post, The New York Times, Al Jazeera, Facebook, Google і Twitter.

Інший проект, також запущений французькими та світовими медіа, — CrossCheck. Серед його засновників — AFP, BuzzFeed News, Les Echos, Le Monde, Libération, La Voix Du Nord. Поштовхом до його запуску стали президентські вибори у Франції.

CrossCheck просить читачів ділитися новинами, які викликали в них недовіру, чи питаннями, які їх цікавлять. Далі редакція обирає серед них ті, які розгляне, й потім опублікує результати проведеної перевірки. Окрім медіа проект підтримали FirstDraft та Google News Lab.

В Україні вже понад три роки команда сайту Stopfake.org займається пошуком фальшивих новин і їх спростуванням, а також неодноразово доводила неправдивість повідомлень, абсурдних, аморальних, нелогічних, що не лише поширювалися в

нашій країні, а й формували уявлення про українські події й українців у міжнародній аудиторії. За підсумками 2016 р. було оброблено понад 300 повідомлень, а за весь час існування зроблено понад 1 тис. спростувань.

VoxCheck — ще один факт-чек проект VoxUkraine, започаткований в січні 2016 року. Протягом цього періоду він опублікував 81 статтю та перевірів більше 1000 цитат від 70 українських політиків. Перевірки відбуваються за принципом рецензування.

Спочатку аналітики та волонтери моніторять всі виступи, заяви та дописи політиків і вибирають з них висловлювання, які можна перевірити. Потім вони, користуючись відкритими джерелами, аналізують їх та пишуть перевірку.

Члени редакційної колегії вичитують перевірки та залишають свої зауваження, після чого аналітики враховують ці зауваження та пишуть остаточну перевірку. VoxCheck зібрав 424 висловлювання політиків, що містять факти й піддаються перевірці, та проаналізував їх.

Аналітики дійшли висновку, що українські політики кажуть правду лише в одному випадку з трьох. А 60% із перевірених цитат мали негативну конотацію, тобто спікери робили акцент на негативі.

Також можна розглянути методологію незалежної перевірки фактів проекту PolitiFact під назвою “Truth-O-Meter”, яка базується на ручній перевірці фактів.

Як наслідок, кожне ствердження, або політична стаття у соцмережах оцінюється використовуючи один з шести ступенів правдивості:

- правдиве – ствердження правдиве та нічого важливого не упущено;
- майже правдиве – ствердження правдиве, але потребує уточнення або додаткової інформації;
- частково правдиве – ствердження частково правдиве, але упускає важливі деталі або вириває деякі речі з контексту;
- переважно неправдиве – ствердження містить елемент правди та ігнорує критичні факти які можуть створити інше враження;

- неправдиве – ствердження не точне;
- брехня – ствердження не правдиве та створює обурливе заявлення.

Для того, щоб дати фінальну оцінку статті або публікації у соцмережі, три редактора спільно відповідають на наступні запитання:

- Чи є це твердження буквально та цілком вірним?
- Чи є інший спосіб прочитати це ствердження? Чи залишає це ствердження вільне місце для інтерпретації?
- Чи надав спікер докази? Чи доводить спікер, що це твердження є правдою?
- Яким чином у минулому були розглянуті схожі ствердження?

Наприкінці усі три редактори оцінюють ствердження (де два голоси мають перевагу), іноді залишаючи оцінку як було запропоновано редакторами, а іноді змінюючи її на іншу. Після декількох ітерацій, звіт публікується.

Автоматична система Factmata (<https://factmata.com>) (рис.1.1) станом на сьогодні, знаходиться у бета режимі.

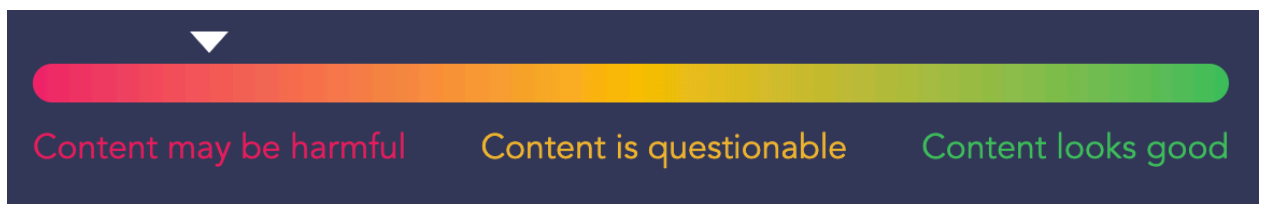


Рисунок 1.1 – Результат аналізу системою Factmata.

Система оцінює вміст за дев'ятьма сигналами (рис.1.2).

- Мова ворожнечі. Мова, що атакує якусь групу на основі таких атрибутів, як релігія, інвалідність або стать;
- Сексизм. Принижуюча або образлива мова проти будь-якої конкретної статі;
- Політичний настрій. Дуже схилена, одностороння точка зору на політичну тему;
- Клікбейт. Сенсаційні заголовки які привертають увагу, а також зазвичай вводять в оману;

- Образи. Зневажливі або образливі зауваження, спрямовані на фізичну або юридичну особу;
- Непристойність. Використання образливих слів та/або зображень;
- Токсичність. Неуважна мова, згадування імен, стереотипів тощо
- Мова ворожнечі за ідентичністю. Зневажливі або образливі зауваження, спрямовані на групу людей / етнічну приналежність.
- Загрози. Висловлюють намір заподіяти шкоду.

This page may contain:

- Political bias** ⓘ 93% likely Strongly biased political language.

This page does not appear to contain:

- Clickbait** ⓘ 44% likely Article headlines which at the expense of being informative, are designed to entice readers into clicking the accompanying link.
- Toxicity** ⓘ 4% likely Demeaning and abusive language in general.
- Obscenity** ⓘ 2% likely Obscene or profane language.
- Sexism** ⓘ 1% likely Demeaning and abusive language targeted towards a particular gender, usually with stereotypes.
- Racism** ⓘ 1% likely Demeaning and abusive language targeted towards a particular ethnicity, usually with stereotypes.
- Insults** ⓘ 0% likely Scornful remarks directed towards an individual.
- Threats** ⓘ 0% likely Expressing a wish or intention for pain, injury, or violence against an individual or group.

Рисунок 1.2 – Результати аналізу за сигналами системи Factmata.

Спробувати свої сили у процесі перевірки фактів вирішив і засновник інтерактивної енциклопедії Вікіпедія Джиммі Вейлз (Jimmy Wales). Його новий проект Wikitribune зараз функціонує в пілотній версії (рис.1.3). У проекті працюють як професійні журналісти, так і волонтери.

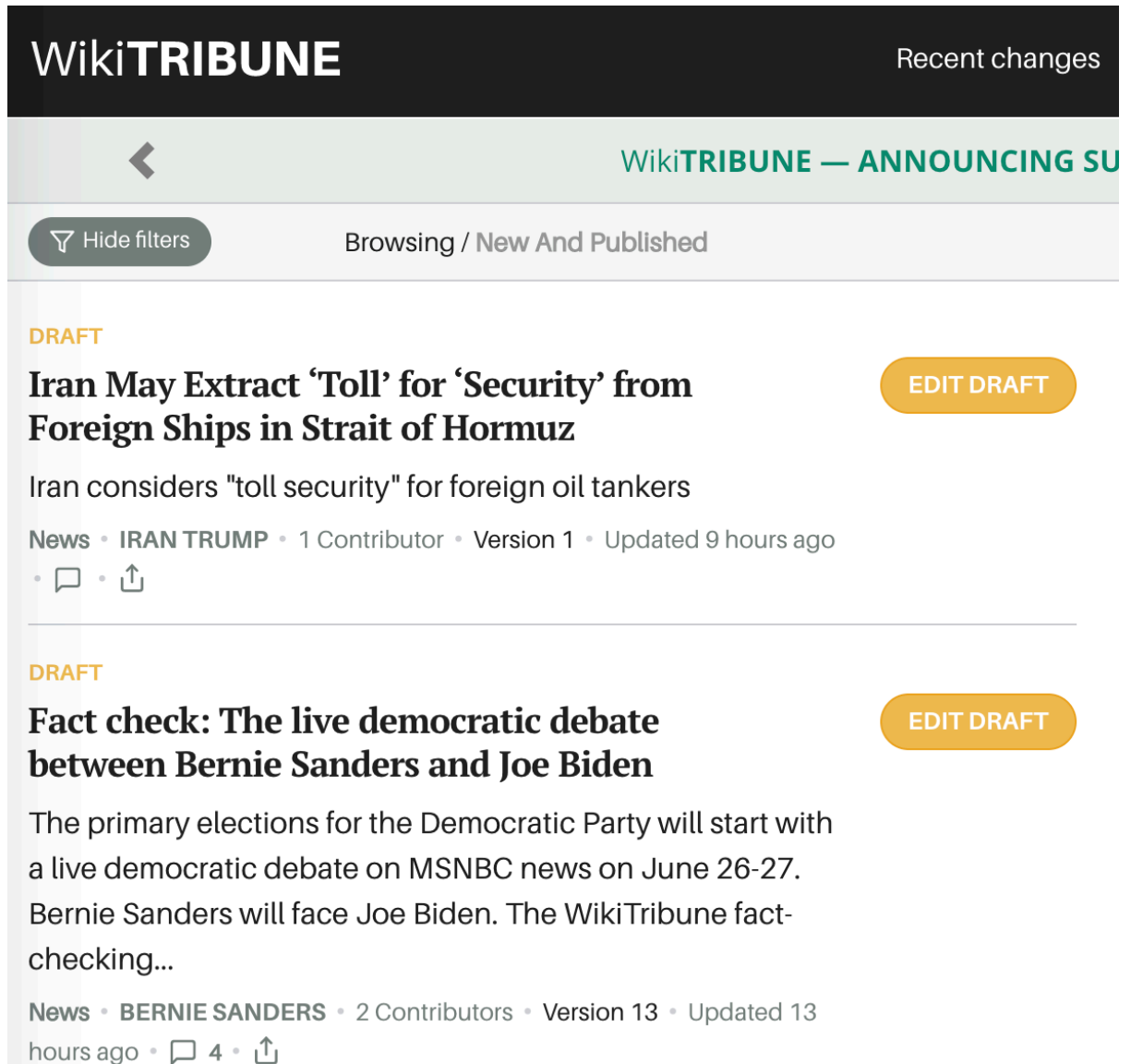


Рисунок 1.3 – Інтерфейс веб-сайту WikiTribune

За задумом, фінансові донори отримують право голосу у виборі тем, які висвітлює Wikitribune, а спільнота читачів може займатися перевіркою фактів та редагувати тексти на сайті.

