

КОМП'ЮТЕРНИЙ АНАЛІЗ ДЖЕРЕЛ ІНФОРМАЦІЇ У РЕГІОНАЛЬНИХ ЗМІ ОДЕЩИНИ

СТЕБЛИНА Наталія,

канд. наук із соц. комунік., доц.,

Донецький національний університет імені Василя Стуса, м. Вінниця, вул. 600-річчя, 21, 21021, Україна, e-mail: n.steblyna@donnu.edu.ua.

У дослідженні запропоновано комп'ютерний метод аналізу «живих» джерел інформації у журналістських текстах, описано його застосування та встановлена ефективність, створено комп'ютерну програму (мова Python). Для перевірки роботи комп'ютерної програми вивчено контент двох регіональних одеських сайтів – «Думская» та «Одесская жизнь». Порівняння роботи програми та обробки текстів вручну показало незначний відсоток помилок (7,8 % та 11 % відповідно). За допомогою програми було окреслено провідні джерела інформації на сайтах «Думская» та «Одесская жизнь». Також було вивчено ситуації, в яких програма помилялася. Як правило, більшість із них виникала через недотримання журналістами професійних стандартів.

Ключові слова: джерела новин, комп'ютерний аналіз, регіональні масмедіа.

NEWS SOURCES COMPUTER ANALYSIS OF REGIONAL ODESA MASS MEDIA

Computer method of human news sources is proposed in the research, it's implementation is described, efficiency is evaluated. A computer program (Python language) is created. For the program testing content of two regional sites is studied: «Dumskaya» and «Odessa Life». The program and manual processing results comparison shown an insignificant percent of mistakes (7,8% and 11% respectively). With the help of the program the leading news sources of «Dumskaya» and «Odessa Life» sites were defined. The cases of the program mistakes were studied as well. As a rule, the majority of the cases can be explained with the professional standards violations by journalists.

Key words: news sources, computer analysis, regional mass media.

Вступ. Як правило, джерела інформації у публікаціях українських ЗМІ обробляються вручну [13, с. 102; 14, с. 94]. Так, наприклад, згідно з методологією Інституту демократії імені Пилипа Орлика, походження матеріалів визначають експерти, аналізуючи джерела інформації у журналістських текстах [9]. Однак, оскільки для посилення на так звані «живі» джерела журналісти використовують типові формулювання на кшталт «сказав Прізвище» чи «за словами Прізвище», існує можливість формалізувати пошук таких даних і, відповідно, створити комп'ютерну програму, яка буде використовувати ці формулювання для пошуку. Це дослідження якраз і є спробою обґрунтування подібного алгоритму аналізу.

Актуальність такої роботи полягає в тому, що автоматизація процедури аналізу джерел інформації дасть змогу опрацьовувати значні корпуси текстів у регіональних ЗМІ, визначати провідних ньюзмейкерів, робити висновки щодо рівня збалансованості висвітлення подій, замовчування певних тем та джерел і, навпаки, просування певних політиків, чиновників.

Новизна дослідження: у роботі пропонується новий алгоритм аналізу джерел інформації (автоматичний, а не ручний), розробляється комп'ютерна програма, а також описується її робота й особливості обробки отриманих результатів.

Дослідження пов'язане із сучасними науковими та практичними завданнями, зокрема у сфері політичних комунікацій, професійних стандартів журналістики, ви-

явлення маніпуляцій у ЗМІ. Так, у низці досліджень, проведених ще в доцифрову добу, було доведено, що навіть в демократичних країнах існує «ієрархія доступу» [16, с. 10] або ж «інформаційна гегемонія» [6], за якої еліти краще представлені в новинах. Автоматизована обробка журналістських текстів із виявленням джерел інформації дасть змогу звернути увагу на те, чи змінюється ця ситуація (диспропорційна презентація політиків, лідерів думок, громадянських активістів) у цифрові часи. Окрім цього, згідно з «індексною теорією» [17], опозиційним голосам приділяється мало уваги, особливо під час кризових ситуацій (війна, катастрофа тощо). Цікаво, що й у цифрову епоху, коли у всіх, хто має смартфон, є можливість долучитися до суспільного діалогу, ця проблема наявності альтернативних голосів у ЗМІ залишається актуальною. Так, через тиск дедлайнів, фінансові проблеми, журналісти все частіше працюють з офісу [4], сповідуючи так звану «екранну культуру» [21], за якої кількість безпосередніх контактів людини зменшується, натомість світ пізнається опосередковано – через екран. При цьому розквітає явище «пресрелізму», що призводить до того, що голоси суспільства не репрезентовані, а представники влади різних рівнів домінують як у друкованих, так і онлайн-масмедіа [12]. Інколи говорять навіть про «кризу журналістики» [19], нездатність ЗМІ виконувати свої суспільні функції за цифрових часів.