У Wikitribune є кілька підрозділів за регіонами (наприклад, США, Європа, Азія) та за тематикою (поточні події, політика, культура).

Відповідно до заявлених планів, на сайті не буде реклами, і доступ читачів до матеріалів буде безкоштовним. Таким чином, робота проекту буде фінансуватися за рахунок добровільних внесків. Вікітрібуна успадковує багато принципів Вікіпедії, такі як вільне редагування та необхідність посилення на авторитетні джерела.

Іншою дуже перспективною системою є Fabula AI. Це стартап який, до поглинання Твіттером у 2019 році, займався розробкою нового класу алгоритмів машинного навчання під назвою геометричне глибинне навчання (Geometric Deep Learning).

Він складається зі згорткових нейронних мереж, здатних до навчання моделей на складних і розподілених наборах даних, таких як соціальні мережі.

Компанія каже, що їх алгоритм може розпізнавати підроблені новини з точністю 93% протягом 2 - 20 годин після публікації. У порівнянні з ручною перевіркою яка зазвичай займає набагато довше часу (у середньому 8 годин).

Геометричний алгоритм глибинного навчання розглядає те, як публікаціями діляться на відміну від фокусування на змісті новин, він вивчає ті моделі, які є унікальними та принциповими до фейкових новин (рис.1.4). В останньому випадку згорткові фільтри (візуалізовані у вигляді кольорового вікна) є деформаційно інваріантними за конструкцією.

Приклади використання включають повністю автоматизований фільтр (тобто без жодного людського рецензента в циклі) задля того, щоб привести в дію систему рейтингу довіри до змісту.

Така система зможе знизити вагу сумнівних історій або навіть повністю заблокувати їх, або для проміжного аналізу змісту для маркування потенційних фейкових новин для уваги користувачів.

На даний момент, стартап є частиною Twitter та публічного API на сьогоднішній день не надає.

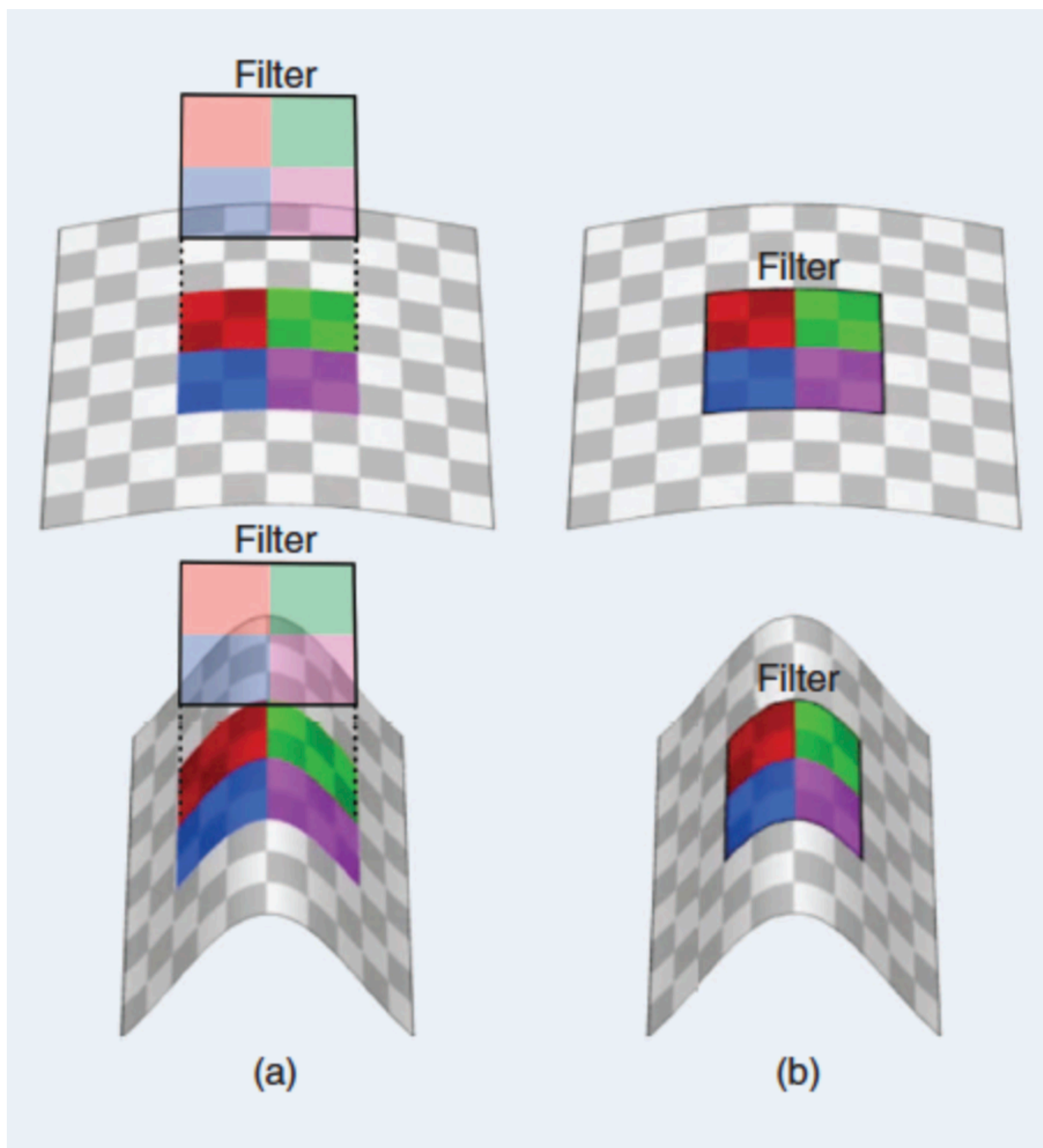


Рисунок 1.4 – Ілюстрація різниці між (a) класичною згортковою нейронною мережею (CNN) застосованою до 3D об'єкту (шахматна поверхня) та (b) геометрична CNN, застосована напряду до кольорової поверхні.

### 1.3 Постановка задачі

Таким чином, на підставі проведеного порівняльного аналізу ручних засобів та методів розпізнавання фейкових новин, слід зробити наступні висновки:

- більшість проектів здійснюють перевірку фактів ручним засобом, використовуючи на це велику кількість людського часу;
- такі засоби дуже важко масштабувати до обсягів генеруємого контенту у соціальних мережах;
- швидкість обробки у існуючих рішень не задовільняє потребам сучасних соціальних мереж для оперативного аналізу поступаючої інформації.

Виходячи з наведених недоліків існуючих аналогів, особливо ручних, можливо зазначити актуальність роботи, яка полягає в тому, що, на відміну від проаналізованих засобів, обраний автоматичний метод матиме змогу робити аналіз своєчасно з моменту появи нового контенту. Також цей метод повинен бути достатньо гнучким для того, щоб мати можливість масштабування до різних обсягів аналізуємої інформації. І саме питання оперативного аналізу генерованої користувачами інформації дає змогу сформулювати мету даної роботи.

Метою роботи є дослідження методів та засобів автоматичного розпізнавання фейкових новин у соціальних мережах. Для цього потрібно проаналізувати існуючі методи та засоби перевірки фактів у інтернет-публікаціях, визначити, які з них найкраще підійдуть для умов задачі автоматичного та оперативного розпізнавання фейкових новин, та обрати алгоритм оцінки достовірності користуючись цими методами.

## 2 АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ ФЕЙКОВИХ НОВИН

### 2.1 Методи розпізнавання фейкових новин у соціальних мережах

Після оновлення внутрішніх алгоритмів, компанія Facebook оголосила, що у стрічці новин вище підіймається релевантний та автентичний контент. Для цього Facebook категоризує сторінки, які часто поширюють недостовірні повідомлення, наприклад, фейкові новини, або повідомлення з клікбейт-заголовками та частіше приховує їхні публікації.

Користувачі можуть самостійно позначати новини, які вони вважають фейковими, та надіслати їх на перегляд модераторам (рис. 2.1).



Рисунок 2.1 – Позначка про “спірний контент” у стрічці новин Facebook.

Співпраця з відомими організаціями які перевіряють факти та запровадження позначок про сумнівне джерело новини.

За перевірку новин для Facebook узялася низка незалежних організацій, що спеціалізуються на перевірці фактів, — ABC News, AP, FactCheck.org, PolitiFact та Snopes. Усі вони є учасниками ініціативи International Fact-Checking Network. Якщо перевірка фактів встановить, що та чи інша новина є сумнівною, соцмережа застосує до неї відповідну позначку.

Компанія Google також вирішила переглянути свої алгоритми, щоб позбутися посилань на фейкові новини та екстремістські висловлювання в результатах пошуку. Авторитетніші джерела інформації пошуковик тепер показує ближче до вершини списку результатів, а ті, які вважаються неавторитетними, були пересунуті донизу.

Нові інструкції для співробітників компанії. Вони стосуються тих модераторів, які оцінюють якість роботи пошуковика. Тепер вони мають позначати в результатах пошуку образливі матеріали, сайти зі спростованими теоріями змови і просто недостовірною інформацією, що відповідно впливатиме на позицію цих ресурсів у результатах пошуку.

Блокування сайтів на сервісі AdSense. Як повідомив пошуковик, блокування на рекламному сервісі AdSense зачіпає ті сторінки, які поширюють недостовірну інформацію.

В компанії наголосили, що регулярно відсіюють таких рекламодавців, але кількість блокувань збільшилася у зв'язку з “різким зростанням кількості фейкових новин та пропагандистських ресурсів”.

У 2019 році Twitter придбав стартап Fabula, що спеціалізується на автоматичному розпізнаванні фейкових новин використовуючи глибинне навчання. Ця розробка послужить основою для дослідницької групи Twitter, яка буде працювати над пошуком нових способів використання машинного навчання через аналіз мови і способів нейро-лінгвістичного програмування, а також системи рекомендацій.

Заснована в 2018 році, Fabula розробила запатентовану систему штучного інтелекту - ефективні алгоритми, які навчаються на великих і складних наборах даних, отриманих з соціальних мереж. Fabula AI допомагає ідентифікувати достовірність будь-якої інформації, яка поширюється в соціальних мережах - незалежно від наміру авторів.