Тому вивчати способи використання джерел інформації у журналістських текстах сьогодні – надзвичайно важливо. Звісно, актуальним лишається й пошук оптимального алгоритму, який би допоміг досліднику вивчити це питання швидко й ефективно. Тим не менш, у науковій літературі, де пропонуються методи комп'ютерного аналізу, під поняттям «джерело новини» чи «джерело інформації» розуміється повідомлення певного ЗМІ (телебачення, інформаційне агентство тощо) [10, с. 89; 11, с. 58; 7, с. 27] або допис у соціальній мережі [8], IP-адреса [2]. Часом досліджуються лінки на джерела [20] (технічно це легко зробити через аналіз html-коду). Тобто йдеться не про джерела інформації у тексті, а про текст як джерело інформації. Джерела новин можуть виділятися й за допомогою процедури анотування текстів – як, наприклад, у дослідженні Н. Шаховської та колег [15]). І у такому випадку, окрім основних елементів тексту, може виділятися джерело інформації, але все ж комп'ютерний алгоритм тут також не є розробленим у достатній мірі.

Тож оскільки методу комп'ютерного аналізу для дослідження «живих» джерел у журналістських текстах, принаймні в українському науковому дискурсі, ще не було запропоновано, зробимо таку спробу. Але все ж наголосимо на тому, що окреслене питання розв'яземо лише частково. У цьому дослідженні йтиметься не про всі джерела інформації у певному тексті, а лише про «живі».

Відразу окреслимо певні обмеження, які виникають під час комп'ютерного аналізу журналістських текстів. Кожен ЗМІ може використовувати свої правила для посилань на джерела, а також загалом – для викладу матеріалу. Тож, з одного боку, маємо професійні стандарти (а з-поміж них стандарт достовірності, що вимагає чіткого посилання на джерело), які мають бути спільними для всіх, а з іншого – унікальний стиль ЗМІ, а також – кожного журналіста, який пише текст. Тому, звісно, повинні розуміти, що комп'ютерна програма не дасть 100%-го результату. І для кожного ЗМІ треба коригувати регулярні вислови, що допомагають шукати потрібні фрагменти тексту.

Якщо говорити про загальнонаціональні масмедіа, наприклад, «Укрінформ» чи «Українську правду», то розв'язання поставленого нами завдання цілком можливе, оскільки тексти там створюються за певним шаблоном. У деяких випадках, як-от для «Української правди», ситуація спрощується через те, що з середини 2019 року видання перейшло на новий формат текстів і тепер у більшості випадків вказує на походження інформації окремим підрозділом у тексті (наприклад, «Джерело: фейсбук-сторінка Дмитра Кулеби»). Деякі світові видання взагалі можуть використо-

вувати для посилання на джерела окремих html-тег, тож за таких умов можливо встановити джерела інформації для кожного тексту (за умови, якщо посилання коректне).

Однак у регіональних ЗМІ ситуація не настільки проста. Деякі з них, наприклад, змішують у стрічці новин як новини, так і матеріали інших жанрів, де посилання на джерела не настільки формалізоване. Деякі мають «власні» професійні стандарти посилань на джерела. Тому для цього дослідження візьмемо саме масмедіа такого типу.

Мета – здійснити комп'ютерний аналіз джерел інформації у ЗМІ одеського регіону за квітень 2020 року та інтерпретувати отримані дані з точки зору дотримання професійного стандарту балансу думок.

Завдання дослідження: розробити комп'ютерну програму для аналізу «живих» джерел у ЗМІ; описати процедуру аналізу отриманих результатів, визначити випадки неправильної ідентифікації джерел і на основі цього виміряти ефективність роботи програми на матеріалі регіональних ЗМІ; опрацювати результати роботи програми, визначити перелік найбільш уживаних джерел інформації та виявити збалансованість чи незбалансованість подання інформації.