Дослідження показали, що фейки в інтернеті поширюються швидше, ніж реальні новини, і ця схема може бути використана для виявлення дезінформації. Це те, на чому фокусується Fabula: вона виявляє особливості поширення інформації в соціальних мережах, і розподіляє “бали істини”.

Також Twitter запровадив кнопку, за допомогою якої користувачі можуть позначати твіти, що містять хибну, шкідливу або оманливу інформацію . Потім ці репорти модератори переглядають вручну (рис 2.2).

**Help us understand the problem. What is going on with this Tweet?**

**I'm not interested in this Tweet**

**It's suspicious or spam**

**It's abusive or harmful**

[Learn more](#) about reporting violations of our rules.

Рисунок 2.2 – Диалог створення скарги у Twitter

## 2.2 Метод сильної евристики

Навіть високоосвічені люди погано відрізняють надійні джерела інформації в мережі від ненадійних, свідчить експеримент у Стенфордському університеті. Дослідники пропонують нові підходи до перевірки фактів, засновані на методі “сильної евристики”.

Вони дослідили [5], як різні люди оцінюють надійність сайтів та шукають інформацію на політичні та соціальні теми.

Для експерименту вони запросили десятьох докторів наук з історії, десятьох фахівців із перевірки фактів і 25 студентів Стенфордського університету та попросили їх виконати шість різноманітних завдань: порівняти сайти, контент, знайти на сайтах певну інформацію.

Для кожного завдання було встановлено ліміт часу — п'ять чи десять хвилин. Науковці посилаються на дослідження компанії Microsoft, згідно з яким користувачі в середньому проглядають сторінку в мережі близько 70 секунд, а це надто короткий проміжок часу для того, щоб її проаналізувати й оцінити контент. Тому вони вирішили встановити більший дедлайн для своїх респондентів. Утім, деякі з них виконували завдання швидше.

Наприклад, учасникам запропонували оцінити публікації з двох сайтів — Американської академії педіатрії та Американського коледжу педіатрів. Академія — відома організація, заснована в 1932 році, яка об'єднує близько 64 тисяч членів із усього світу та має штат із 450 постійних працівників.

Водночас коледж був заснований у 2002 році групою батьків, має від 200 до 500 членів і лише одного постійного працівника.

Ця організація, зокрема, відома своєю позицією проти всиновлення дітей одностатевими парами та підтримку репаративної терапії, яка згодом була заборонена для неповнолітніх у дев'яти штатах США (суть терапії — у спробах

змінити сексуальну самоідентифікацію людини на гетеросексуальну).

За результатами дослідження (рис.2.1), 64% студентів, які взяли участь у дослідженні, визнали сайт Американського коледжу педіатрів таким, що викликає більше довіри, ніж сайт Американської академії педіатрії. При цьому на сайті коледжу його анти-ЛГБТ позиція зовсім не приховується. Серед істориків цей показник був меншим — 10%; а 40% вирішили, що обидва сайти варті довіри.

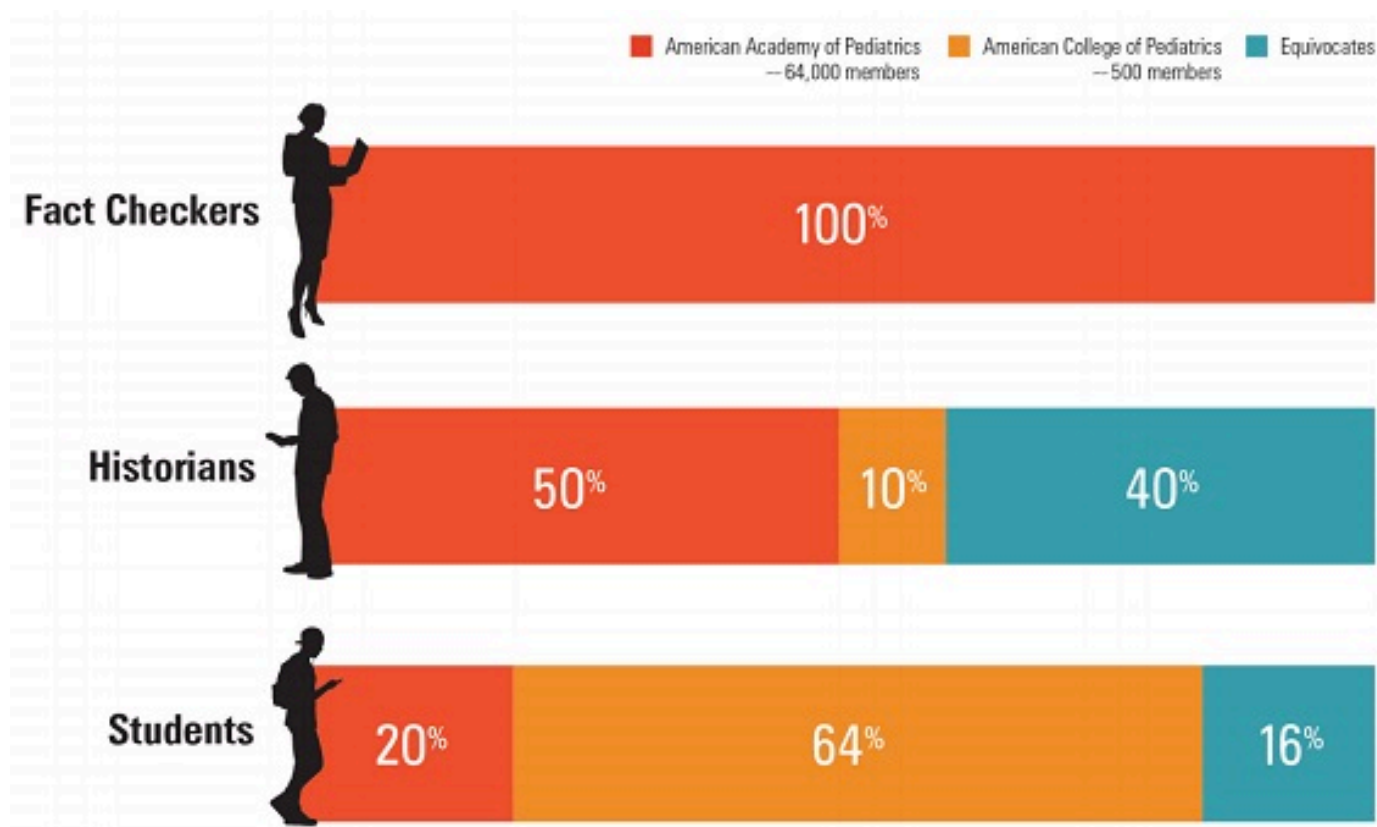


Рисунок 2.1 – Результати дослідження.

Проблем не виникло лише у групи професійних перевіряльників фактів. Їхній 100% результат, який показав експеримент, дослідники пояснюють тим, що перевіряльники фактів не просто оцінювали інформацію, але розглядали її під різними кутами, в контексті.

Такий підхід вони порівнюють із правилом мандрівників, які, перш ніж іти кудись у незнайомій місцевості, намагаються зрозуміти сторони горизонту і в якому

напрямку треба рухатися.

В дослідженні йдеться, що історики та студенти часто ставали жертвами таких особливих деталей веб-сайтів, як офіційні логотипи, доменні імена, наявність чи відсутність банерної реклами.

Вони читають сайти вертикально й так оцінюють їхню надійність. На відміну від них, перевіряльники швидко проглядають контент сайту і відкривають додаткові вкладки у браузері, щоби перевірити його надійність в інших джерелах.

У порівнянні з іншими групами вони дійшли до більш обґрунтованих висновків і за менший час.

Сем Вінберг та Сара Макґрю визнають, що їхнє дослідження потребує більшої вибірки тем та створення для користувачів умов, більш наближених до реальних, у яких більшість із нас читають статті в інтернеті.

Тим не менше, вони наголошують, що в експерименті брали участь талановиті та розумні люди, але різниця між тим, як вони читають контент у мережі, виявилася дуже великою. “Лише два з десяти істориків прискіпливо оцінювали інформацію в інтернеті”, — зазначається в підсумках до експерименту.

Дослідники вважають, що є два підходи, які допомогли групі перевіряльники краще виконувати завдання: це оцінка напрямку (*taking bearings*) та додаткове читання (*lateral reading*) (рис.2.2).

Перш ніж занурюватися в незнайомий контент, вони робили додатковий пошук та уявляли приблизний план того, як вони будуть аналізувати й оцінювати нову інформацію.

Так наприклад, історики спершу читали запропоновані їм статті й лише після цього починали цікавитися сайтами та організаціями, які їх опублікували.

Студенти здебільшого демонстрували так зване вертикальне читання: прогортали сторінки до кінця, оцінюючи дизайн сайту та його контент і не відкривали додаткових вкладок у браузері, щоби пошукати інформацію в інших джерелах.



Рисунок 2.2 – Схема додаткового читання одного з студентів.

Завдання полягало в тому, щоб оцінити статтю про збільшення мінімальної заробітної плати у США. Студент приділив самій статті лише шість секунд, після цього протягом дев'яти секунд вивчав розділ “Про нас” сайту [minimumwage.com](http://minimumwage.com), де з’ясував, що він є проектом Інституту політики зайнятості (Employment Policies Institute, EPI).

Далі експерт перейшов в Google, де пошукав інформацію про EPI. Він знайшов матеріали в The New York Times та National Public Radio, які свідчили про те, що ця організація не варта довіри. При цьому він не читав всіх статей повністю, а виконував швидкий пошук на сторінках за допомогою комбінації клавіш Ctrl + F. На все завдання в нього пішло трохи більше двох хвилин.

“Як це не парадоксально, але ключова особливість додаткового читання — це те, що не потрібно перечитувати великий обсяг нерелевантного тексту. Потрібно лише знати, як побудований пошук в інтернеті, знання стратегій ефективного пошуку та навігації в мережі”, — зазначають Вінберг та Макгрю. Вони також помітили, що їхні піддослідні часто стають жертвами евристики репрезентативності (representativeness heuristic).

Евристика — це спрощений спосіб розуміння речей, без наукових чи аналітичних розрахунків. І професійний дизайн сайту Американського коледжу

педіатрів збив із пантелику багатьох учасників саме тому, що вони очікували якісного оформлення від серйозного наукового сайту. У запропонованій їм статті про цькування дітей у школах був зміст, підзаголовки та список використаної літератури з посиланнями на видання, назви яких звучать авторитетно. Респонденти оцінили вірогідність того, що ця стаття й цей сайт варті довіри на основі цих зовнішніх ознак.