Об'єкт дослідження – регіональні онлайн-ЗМІ (популярні сайти Одеси).

Предмет дослідження – формалізований (комп'ютерний) аналіз «живих» джерел інформації у ЗМІ.

Для дослідження візьмемо два сайти Одеси: «Думская» із 2.85 млн переглядів за місяць (інтернет-видання) та «Одесская жизнь» (інтернет-видання та газета) із 234 тис. переглядів. Таким чином, ми зіставимо в нашому аналізі найбільш популярне видання, на редакційну політику якого здійснює вплив власник, і найбільш популярне незалежне видання. Як відомо, перший сайт пов'язують з українським народним депутатом О. Гончаренком, другий сайт є не асоціюється ні з ким із політиків і належить підприємцю Г. Чабанову [1]. Обидва сайти мають стрічку новин, що періодично оновлюється. У цій стрічці представлені не тільки інформаційні, але й аналітичні повідомлення. Штат видання складають власні кореспонденти. Тож подібний добір ЗМІ дасть змогу отримати релевантні дані, а також зіставити результати обох сайтів.

Одеський регіон було взятий тому, що тут кількість регіональних ЗМІ – одна з найбільших в Україні (більше 100 новинних ресурсів).

Для кожного з сайтів була сформована вибірка. Для завантаження матеріалів видання «Одесская жизнь» використовувався веб-архів, бо на момент написання статті було неможливо отримати тексти прямо з сайту. Таким чином, через веб-архів і було досягнуто потрібної мети: за квітень завантажено 52 тексти новин. Архів зберіг скріншоти сайту за кожен день, однак час збереження був різним: ніч, ранок, день і вечір. Тому бралися поперемінно нічні, ранкові та інші тексти. Із сайту «Думская» вдалося завантажити 1021 текст, і вже з цих текстів було сформовано вибірку (враховано кожен десятий текст) – усього 102 публікації (текстів у вибірці більше, тому що цей сайт розміщує більше новин, ніж попередній).

Отже, обсяг вибірки: 52 тексти із сайту «Одесская жизнь» та 102 тексти із сайту «Думская».

Методи дослідження: контент-аналіз із застосуванням комп'ютерної програми. Елементами аналізу виступили живі джерела інформації у текстах. Окрім цього, використовувався аналіз текстів на дотримання професійних стандартів, зокрема балансу думок – для визначення пропорційності уваги до живих джерел різних типів (офіційні джерела, фахівці, громадянські активісти, пересічні громадяни тощо). А також компаративний аналіз – для порівняння текстів двох видань.

Кожен текст оброблявся комп'ютерною програмою, після цього – вручну. Для комп'ютерного аналізу розроблялися регулярні вислови (модуль re, мова програму-

вання Python), які фіксували найбільш поширені формулювання для позначення живих джерел інформації.

Результати й обговорення. Спочатку опишемо регулярні вислови, які використовувалися для пошуку інформації.

1) Посилання після цитати.

Це найпоширеніший тип посилання, який, до того ж, знайти найлегше. Для цього програма шукає лапки, кому після лапок, тире та слово з великої літери наприкінці речення. Наприклад: *«Зроблена вакцинація проти COVID19. Зробив це на передовій з нашими солдатами в якості головнокомандуючого. Той же Oxford/AstraZeneca (Covishield) з Індії, який був доставлений першим в Україну і прийнятий мільйонами людей у світі. Вакцина дозволить нам знову жити без обмежень»*, – написав Зеленський.

2) Посилання в ліді.

За стандартами, у ліді має бути розміщено посилання на першоджерело чи джерело. І, як правило, журналісти також використовують для цього стандартні вирази: «про це повідомив», «як повідомив» тощо. Тож знаходити такі посилання можна з допомогою прийменників та наступного дієслова. А потрібне прізвище джерела стоїть або перед крапкою, або перед комою, або перед прийменником в/у.

3) Посилання в складнопідрядному реченні.

Ці посилання використовуються в тексті новини й містять дієслово з комунікативною інтенцією («сказав», «заявив» тощо) та сполучник «що». У таких формулюваннях потрібне прізвище стоїть перед дієсловом. Наприклад: Раніше Іван Русев повідомляв, що на кургані в Татарбунарському районі Одеської області розпустилися рідкі рослини.