Вінберг та Макгрю посилаються на роботу Герда Гігеренцера (Gerd Gigerenzer) та Вольфганга Гайсмаєра (Wolfgang Gaissmaier), в якій аналізується використання евристики при прийнятті рішень. Вони виділили слабку евристику (наприклад, врахування домену та дизайну сайту при оцінці його надійності) та сильну евристику (пошуки додаткової інформації про сайт в мережі).

На думку дослідників зі Стенфорду, додаткове читання, яке продемонструвала група перевіряльників під час експерименту, цілковито підпадає під поняття сильної евристики. А те, що перевіряльникам, які використовували цей підхід, потрібно було менше часу, щоби впоратися із завданнями, відповідає й іншому виклику.

Користувачі щодня отримують величезні обсяги інформації і в них просто немає часу довго й прискіпливо розбиратися з кожною статтею й кожним постом у соцмережах.

Заступниця редактора NiemanLab Лаура Хазард Оуен (Laura Hazard Owen) розглядає [6] результати дослідження Вінберга та Макгрю як основу для нового типу навчання медіаграмотності. За її прогнозами, він ґрунтуватиметься на евристиці, розумінні того, як оцінювати та знаходити інформацію, а не на вивченні ознак, за якими можна розпізнати фейк.

Відтак, Лаура Хазард Оуен вважає, що класичний спосіб навчання медіаграмотності (дев'ять порад від The Huffington Post) не дуже ефективний. Натомість дослідникам варто зосередитися на пошуках нових і дієвіших методів навчання різних груп людей навичок, які допоможуть їм об'єктивно оцінювати інформацію, яку вони споживають.

Журналістка Брук Борел (Brooke Borel), яка працювала у якості перевіряльника фактів й написала з цієї теми книжку *The Chicago Guide to Fact-Checking*, вважає, що не зважаючи на всі зусилля соцмереж і перевіряльників фактів фейкові новини постійно адаптуються. “Я знаю, що може зробити перевірка фактів, і наскільки це важливо. Але для боротьби з фальшивими новинами цього просто недостатньо”, — пише Борел.

На її думку, перш за все медіа мають менше зловживати “жовтими” новинами та заголовками. Таким чином фейковим новинам буде складніше губитися в загальному потоці інформації.

Експертка Брук Борел також вказує на те, що соцмережі й інші платформи мають зробити прозорішою роботу своїх алгоритмів, щоби стали зрозумілими механізми, завдяки яким деякі новини активно поширюються і стають вірусними, а інші — ні.

Вона посилається також на професорку Енді Такер (Andie Tucher) із Колумбійського університету, яка працює над історією фейкових новин у США. Такер вважає, що цикл фейкових новин почав змінюватися таким чином, що люди почали втомлюватися від них. І багато людей вирішують вийти за його межі.

Тоді на допомогу приходять розробки у сферах автоматичного розпізнавання фейкових новин.

## 3 ОПИС ОБРАНОГО МЕТОДА

### 3.1 Метод розпізнавання фейкових новин на основі зобраних знань

При роботі з фейковими новинами з опираючись знання, потрібно прагнути проаналізувати та / або виявити фальшиві новини, використовуючи процес, відомий як перевірка фактів.

Вперше, перевірка фактів використовувалась в журналістиці, спрямована на оцінку автентичності новин шляхом порівняння отриманих знань із змісту новин (наприклад тверджень) з відомими фактами (тобто, правдою).

У цьому розділі буде розглянуто традиційну перевірку фактів (також відому як ручну перевірку фактів) і як її можна трансформувати у автоматичний засіб аналізу та виявлення фейкових новин (тобто автоматичної перевірки фактів).

#### 3.1.1 Ручна перевірка фактів



Загалом, ручну перевірку фактів можна поділити на:

- експертну (expert-based) (рис.3.1);
- перевірку фактів спільнотою людей (crowd-sourced) (рис.3.2).

Експертна перевірка фактів покладається на експертів з якоїсь теми (тобто перевіряючих факт) для перевірки змісту даної новини. Експертна перевірка фактів часто проводиться невеликою групою висококваліфікованих перевіряльників, єю легко керувати та вона приводить до високоточних результатів, але є дорогим засобом та погано масштабується із збільшенням обсягу перевіряємих новин.

Перевірка спільнотою людей зазвичай являє собою відкриту платформу до якої може залучитися майже кожен щоб зробити свій внесок у перевірку фактів.

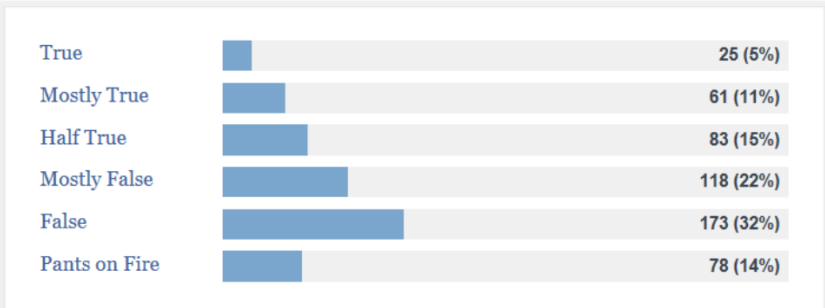
## Donald Trump's file

### Republican from New York

Donald Trump was elected the 45th president of the United States on Nov. 8, 2016. He has been a real estate developer, entrepreneur and host of the NBC reality show, "The Apprentice." Trump's statements were awarded PolitiFact's 2015 Lie of the Year. Born and raised in New York City, Trump is married to Melania Trump, a former model from Slovenia. Trump has five children and eight grandchildren. Three of his children, Donald Jr., Ivanka, and Eric, serve as executive vice presidents of the Trump Organization.

### The PolitiFact scorecard



| Category      | Count | Percentage |
|---------------|-------|------------|
| True          | 25    | 5%         |
| Mostly True   | 61    | 11%        |
| Half True     | 83    | 15%        |
| Mostly False  | 118   | 22%        |
| False         | 173   | 32%        |
| Pants on Fire | 78    | 14%        |

Рисунок 3.1 – (Expert-based) PolitiFact: графік оцінювання.

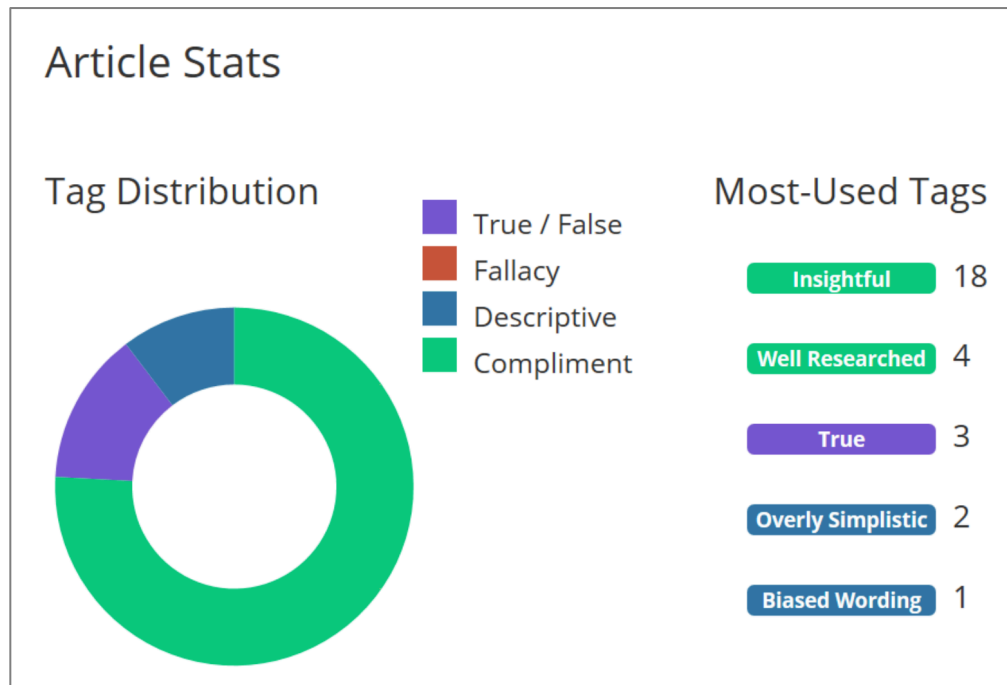


Рисунок 3.2 – (Crowd-sourced) Fiskkit: поширення тегів

Останнім часом також з'явилося багато організацій, які роблять експертний аналіз новин. У таблиці 2 наведено деталі стосовно найвідоміших веб-сайтів та організацій.

Таблиця 2 – Порівняння організацій які використовують експертну оцінку

| Організація                      | Галузь аналізу   | Аналізуємий контент                          | Варіанти оцінювання  |
|----------------------------------|--|--|--|
| PolitiFact                       | Політика США   | Твердження                                   | Правда; Практично правда; Частково правда; Практично неправда; Неправда; Брехня  |
| The Washington Post Fact Checker | Політика США   | Твердження                                   | Один Піноккіо; Два Піноккіо; Три Піноккіо; Чотири Піноккіо, Відмітка Джузеппе; Перевернутий Піноккіо; Вердикт ще не видано                             |
| FactCheck                        | Політика та інші соціальні питання                     | Реклама, дебати, доклади, інтерв'ю та новини | Правда; Відсутність доказів; Неправда  |
| Snopes                           | Політика, релігія,                                     | Новини та відео                              | Правда; Практично правда; Суміш; Практично неправда; Не доказано; Застаріло; Невірно; Правильне віднесення; Неправильне донесення; Шахрайство; Легенда |
| TruthOrFiction                   | Політика, релігія, медицина, тощо                      | Email чутки                                  | Правда; Фікція; тощо   |
| FullFact                         | Економіка, здоров'я, освіта, злочини, іміграція, закон | Статті                                       | Неоднозначність (нема чотких оцінок)   |
| NoaxSlayer                       | Неоднозначність  | Статті та повідомлення                       | Містифікація, шахрайство, злісна неправда, фальшиве попередження, фейкові новини, омана, гумор, спами, тощо  |

Деякі організації надають додаткову інформацію, наприклад, PolitiFact надає “систему показників PolitiFact”, в якій представлені статистичні дані про достовірність розподілу всіх заяв, пов'язаних з певною темою. Ця інформація може допомогти визначити достовірні теми, які потребують подальшого вивчення для перевірки.