4) Посилання з дієсловами, що містять комунікативну інтенцію.

Це дієслова (наприклад, «назвав», «вимагає», «пропонує») без сполучника. Розпізнати таке джерело можна, уклавши список подібних дієслів. У такому випадку програма також шукатиме прізвище перед дієсловом. Наприклад: *Президент України Володимир Зеленський запропонував холдингу DP World з Об'єднаних арабських еміратів інвестувати в розвиток українських портів*

```
import re
for od in range(355):
    exampleFile = open(r'news_item (%s).txt' % (od + 1), encoding='utf-8')
    exampleFile1 = exampleFile.read()
    with open(r'sources.txt', 'a', encoding='utf-8') as file:
        source_r = re.compile(r', [—-] \w+[ија]н\w? {[A-R]E3ГA-Z}[w+]}')
        i = source_r.findall(exampleFile1)
        print('\n'.join(i), file=file)
        source_r = re.compile(r', — \w+[ея]т {[A-R]E3ГA-Z}[w+]}')
        e = source_r.findall(exampleFile1)
        i.extend(e)
        source_r = re.compile(r', [—-] \w+[ея]т\w? .*? {[A-R]E3ГA-Z}[w+]}')
        f = source_r.findall(exampleFile1)
        i.extend(f)
        source_r = re.compile(r'[Oo]б атом .*? {[A-R]E3ГA-Z}[w+] {[A-R]E3ГA-Z}[w+]}[-фон]')
        n = source_r.findall(exampleFile1)
        i.extend(n)
        source_r = re.compile(r'Как \w+[ија]н\w? .*? {[A-R]E3ГA-Z}[w+]}[-,вно]')
        q = source_r.findall(exampleFile1)
        i.extend(q)
        source_r = re.compile(r'[Пп]о словам {[A-R]E3ГA-Z}[w+]}')
        s = source_r.findall(exampleFile1)
        i.extend(s)
        source_r = re.compile(r'[Пп]о мнению {[A-R]E3ГA-Z}[w+]}')
        t = source_r.findall(exampleFile1)
        i.extend(t)
        source_r = re.compile(r'со ссылкой на .*? {[A-R]E3ГA-Z}[w+]}')
        v = source_r.findall(exampleFile1)
        i.extend(v)

source_r = re.compile(r'([A-R]E3ГA-Z}[w+]) \w+[ија]н\w?, что')
x = source_r.findall(exampleFile1)
i.extend(x)
source_r = re.compile(r'([A-R]E3ГA-Z}[w+]) \w+[иое]т, что')
y = source_r.findall(exampleFile1)
i.extend(y)
source_r = re.compile(r', [сзопн]\w+[аи]н\w? .*? {[A-R]E3ГA-Z}[w+]}')
B = source_r.findall(exampleFile1)
i.extend(B)
source_r = re.compile(r'о данным .*? {[A-R]E3ГA-Z}[w+]}')
H = source_r.findall(exampleFile1)
i.extend(H)
source_r = re.compile(r'[Пп]о информации .*? {[A-R]E3ГA-Z}[w+]}')
J = source_r.findall(exampleFile1)
i.extend(J)
source_r = re.compile(r'([A-R]E3ГA-Z}[w+]) \w+[зв]ан\w?')
L = source_r.findall(exampleFile1)
i.extend(L)
source_r = re.compile(r'копеекноид\w+ .Одессо\w+ жикан')
T = source_r.findall(exampleFile1)
i.extend(T)
source_r = re.compile(r'копеекноид\w+ .Думо\w+')
U = source_r.findall(exampleFile1)
i.extend(U)
V = ('\n'.join(i))
r = V.split('\n')
rr = set(r)
print('\n'.join(rr), file=file)
print('-----%s-----' % (od + 1), file=file)
```

Рисунок 1 – Комп'ютерна програма для аналізу «живих» джерел інформації у текстах регіональних ЗМІ

5) Вислови «за словами», «за даними», «за інформацією».

Ці вислови також часто використовують журналісти, відтак знайти потрібне прізвище можна з використанням цього виразу та останньої власної назви перед комою. Наприклад: *За словами директора департаменту дорожнього господарства Андрія Сандула, оновлення та ремонту потребують не тільки одеські дороги.*

На основі окреслених регулярних висловів було укладено комп'ютерну програму (рис. 1).

Тепер опишемо процедуру отримання результатів. Спочатку було проведено автоматизований збір даних, після цього тексти ЗМІ оброблялися вручну. Після порівняння результатів для сайтів «Думская» та «Одесская жизнь» було отримано такі результати (табл. 1):

Таблиця 1. Ефективність роботи програми

Сайт	«Живі» джерела виявлені правильно	«Живі» джерела виявлені неправильно або частково неправильно
Думская	94 (92,2%)	8 (7,8%)
Одесская жизнь	46(88%)	6 (11%)

Отже, на підставі порівняння автоматизованого та ручного збору даних можемо зазначити, що результати комп'ютерного аналізу не дають змоги визначення усіх джерел новин, Але все ж похибка – незначна. Таким чином, головних героїв новин, яких цитують журналісти, виявити можливо.

Розглянемо випадки неправильного визначення джерел:

1) розміщення не новинних жанрів у стрічці (замальовок, нарисів, колонок, коментарів тощо). Згідно з журналістськими стандартами, новинний і не новинний контент має бути чітко розмежований (див. «Редакційні настанови «Bi-Bi-Ci» [18]). Обидва видання все ж не дотримуються цих рекомендацій, тому, звісно, тут регулярні вислови не працюють.

Наприклад, в одному з текстів видання «Одесская жизнь» розповідається про візит журналістів на базар за черешнею. У тексті не названо прізвищ джерел інформації, тільки зазначено їхній статус («продавчиня», «покупець» тощо). Також немає вислову «кореспондент видання», який би могла зафіксувати програма. Зазначено тільки «відправились ми», «вивчаємо тему», «перше, що я побачила».

2) немає збігу зі словником дієслів, що висловлюють комунікативну інтенцію. Знову ж таки, згідно із журналістськими стандартами, у новинних жанрах журналісти повинні використовуватися обмежена кількість дієслів на позначення заяв або інших коментарів джерел. Найбільш нейтральними вважаються «казати», «говорити» [5]. Однак все ж українські журналісти не завжди дотримуються цього правила й використовують інші дієслова, тому до комп'ютерної програми було додано ще й такі: «повідомити», «написати», «назвати», «просити», «вимагати», «коментувати», «висловити», «пропонувати». Тим не менш, інколи журналісти аналізованих видань використовували й інші дієслова. Наприклад: «згадує», «сподівається». Тобто такі дієслова, які в одних випадках можуть виражати комунікативну інтенцію, а в інших – ні. Тому їх не було додано до словника, відтак – і джерела цитати не було ідентифіковано.

3) подвійні посилання – інколи журналісти посилалися відразу на два джерела, приписуючи їм одну й ту ж саму цитату. Це також може розглядатися як порушення професійних стандартів, оскільки одна цитата має стосуватися одного конкретного

джерела. У таких випадках програма або знаходила останнє згадане джерело, або не знаходила нічого.

Тепер розглянемо, як саме можна використовувати подібний аналіз. Ось як виглядає результат роботи програми (рис. 2).

Абрамченко	
Колдорара	
Правоохранитель	
-----1-----	
Николай	
Мы	
корреспондент «Думской	
-----2-----	
Мачулка	
корреспондент «Думской	
-----3-----	
-----4-----	
-----5-----	
Зеленский	
Он	
Хомчак	
-----6-----	
ГУНП	
-----7-----	
-----8-----	
-----9-----	
-----10-----	
Хомчак	
-----11-----	
Позднякова	
Она	
-----12-----	
-----13-----	
корреспондент «Думской	
-----14-----	
-----15-----	
Геннадий	
-----16-----	
-----17-----	
Украины	
-----18-----	
-----19-----	
корреспондент «Думской	
-----20-----	
Беляков	
Белякова	
-----21-----	
Шыгала	
-----22-----	
-----23-----	
Ножевникова	
Ляшко	
-----24-----	

У першому тексті маємо два джерела: мера міста Ізмаїл Абрамченка та лідера ромської громади – Колдорара. Програма також знайшла іменник «правоохоронець». Воно стосувалося головного поліцейського району Кучугури. Програма не зафіксувала перше згадування через дієслово «обурився», якого немає у словнику, бо воно не є нейтральним і не обов’язково виражає комунікаційну інтенцію («Кучугура обурився тим, що»). Проте знайшла слово «правоохоронець», бо воно було написано з великої літери, а після нього йшло дієслово «заявив».