Іншим прикладом є NoaxSlayer, який відрізняється від більшості веб-сайтів, що перевіряють факти на автентичність, оскільки він додатково класифікує статті та повідомлення, наприклад, містифікації, спама та фальшиві новини. Хоча веб-сайт не дає чітких визначень для цих категорій, його інформація може бути потенційно використана як основна істина для порівняльних досліджень фейкових новин. На додаток до цього переліку, повний перелік веб-сайтів для перевірки фактів надає лабораторія репортерів з Дюкського університету (<https://reporterslab.org/fact-checking>), де перелічено понад двісті веб-сайтів для перевірки фактів у різних країнах і мовах. Як правило, ці веб-сайти можуть надавати корисні дані к якості основи для виявлення фейкових новин.

Ці веб-сайти надають детальний експертний аналіз стосовно вже перевіреного контенту (наприклад, що є хибним і чому воно є хибним), а також надають важливу інформацію стосовно різних аспектів аналізу фейкових новин.

Однак, на сьогоднішній день такі дані ще не були використані належним чином.

Ручна перевірка фактів спільнотою людей спирається на велику кількість осіб, які діють як перевіряльники фактів (тобто, колективний інтелект). У порівнянні з перевіркою фактів на експертної оцінки, перевірка фактів спільнотою людей є відносно важкою, менш надійною та точною через політичні погляди перевіряючих та їх суперечливих аотацій, однак має кращу (хоча і недостатню) масштабованість.

Таким чином, у перевірці фактів спільнотою людей часто необхідно:

- фільтрувати ненадійних користувачів;
- вирішувати суперечливі результати з перевірки фактів.

Причому обидва вимоги стають більш критичними, оскільки зростає кількість перевіряючих фактів. Тим не менш, платформи зі спільнотами частіше за все дозволяють перевіряючим надавати більш детальну інформацію (наприклад, їхні настрої або позиції), які можуть бути додатково вивчені для покращення ідентифікації фейкових новин.

Веб-сайти з перевірки фактів з спільнотою людей, на відміну від експертної оцінки фактів, все ще перебувають у ранній стадії розробки. Прикладом може служити Fiskkit, де користувачі можуть завантажувати статті, надавати рейтинги реченням у статтях і вибирати теги, які найкраще описують його.

Дані джерела статей допомагають:

- розрізнити типи контенту (наприклад, новини проти не новин);
- визначити його довіру.

Теги, класифіковані у кількох категоріях дозволяють знаходити та аналізувати моделі у фейкових та не фейкових публікаціях (рис.3.2).

Незважаючи на те, що веб-сайтів які спеціалізуються на перевірці фактів спільнотою людей не дуже багато, ми можемо вважати, що буде з'являтися більше платформ або інструментів які використовують цей принцип, оскільки основні техногіганти та соціальні мережі вже усвідомили важливість ідентифікації фейкових новин (наприклад, Google, Facebook, Twitter тощо).

### 3.1.2 Автоматична перевірка фактів

Перевірка фактів вручну (за допомогою експертів або більшої кількості людей) практично не масштабується завдяки збільшенню обсягу новоствореної інформації, особливо це актуально в соціальних мережах.

Для того, щоб метод можна було масштабувати були розроблені автоматичні методи перевірки фактів спираючись на методику отримання інформації (Information Retrieval або IR) та технології обробки природної мови (Natural Language Processing або NLP), а також на теорію графів [7].

Для перегляду цих методів ми спочатку надаємо єдине стандартне представлення знань які можуть бути оброблені машинами автоматично і були широко прийняті у відповідних дослідженнях [8].

Введемо таке поняття як “Знання”. Це буде набір з (Предмет, Предикат, Об’єкт) (Subject, Predicate, Object або SPO) трійки сутностей добутих з публікації, яка добре представляє дану інформацію.

Наприклад, знання в реченні “Дональд Трамп є президентом США” може бути: (DonaldTrump, Profession, President).

На основі наступної репрезентації знань ми матимемо можливість створити інструмент автоматичної перевірки фактів для оцінки автентичності новин.

Далі, деталізуємо автоматичний процес перевірки фактів, можливі завдання в межах процесу, а також деякі відкриті питання в автоматичній перевірці фактів.

Загальний процес автоматичної перевірки фактів відображено на рисунку 3.1.

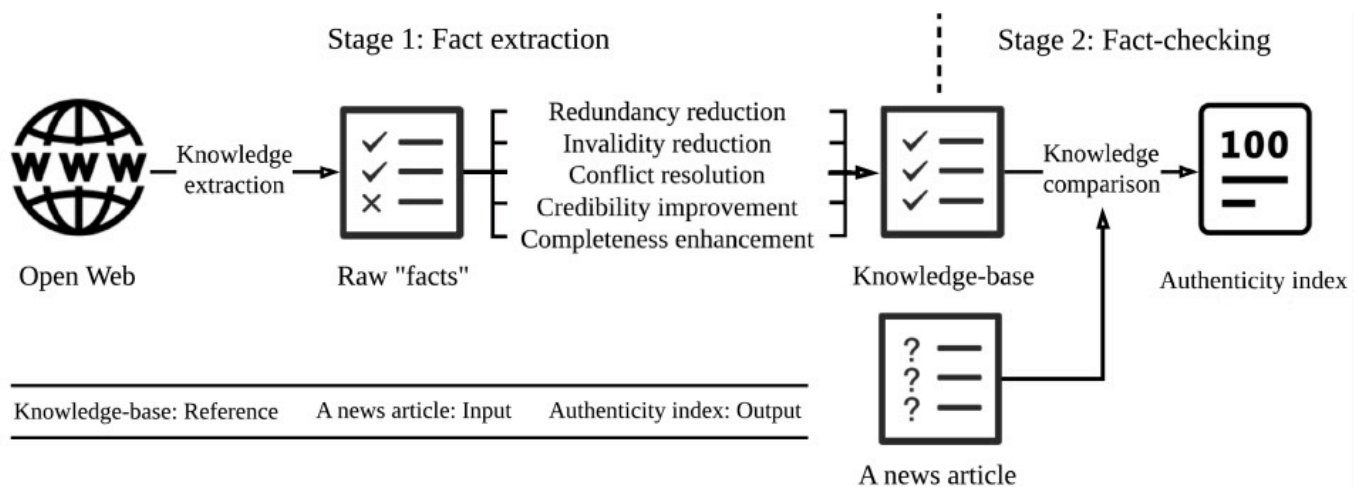


Рисунок 3.1 – Ілюстрація процесу автоматичної перевірки фактів.

Його можна розділити на два етапи:

- fact extraction або добування фактів (також відомий як knowledge-base construction);
- fact checking або перевірка фактів (також відомий як knowledge comparison).

На етапі добування фактів, знання збирається загалом з відкритої мережі інтернет, що на виході створює великий обсяг неструктурованої інформацію у формі онлайн-документів.

Добуті знання використовуються для побудови бази знань (Knowledge Base або KB) або графу знань (Knowledge Graph), кожен з яких після належного очищення даних містить набір фактів (тобто, справжні знання).

Під час перевірки фактів автентичність змісту новин, що підлягають перевірці, визначається шляхом зіставлення знань, отриманих зі змісту новин, до фактів, що зберігаються в побудованій базі знань або графу знань.

### 3.1.3 Open Web та збір знань

Щоб зібрати факти, знання часто витягуються з відкритої мережі інтернет як “необроблені факти”, знання, які є зайвими, застарілими, конфліктними, ненадійними або неповними. Ці необроблені факти далі обробляються та очищаються шляхом завдань обробки знань для побудови бази знань або графа знань.

Збір знань (Knowledge Extraction), також відоме як екстракція відносин, має на меті збирати необроблені факти з мережі інтернет.

Загалом існує чотири (але не обмежуючись) типи веб-контенту: текст, табличні дані, структуровані сторінки та анотації, які містять реляційну інформацію і можуть бути використані для вилучення знань різними екстракторами.

Вилучення знань можна додатково класифікувати на вилучення знань з одним джерелом або з відкритим кодом. Вилучення знань з одного джерела, яке спирається на одне порівняно надійне джерело (наприклад, Вікіпедія) для отримання знань, є відносно ефективним, але часто призводить до отримання неповного знання.

З іншого боку, добування відкритих знань має на меті злиття знань з різних джерел, що призводить до меншої ефективності, але більш повного знання.

### 3.1.4 Завдання з обробки зібраних знань

Збір знань з мережі інтернет створює набір необроблених фактів. Знання в необроблених фактах може бути надмірне, наприклад, (DonaldJohnTrump, profession, President) є надмірне, коли (DonaldTrump, profession, President) так як DonaldTrump і DonaldJohnTrump відносяться до одного і того ж суб'єкта; недійсне, наприклад, (Britain, joinIn, EuropeanUnion) є застарілим і повинно бути видалено або оновлено; суперечливе, наприклад, (DonaldTrump, bornIn, NewYorkCity) та (DonaldTrump, bornIn, LosAngeles) є парою з конфлікуючими знаннями; нереалістичне, наприклад, знання, отримані з The Onion, організації сатиричної новини, слід вважати ненадійними знаннями і неповними.

Для вирішення цих питань та очищення необроблених фактів необхідно виконати наступні п'ять завдань:

- резолюція об'єкта (Entity Resolution) для зменшення надмірності. Резолюція об'єкта, також відома як дедуплікація (deduplication) або зв'язок між записами (record linkage) - це завдання ідентифікації всіх згадок, які стосуються однієї і тієї ж сутності реального світу в межах бази знань або між кількома базами знань. Найбільш пов'язаним з перевіркою фактів є відносна резолюція сутності (relational entity resolution), де поточні методи

- базуються або на відстані, або на основі залежності. Резолюція об'єктів часто вимагає обчислення попарних подібностей, що є дорогим з обчислюваної точки зору. Методи блокування (або індексації) часто використовуються для вирішення цієї обчислювальної складності;
- запис часу (Time Recording) для видалення застарілих знань. Правильність деяких фактів залежить від конкретного часового інтервалу. Одним із способів вирішення цієї проблеми є конструювання типу Compound Value Type (CVT), що дозволяє фактам мати анотації на початкову та кінцеву дати. Інший шлях, це уточнити поточні факти, додавши до них додаткові твердження. Проте дослідження з перевірки фактів, які розглядали своєчасність, обмежені на сьогоднішній день, враховуючи важливість своєчасності в новостях та раннього виявлення підроблених новин, де факти можуть швидко оновлюватися.
  - злиття знань (Knowledge Fusion) для обробки суперечливих знань. Конфліктуючі знання є поширеним при використанні вилучення знань з відкритим вихідним кодом, у той час як воно менш часте при прийнятті вилучення з одного джерела. Для розв'язання конфлікту знань часто визначаються значення для фактів. Як альтернатива, цілісні методи з машинного навчання можуть бути використані для об'єднання декількох алгоритмів виявлення істин, ефективно виявляючи справжні значення від конфліктуючих. Достовірність веб-сайтів, з яких береться знання, часто є важливим критерієм, який використовується в процесі синтезу знань;
  - оцінка достовірності (Credibility Evaluation) для підвищення достовірності знань. Оцінка достовірності знань зосереджена на аналізі добутого знання з обраного веб-сайту, де підходи, засновані на контрольованому навчанні і статистичний висновок були розроблені для аналізу веб-контенту та / або посилань з іншими веб-сайтами;

- прогнозування посилань (Link Prediction) для виведення нових фактів. Сирі факти, отримані з онлайн-ресурсів, зокрема з використанням одного джерела, часто дуже далекі від закінчених. Отже, надійне виведення нових фактів на основі існуючих фактів є необхідним для вдосконалення побудови баз знань. Сучасні методи, які можуть допомогти передбачити нові факти, можна класифікувати за трьома групами на основі їх припущень: приховані моделі характеристик, які припускають існування трійок бази знань умовно незалежними, що дають приховані особливості та параметри, моделі характеристик графів, які припускають, що існування трійок є умовно незалежними даними спостережуваними характеристиками та параметрами графа (наприклад, алгоритм ранжирування контурів (Path Ranking Algorithm або PRA)), і Маркове випадкове поле (Markov Random Field або MRF), які припускають, що існуючі трійки мають місцеві взаємодії.

Отже, виконання цих п'яти завдань допоможе створити базу знань для подальшого аналізу.

### 3.1.5 Побудова бази знань

База знань формується за допомогою “очищених” знань (наприклад, набір трійок SPO). суб'єкти або об'єкти в SPO трійки) представлені у вигляді вузлів, а відносини (тобто предикати з трійки SPO) представлені у вигляді ребер.

Бази знань або графи знань є підходящими кандидатами для надання основної істини дослідженням фейкових новин, тобто ми можемо обґрунтовано припустити, що існуючі трійки в базі знань або графу знань відображають справжні факти. Однак для неіснуючих трійок існують три загальні припущення:

- припущення “закритого світу” (Closed-world Assumption): неіснуючі трійки вказують на помилкові знання. Хоча це припущення спрощує автоматичну перевірку фактів, це досить небезпечно, оскільки бази знань часто є неповними;
- припущення “відкритого світу” (Open-world Assumption): неіснуючі трійки вказують на невідоме знання, яке може бути або істинним, або помилковим. Це припущення може призвести до більш точних результатів перевірки фактів, але результати залежать від того, як автентичність для неіснуючих трійок виводиться з існуючих;
- місцеве припущення “закритого світу” (Local Closed-world Assumption): автентичність неіснуючої трійки ґрунтується на наступному правилі: припустимо, що  $T(s, p)$  - сукупність існуючих трійок для даного суб'єкта  $s$  і предиката  $p$ . Для будь-якого  $(s, p, o) \in T(s, p)$ , якщо  $|T(s, p)| > 0$ , ми говоримо, що трійка невірна; якщо  $|T(s, p)| = 0$ , достовірність трійки  $(s, p, o)$  невідома.

В останні роки замість побудови бази знань з відкритої мережі інтернет для кожної задачі з перевірки фактів, було побудовано кілька великомасштабних графів знань, наприклад, YAGO, Freebase18, NELL, PATTY, DBpedia, DeepDive.

Однак слід зазначити, що існуючі графи знань не є достатніми у якості джерела істини, особливо для новин, і можуть бути покращені у подальшому.

Наприклад, у зв'язку з виходом новин у зазначений час, ці новини часто неможливо спроектувати на “загальні знання”, бо вони відносяться більш до останніх подій.

Отже, динамічна база знань може значно підвищити точність перевірки фактів новин. Крім того, було підтверджено, що фальшиві новини розповсюджуються швидше, ніж справжні новини, що надає великого значення швидкій перевірці фактів.

Сучасні дослідження по створенню баз знань зосереджені на побудові бази знань з якомога більшою кількістю фактів. Проте швидка перевірка фактів новин вимагає не тільки виявлення частин перевірених новин, які є достовірними для перевірки, але й бази знань, яка зберігає лише стільки “цінних” фактів, наскільки це можливо (тобто процес спрощення бази знань).

Швидкість перевірки фактів також дуже пов'язаний зі стратегіями та методами, які використовуються для перевірки фактів новин

### 3.1.6 Перевірка фактів

Щоб оцінити достовірність новин, нам необхідно далі порівнювати знання, отримані зі змісту новин, що підлягають перевірці (наприклад, трійки SPO), з фактами, що зберігаються в побудованій або існуючій базі знань або графіку знань, тобто з правдою.

Як правило, стратегія перевірки фактів для трійці SPO (Subject, Predicate, Object) полягає в тому, щоб оцінити можливість того, що вершина, позначена як Predicate існує з вузла, позначеного Subject на вузол, що представляє об'єкт у графі знань. А саме:

- розміщення об'єкта (Entity locating). На цьому кроці Subject (Об'єкт) збігається з вузлом на графіку знань, який представляє ту ж сутність, що і Subject (Об'єкт). Зауважимо, що представлення одного і того ж об'єкта не є еквівалентом представлення того ж самого рядка, наприклад, Donald J. Trump і Donald John Trump обидва представляють одну і ту ж сутність. Отже, методи ідентифікації об'єктної можуть бути використані для визначення правильних відповідностей;
- перевірка відношення (Relation verification). Трійка (Subject, Predicate,

- Object) вважається істиною, якщо вершина, позначена Predicate, з вузла, що представляє Subject, до того, що представляє об'єкт, існує у графі знань. В іншому випадку, його автентичність є помилковою, заснованою на припущенні закритого світу, або визначається після наведення знань;
- наведення знань. Коли трійка (Subject, Predicate, Object) не існує у графі знань, ймовірність того, що вершина, позначена предикатом, існуватиме з вузла, що представляє предмет, до того, що представляє об'єкт, можна обчислити, наприклад, за допомогою методів передбачення зв'язку, таких як семантична близькість, дискримінаційний предикатний шлях, або LinkNBed.

Тепер можна надати формальне визначення для виявлення фейкових новин (тобто оцінки автентичності новин).

Припустимо, що стаття, що підлягає перевірці, представлена у вигляді набору відомостей про знання (тобто SPO-трійки)  $(s_i, p_i, o_i)$ ,  $i = 1, 2, \dots, n$ .

Нехай  $G_{KB}$  посилається на граф знань, що містить набір фактів (тобто, правда), позначений як,  $(s_{t_j}, p_{t_j}, o_{t_j})$ ,  $j = 1, 2, \dots, m$ . Завданням оцінки достовірності кожної трійки  $(s_i, p_i, o_i)$  є ідентифікація функції  $\mathcal{A}$ , яка присвоює значення автентичності  $A_i \in [0, 1]$  відповідному  $(s_i, p_i, o_i)$  шляхом порівняння його з кожним  $(s_{t_j}, p_{t_j}, o_{t_j})$  графом знань, де  $A_i = 1$  вказує, що трійка істина, а  $A_i = 0$  вказує, що вона є помилковою.

Остаточний індекс автентичності  $A \in [0, 1]$  статті, що підлягає перевірці, отримується шляхом агрегування всіх  $A_i$  використовуючи формулу (3.1)

$$\begin{aligned} \mathcal{A}: (s_i, p_i, o_i) &\xrightarrow{G_{KB}} A_i, \\ A &= \mathcal{G}(A_1, A_2, \dots, A_n), \end{aligned} \tag{3.1}$$

де  $\mathcal{G}$  є функцією вибору агрегації.

Стаття, що перевіряється, є істинною, якщо  $A = 1$ , і є незаперечною помилкою, якщо  $A = 0$ .

Зокрема, функція  $\mathcal{A}$  може бути сформульована за формулою (3.2)

$$\mathcal{A}((s_i, p_i, o_i), G_{KB}) = P(\text{вершина } p_i \text{ зв'язана з } s'_i \text{ на } o'_i \text{ у } G_{KB}), \quad (3.2)$$

де  $P(\cdot)$  позначає ймовірність, а  $s'_i$  та  $o'_i$  є зв'язаними сутностями з  $s_i$  та  $o_i$  у  $G_{KB}$  відповідно з формули (3.3).

$$\begin{aligned} s'_i &= \arg \min_{s_{t_j}} \|s_i - s_{t_j}\|, \\ o'_i &= \arg \min_{o_{t_j}} \|o_i - o_{t_j}\| \end{aligned} \quad (3.3)$$

### 3.2 Метод розпізнавання фейкових новин на основі їхнього розповсюдження

При вивченні фейкових новин з точки зору шаблонів їхнього розповсюдження, використовують інформацію пов'язану з поширенням таких новин, таку як вони поширюються і які користувачі їх поширюють.

Тут ми спочатку представляємо:

- емпіричні закономірності;
- математичні моделі розповсюдження фейкових новин.

Перш ніж впровадити такі моделі або засоби, необхідно відповісти на ряд фундаментальних питань. Наприклад, як можна описати або представити поширення фейкових новин?

Які заходи необхідні для характеристики розповсюдження фейкових новин? Чи існує різниця між поширенням фейкових новин і регулярними новинами (наприклад, з точки зору відповідних заходів)?

Чи існує різниця між поширенням фейкових новин у різних областях (наприклад, політика, економіка та освіта), темах (наприклад, природні катастрофи, президентські вибори та здоров'я), веб-сайтах (наприклад, Twitter, Reddit, WeChat) або мови (наприклад, англійська, китайська, і російська)?

Далі наведено кілька прикладів багатьох питань, пов'язаних з розповсюдженням фейкових новин, відповіді на які ще залишаються відкритими.