Тексти № 2, 3, 14, 20 написані з посиланням на власного кореспондента. У другому тексті згадано місцевого жителя Миколу, який прокоментував ситуацію. У третьому – посилання на водія трамваю Мачулку. У тексті № 14, окрім власного кореспондента, було згадано артиста Делієва, однак його прізвище не було зафіксовано, оскільки в першій згадці не було дієслова з комунікативним наміром, а в другому випадку був ужитий займенник: «Артист Георгій Делієв все-таки вийшов разом із друзями та колегами до Дюка. За його словами, «ніхто нікого не кликав...». Текст № 20 – на спортивну тематику, окрім кореспондента, ніяких джерел інформації не містив.

Тексти № 4, 5, 8, 9, 10, 13, 15, 17, 19, 23 не містили «живих» джерел. У них журналісти посилалися на пресслужбу, міські департаменти, документи тощо. Також було знайдено й недостовірні посилання: «Розповідають організатори» (без зазначення того, хто саме розповідає). Або ж нечітке посилання на джерело: «Пояснив «Думській» один з депутатів міськради, який проголосував за це рішення». У цьому випадку не зрозуміло, на яких умовах цей депутат коментував рішення й чому його прізвище не назване. Один із текстів (№19) був коментарем, у ньому також не містилося посилань на «живі» джерела.

Тексти № 6, 11, 12, 16, 21, 22, 24 містили цитати Президента Зеленського, головнокомандувача ЗСУ Хомчака, одесита Геннадія, директора зоопарку Белякова, директорки

Рисунок 2 – Вивід результатів програми

представлені місцеві чиновники та активісти. Також було вивчено ситуації, в яких програма помилялася. Як правило, більшість неточностей виникало через недотримання журналістами тих або інших професійних стандартів: розміщення в стрічці новин не новинних форматів, недостовірні посилання тощо.

У подальшому видається можливим використання запропонованої програми для вивчення контенту (новин) інших ЗМІ. А також – доопрацювання програми для пошуку інших типів джерел. Недоліками програми є те, що її потрібно переналаштувати для кожного нового сайту. Адже хоч формулювання для позначення джерел і є здебільшого універсальними, кожне масмедіа може мати свої стилістичні особливості. Також програма не розпізнає прізвища, інші власні назви та просто слова, написані з великої літери, тож під час обробки результатів потрібно відфільтрувати це. Однак усе ж і цю проблему можна розв'язати, розробивши додатково список стоп-слів для таких випадків.

1. Васіна О. Кому належать одеські ЗМІ. Аналіз Інституту масової інформації. Медіасцієнс. URL: <https://ms.detector.media/mediaanalitika/post/14936/2015-11-24-komu-nalezhat-odeski-zmi-analiz-instytutu-masovoi-informatsii/> (дата звернення: 28.06.2021).

2. Головенько В. Система виявлення джерела інформаційного впливу в інтернет-трафіку. URL: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/25942/%D0%A2%D0%B5%D0%B7%D0%B8%20%D0%93%D0%BE%D0%BB%D0%BE%D0%B2%D0%B5%D0%BD%D1%8C%D0%BA%D0%BE%20%D0%91%D0%A1-18%D0%BC.pdf?sequence=1&isAllowed=y> (дата звернення: 28.06.2021).

3. Дбайлива влада, мовчазні громадяни та професійні стандарти регіональної журналістики. Аналітичний звіт за травень 2021 року (частина 3). Детектор медіа. URL: <https://idpo.org.ua/reports/4136-dbajлива-vlada-movchazni-gromadyani-ta-profesijni-standarti-regionalno%d1%97-zhurnalistiki-analitichnij-zvit-za-traven-2021-roku-chastina-3.html> (дата звернення: 28.06.2021).

4. Дейвіс Н. Новини пласкої землі. Київ : Темпора, 2011. 548 с.