Отже, щоб забезпечити єдине представлення стосовно поширення фейкових новин, ми спочатку визначимо поняття каскад фейкової новини як дерево або деревоподібна структура, яка представляє поширення певної фейкової статті у соціальній мережі користувачів (рис.3.2) де глибина дорівнює 3, ширина (розповсюдження) 1,2,3,1 та розмір дорівнює 7.

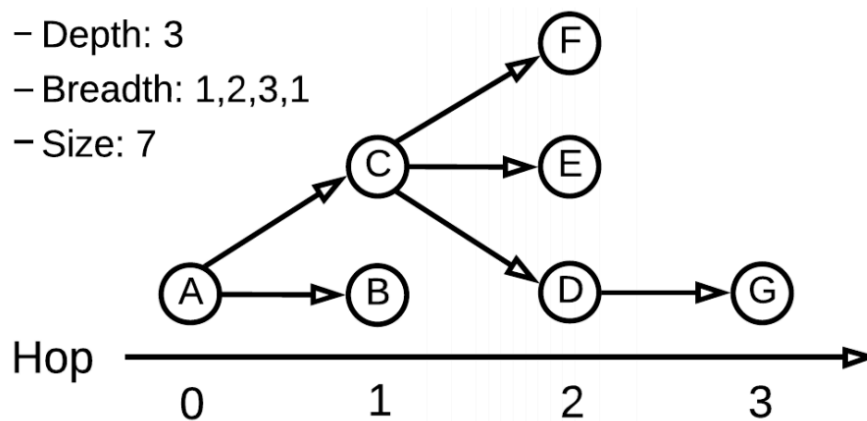


Рисунок 3.2 – Ілюстрація каскаду фейкової новини на основі кроків.

Кореневий вузол каскаду відображує користувача, який вперше опублікував фейкову новину (тобто, творця або ініціатора).

Інші вузли каскаду являють собою користувачів, які згодом розмістили статтю шляхом її пересилання / розміщення одразу після її розміщення батьківськими вузлами, з якими вони з'єднані за допомогою ребер.

Каскад фейкової новини може бути представлений у термінах кількості кроків наприклад, скачок (або Нур) фейкової новини, яку вона пройшла (рис.3.2) або часі, скільки разів вона була опублікована (рис.3.3)

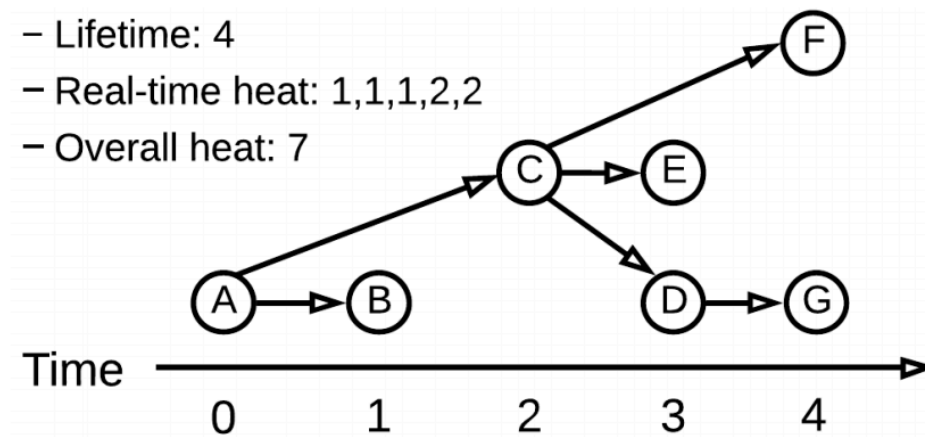


Рисунок 3.3 – Ілюстрація каскаду фейкової новини на основі часу.

Каскад, оснований на кількості стрибків зазвичай є стандартним деревом, яке дозволяє робити заміри, такі як:

- глибина: максимальна кількість кроків (стрибків) фейкова новина пройшла в межах каскаду;
- ширина (при  $k$  стрибках): кількість користувачів, які отримали фейкову новину  $k$  кроків (стрибків) після того, як вона з'явилась у каскаді;
- розмір: загальна кількість користувачів у каскаді.

Заснований на часі каскад фейкової новини є також деревоподібною структурою, що дозволяє робити наступні заміри:

- термін життя: найдовший інтервал, протягом якого поширюється фейкова новина;

- розігрів у реальному часі (в момент часу  $t$ ): це кількість користувачів, що публікують / пересилають фейкову новину у момент часу  $t$ ;
- загальний розігрів: загальна кількість користувачів, які переслали / розмістили фейкову новину.

Треба зауважити, що конкретні фейкові новини можуть призвести до декількох одночасних каскадів через декількох ініціюючих користувачів.

Крім того, часто в межах каскаду фейкової новини, вузли (користувачі) представлені рядом атрибутів і додаткової інформації, наприклад, чи вони підтримують або протиставляють стосовно фейкової новини, інформацію про профіль, попередні повідомлення та їхні коментарі.

### 3.2.1 Шаблони розповсюдження фейкових новин

Моделі розповсюдження фейкової новини можуть бути розділені на шаблони які тільки описують поширення фейкової новини і ті, які порівнюють розповсюдження фейкової новини з регулярними новинами.

Обидва типи шаблонів часто забезпечують заміри отримані з каскадів фейкових новин, наприклад, фейкові новини подорожують все далі і швидше, ніж справжні новини.

Як обговорювалося, такі заміри залежать від того, як розповсюдження фейкових новин було представлено, використовуючи каскад оснований на кроках чи на часі:

- непідтверджені новини часто з'являються знову (рис.3.4). Як було розглянуто, непідтверджені новини мають тенденцію багаторазово з'являтися в період сплеску дискусій, тоді як підтверджена новина, як правило, має тільки один помітний сплеск. Схема стає ще більш помітною

та зрозумілою після контролю за загальною кількістю публікацій протягом дня (рис.3.5);

- фейкові новини розповсюджуються далі, швидше і ширше, ніж правдиві новини, оскільки глибина каскаду, максимальна ширина і розмір каскадів фальшивих новин, як правило, більше, ніж справжніх новин, тоді як час, необхідний для каскадів фейкових новин, досягає будь-якої глибини і розміру менше, ніж для каскадів правдивих новин (рис.3.5);
- фейкові новини пов'язані за політикою розповсюджуються далі, швидше і ширше, ніж звичайні фейкові новини серед інших областей, де тенденції між фейковими політичними новинами та просто фейковими новинами схожі до тенденцій між помилковими новинами та справжніми новинами [10] (рис.3.6).

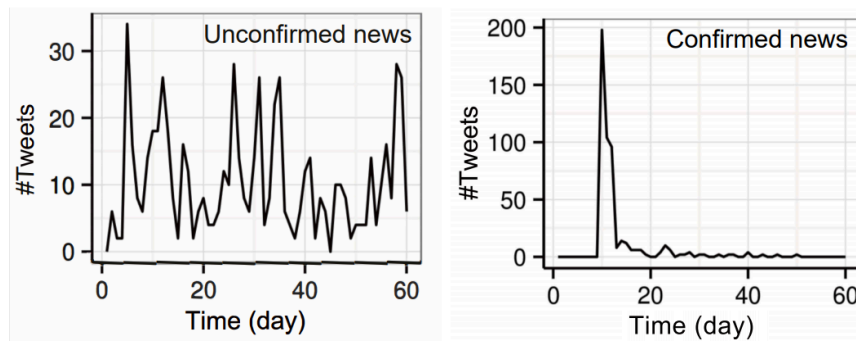


Рисунок 3.4 – Порівняння графіків появи новин протягом дня.

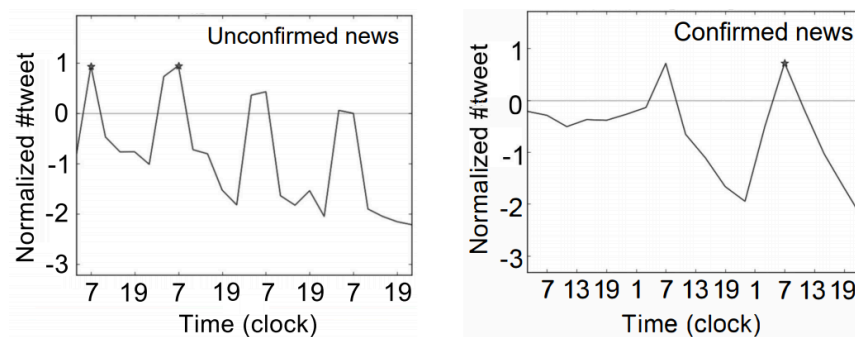


Рисунок 3.5 – Порівняння графіків появи новин протягом часу.

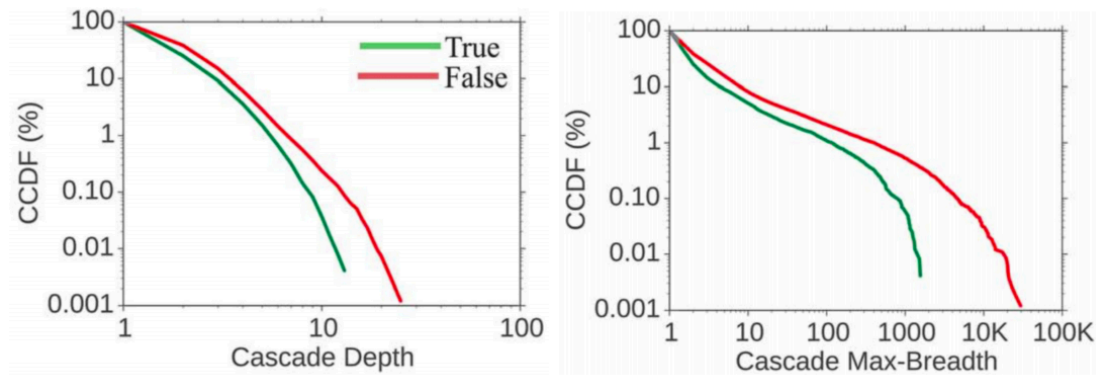


Рисунок 3.6 – Графіки кореляції глибини каскаду з його шириною.

Аналіз показав, що фейкові новини подорожують все далі і швидше, ніж справжні новини. Також політичні фейки мають ще більшу швидкість розповсюдження у порівнянні з іншими фейковими новинами.

### 3.3 Метод розпізнавання фейкових новин на основі стилю написання

Подібно до того, як фейкові новини аналізуються використовуючи знання, вивчення недостовірних публікацій з точки зору стилю написання також стосується аналізу змісту цих новин.

Проте дослідження, засновані на знаннях, мають на меті оцінити достовірність якоїсь новини, тоді як дослідження засновані на стилях, мають на меті оцінити намір новин, тобто чи є намір ввести в оману громадськість чи ні?

Формально, поняття “стиль” щодо фейкових новин можна визначити як набір кількісно визначених характеристик, які можуть якимось чином характеризувати фейкові новини і відрізнити фейкові новини від істини.

Хоча розвиток досліджень стосовно стилю написання фейкових новин все ще перебуває на ранніх стадіях, аналіз та виявлення обману вже давно є активним

напрямком досліджень, які загалом зосереджені на загальному стилі оманливого (тобто навмисно неправдивого) змісту різних типів інформації.

Далі розглянемо існуючі дослідження стосовно обману, які охоплюють різні типи інформації, наприклад, спілкування онлайн, відгуки, а також новини. Ці дослідження зазвичай не враховують притаманних новинам характеристик стилю їхнього написання.

### 3.3.1 Аналіз і виявлення обману

Аналіз обману спрямований на дослідження стилю написання змісту різних типів інформації, наприклад, онлайн-комунікації, відгуки, заяви, есе та текст, зображення та відео, а також фейкові новини.

Дослідження обману стосуються головним чином теорій стилю обману, тобто того, чому стиль змісту може допомогти дослідити обман, особливостей та моделей, заснованих на стилі, які можуть добре представляти та захоплювати обман, та стратегії виявлення обману

Тобто як можна використовувати стиль для виявлення фейкових новин та інших типів дезінформації.

Інтуїтивно, стиль змісту оманливої інформації (наприклад, фейкові новини), який має на меті обдурити читачів (наприклад, з перебільшеними виразами та сильними емоціями), має дещо відрізнитися від правдивої.

Дійсно, криміналістичні психологічні дослідження (наприклад, гіпотеза Ундеуча, моніторинг реальності, теорія міжособистісних обманів і чотирифакторна теорія) показали, що твердження, отримані на основі фактичного досвіду, відрізняються за змістом і якістю від тих, які базуються на фантазіях.

Ці інтуїції та фундаментальні теорії мотивують і роблять можливим дослідження стилю обману, будь то для тверджень або заяв, онлайн-повідомлень, онлайн-оглядів або новостей.

Результати виявлення обману з використанням стилю змісту, ще більше підтвердили обґрунтованість цих теорій, причому показники точності виявлення обману в проведених експериментах коливаються в межах від 60% до 90%.

Як зазначено вище, стиль змісту зазвичай представлений набором кількісно визначених характеристик, часто можливостей для машинного навчання.

Як правило, ці функції можуть бути згруповані в функції мови на основі атрибутів (attribute-based language features) або функції мови на основі структури (structure-based language features).

### 3.3.2 Функції мови на основі атрибутів

Також відомі як теоретично-орієнтовані функції мови, в основному надихаються або безпосередньо впливають з пов'язаних вище згаданих теорій обману.

Наприклад, ознака “сенсорний коефіцієнт” фіксує явище, що пояснюється моніторингом реальності, що недостовірні події виражаються більш низькими рівнями сенсорної інформації порівняно з істинними подіями.

Базуючись на нашому дослідженні, атрибутивні мовні особливості, що описують стиль змісту, можуть бути згруповані за десятьма паралельними вимірами: кількість, складність, невизначеність, суб'єктивність, невідповідність, почуття, різноманітність, неформальність, специфічність і читабельність.

Хоча Функції мови на основі атрибутів можуть бути дуже доречними, зрозумілими та передбачуваними, вони часто бідні (або менш гнучкі) у кількісному

визначенні стилю вмісту обману порівняно зі структурно-орієнтованими. Зокрема, функції, що базуються на атрибутах, часто вимагають додаткових рівнів кількісного визначення або обчислення, що може зайняти багато часу і надає більшого значення правильній оцінці характеристик і фільтрації для виявлення обману.

Щоб дослідити, які категорії ознак на основі атрибутів є найбільш релевантними для виявлення обману, був проведений аналіз досліджень, що використовують такі особливості в різних категоріях.

Обманна інформація демонструє більш високі рівні кількості, непрямоті, неформальності при цьому нижчі рівні різноманітності і специфічності. Вважається, що існує сильна потреба у більш систематичних дослідженнях з визначення найбільш інформативних та теоретично орієнтованих функцій, які можуть найкраще охопити обман у змісті.

### 3.3.3 Функції мови на основі структури

Структурно-орієнтовані мовні характеристики описують стиль вмісту з чотирьох мовних рівнів:

- лексикон;
- синтаксис;
- семантика;
- дискурс.

Структурно-орієнтовані характеристики також відомі як техніко-орієнтовані характеристики (*technique-oriented features*), оскільки їх кількісне визначення здебільшого спирається на зрілі технології обробки природних мов (НЛП), отже, вони не залежать від тематики досліджень або областей.

Зокрема, основним завданням на рівні лексики є оцінка частотної статистики букв, слів тощо, які можуть бути проведені відповідним чином за допомогою n-грамових моделей.

На синтаксичному рівні завдання дрібного синтаксису виконуються за допомогою модулів Part-of-Speech (POS)-таггерів, що полегшують розмітку та аналіз POS. Завдання глибокого рівня синтаксису виконуються програмою Probabilistic Context-Free Grammars (PCFG) з деревами розбору, які дозволяють робити аналіз контекстно-вільної граматики (CFG).

На семантичному рівні доволі часто застосовується Linguistic Inquiry and Word Count (LIWC) для забезпечення близько 80 семантичних класів для семантичних характеристик.

Нарешті, Rhetorical Structure Theory (RST) і засоби риторичного розбору зв'язують риторичні відносини як характеристики на рівні дискурсу [9]. Хоча в дослідженнях обману ці особливості є менш доречними, зрозумілими та передбачуваними, їх обчислення є легким порівняно з атрибутивними.

Щоб оцінити відносну необхідність і важливість ознак на різних мовних рівнях, ми далі аналізуємо продуктивність (точність) серії досліджень виявлення обману, які включають особливості на більш ніж одному мовному рівні, що показано

Для кожного дослідження, треба виділити особливості в межах або між рівнями мови, які досягають оптимальної продуктивності в явній формі, і підкреслюємо найвищу продуктивність, яку можуть досягти функції на одному рівні мови, що дозволяє зробити такі висновки:

- в рамках єдиного мовного рівня функції лексиконного рівня майже завжди працюють краще порівняно з синтаксичними, семантичними або дискурсивними характеристиками, тобто для досліджень, які включають характеристики мовних рівнів, 11/14 з них досягають кращої продуктивності на рівні лексики;

- поєднання функцій між мовними рівнями майже завжди призводить до кращих результатів, ніж використання функцій на одному рівні мови, тобто, з дванадцяти досліджень, що містять характеристики в межах та між мовними рівнями, вісім з них працюють краще за допомогою функцій крос-мовного рівня.

Хоча такі аналізи дають певні погляди, очевидно, що необхідні більш ретельні експерименти для систематичної оцінки відносної важливості структурних функцій на різних мовних рівнях.

## ВИСНОВКИ

З ростом швидкості створення та поширення новин у соціальних мережах, сьогодні дуже перспективною є розробка автоматичних систем перевірки фактів серед публікацій в соціальних медіа.

У ході роботи було проведено аналіз предметної області, знайдено та досліджено автоматично засоби і алгоритмів розпізнавання фейкових новин у соцмережах.

Був проведений аналіз предметної області, в процесі якого були визначені основні проблеми, що існують в розглянутій галузі. На підставі аналізу предметної області була проведена постановка завдання.

Проаналізовані проблеми сучасних підходів до розпізнавання фейкових новин такі як: великі витрати людського часу на ручну роботу, маскування та постійна адаптація фейкових новин. На основі проведеного аналізу оптимальних алгоритмів і методів виявлення і розпізнавання фейкових новин був обраний метод, який дозволяє зменшити ймовірність помилкових виявлень і підвищити якість оцінки правдивості публікацій.

У результаті роботи було розглянуто сучасні методи розпізнавання та боротьби з фейковими новинами у соцмережах. Проаналізовано ручні та автоматичні методи оцінювання, та досліджено ефективність кожного з методів. Був обраний перспективний метод для задачі розпізнавання недостовірних даних серед публікацій в соціальних мережах.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Інформаційна війна: пропаганда російських ЗМІ про події на Сході України // Інформаційне агентство “Корупція інфо”. – 2014. URL: <http://www.korupsiya.com/?p=18255> (дата звернення 19.02.2019).
2. Моль А. Социодинамика культуры // А. Моль: Прогресс, 1973. – 224 с
3. Леонов Н. С. Информационно-аналитическая работа в загранучреждениях // Николай Сергеевич Леонов. – М., 1996. – 173 с.
4. Як розпізнати фейкову новину в соцмережах – рекомендації The Huffington Post // Media Sapiens. – 2016. – 23.11. URL: <http://osvita.mediasapiens.ua/> (дата звернення 15.03.2019).
5. Even smart people are shockingly bad at analyzing sources online. This might be an actual solution // NiemanLab. – 2017. – 13.10. URL: <https://www.niemanlab.org/2017/10/even-smart-people-are-shockingly-bad-at-analyzing-sources-online-this-might-be-an-actual-solution> (дата звернення 08.04.2019)
6. A year in fake news, and what to look forward to in 2018 // NiebanLab. – 2017 – 15.12. URL: <https://www.niemanlab.org/2017/12/a-year-in-fake-news-and-what-to-look-forward-to-or-how-to-tune-out-in-2018> (дата звернення 08.04.2019)
7. Sarah Cohen, James T Hamilton, Fred Turner. Computational journalism. ACM 54, 2011. 66–71
8. Maximilian Nickel, Kevin Murphy, Volker Tresp, Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. 2016. 11–33
9. D Pisarevskaya. Rhetorical Structure Theory as a Feature for Deception Detection in News Reports in the Russian Language. In Artificial Intelligence and Natural Language & Information Extraction, Конф. ФРУКТ, Санкт-Петербург, Росія. 2015.
10. Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science 359, 6380 (2018), 1146–1151.