5. Довженко О. Зізнався, заявив, пообіцяв: чому слід добирати слова в новинах. Medialab. URL: <https://medialab.online/news/terevenyut/> (дата звернення: 28.06.2021).

6. Лалл Д. Масмедіа, комунікація, культура: глобальний підхід. Київ : К. І. С., 2002. 264 с.

7. Ланде Д., Дармохвал О. Оперативність і передрук в мережевих джерелах новин. Правова інформатика. № 1(37). 2013. С. 24–29.

8. Мельник В., Стьопочкіна І. Виявлення джерела шкідливої інформації в соціальних мережах. URL: <https://ela.kpi.ua/bitstream/123456789/25295/1/S.118-121.pdf> (дата звернення: 28.06.2021).

9. Методологія оцінювання якості контенту в регіональних друкованих та он-лайн виданнях. ІДПО. URL: <https://idpo.org.ua/wp-content/uploads/2021/methodology-2021.pdf> (дата звернення: 28.06.2021).

10. Петренко О. С. Джерела інформації про партії та кандидатів-мажоритарників у ході передвиборчої кампанії 2012 року (на прикладі мешканців м. Луганська). Вісник Луганського національного університету імені Тараса Шевченка. № 11(1). 2013. С. 88–100.

11. Соболев А. Виявлення в глобальній мережі Інтернет інформаційних джерел, які розповсюджують недостовірну інформацію. Реєстрація, зберігання і обробка даних [Registration, storage and data processing]. № 21(3). 2019. 56–68.

12. Стеблина Н. Регіональні сайти – для кого і на кого працюють. Детектор медіа. URL: <https://detector.media/regionalna-presaeonlain/article/165060/2015-07-20-regionalni-sayty-dlya-kogo-i-na-kogo-pratsyuyut/> (дата звернення: 28.06.2021).

13. Стеблина Н. Репрезентація політичного режиму України у новинах онлайн-видання «Українська правда». Політичне життя. № 4. 2019. С. 99–108.

14. Чабаненко М. Фейсбук і агенційна журналістика в інтернеті (на прикладі новин сайту УНІАН). Наукові записки Інституту журналістики. № 1. 2017. 93–97.

15. Шаховська Н., Нога Р., Вовк О. Інтелектуальна система анотування новин для оцінювання достовірності їх джерел. Моделювання та інформаційні технології. № 72. 2014. С. 119–123.

16. Atton C. Alternative media. London : Sage, 2006. 185 p.

17. Bennett W. L., Lawrence R. G. Livingston S. When the Press Fails, Political Power and the News Media from Iraq to Katrina. Chicago : Chicago Press, 2007. 251 p.

18. Editorial Guidelines. BBC. URL: <https://www.bbc.co.uk/editorialguidelines/guidelines> (дата звернення: 28.06.2021).

19. Gibson, J. 2017. A crisis of relevance. Columbia Journalism Review. URL: https://www.cjr.org/special_report/news-janine-gibson-newspapers-audience.php (дата звернення: 28.06.2021).

20. Mehler A., Bao Y., Li X., Wang Y., Skiena S. Spatial analysis of news sources. IEEE Transactions on Visualization and Computer Graphics. No 12(5). 2006. P. 765-772.

21. van Dejk J. A. The Network Society. Social Aspects of New Media. London : Thousand Oaks, 2006. 292 p.

1. Vasina, O. (2015), «Who owns Odesa mass media. Institute of Mass Information Analysis», Mediasapience [Media sapiens], available at: <https://ms.detector.media/mediaanalitika/post/14936/2015-11-24-komu-nalezhat-odeski-zmi-analiz-instytutu-masovoi-informatsii/> (accessed: 28 June 2021).

2. Holovenko, V. (2019), «Information influence system of identification on the Internet-traffic», URL: [aztyy \[in Ukrainian\]](http://aztyy.in.ua/), (accessed: 28 June 2021).

3. IDPO, (2021), «Careful government, silent citizens and the professional standards of regional journalism. Analytic report for May 2021 (part 3)», Detector media [Media detector], available at: <https://idpo.org.ua/reports/4136-dbajлива-vlada-movchazni-gromadyani-ta-profesijni-standarti-regionalno-d1%97-zhurnalistiki-analitichnij-zvit-za-traven-2021-roku-chastina-3.html>, (accessed: 28 June 2021).

4. Davies, N. (2011), Flat Earth News, Kyiv, Tempora, 2011, 548 p.

5. Dovzhenko, O. (2019), «Confessed, stated, promised: why is it important to select words in news», Medialab, available at: <https://medialab.online/news/terevenyty/>, (accessed: 28 June 2021).

6. Lull, D. (2002), Mass Media, communication, culture: global approach, Kyiv, K. I. S., 264 p.

7. Lande, D. & Darmokhval, O. (2013), «Operativity and reprints in network news sources», Pravova informatyka [Legal informatics], vol. 1, no. 37, pp. 24–29.

8. Melnyk, V. & Stepochkina, I. (2018), «Sources of harmful information identification in social networks», available at: <https://ela.kpi.ua/bitstream/123456789/25295/1/S.118-121.pdf>, (accessed: 28 June 2021).

9. IDPO, (2019), «Evaluation of content quality methodology in regional print and online media», available at: <https://idpo.org.ua/wp-content/uploads/2021/methodology-2021.pdf>, (accessed: 28 June 2021).

10. Petrenko, O. (2013), «Sources of information about parties and majoritarian deputies during pre-election campaign 2012 (on example of Luhansk inhabitants)», Taras Shevchenko Luhansk national university Herald [Bulletin of Luhansk Taras Shevchenko National University], vol. 11, no. 1, pp. 88–100.

11. Sobolev, A. (2019), «Sources of unreliable information dissemination identification in global network Internet», Data Recording, Storage & Processing, vol. 21, no. 3, pp. 56–68.

12. Steblyna N. (2015), «Regional sites – for whom are they working», Detector media [Media detector], available at: <https://detector.media/regionalna-presaoain/article/165060/2015-07-20-regionalni-sayty-dlya-kogo-i-na-kogo-pratsyuyut/>, (accessed: 28 June 2021).

13. Steblyna, N. (2019), «Ukrainian political regime representation in news of online media «Ukrayins`ka Pravda»», Politychne zhyttya [Political Life], vol. 4, pp. 99–108.

14. Chabanenko, M. (2017), «Facebook and journalism of news agencies on the Internet (on the example of UNIAN site's news)», Naukovi zapysky Instytutu zhurnalistyky [Scientific Notes of the Institute of Journalism], vol. 1, pp. 93–97.

15. Shakhovska, N., Noga, P., & Vovk, O. (2014), «Intellectual system of news annotation for news reliability evaluation», Modelyuvannya ta informacijni tehnologiyi [Modeling and Information Technologies], vol. 72, pp. 119–123.

16. Atton, C. (2006), Alternative media, London, Sage, 185 p.

17. Bennett, W.L., Lawrence, R.G. Livingston, S. (2007), When the Press Fails, Political Power and the News Media from Iraq to Katrina, Chicago, Chicago Press, 251 p.

18. Editorial Guidelines, BBC, available at: <https://www.bbc.co.uk/editorialguidelines/guidelines>.

19. Gibson, J. (2017), «A crisis of relevance. Columbia Journalism Review», available at: https://www.cjr.org/special_report/news-janine-gibson-newspapers-audience.php (accessed: 28 June 2021).

20. Mehler, A., Bao, Y., Li, X., Wang, Y., & Skiena, S. (2006), «Spatial analysis of news sources», *IEEE Transactions on Visualization and Computer Graphics*, no. 12, vol. 5, pp. 765-772.

21. van Dejk J.A., (2006), *The Network Society. Social Aspects of New Media*, London, Thousand Oaks, 2006, 292 p.

UDC 007: 645.063

NEWS SOURCES COMPUTER ANALYSIS OF REGIONAL ODESA MASS MEDIA

Steblyna Nataliia, PhD (Social Communications), Assistant Professor,
Vasul Stus Donetsk National University, Vinnytsya, 21, 600-richchya Str., 21021, Ukraine, e-mail: n.steblyna@donnu.edu.ua.

ORCID: <https://orcid.org/0000-0001-9799-9786>

Introduction. As a rule, news sources are identified in journalistic texts manually. However as far as journalists use typical wordings for news sources references, it's possible to create regular expressions and to design a computer program for automatic identification of the data.

Relevance of the study. The program is important for news sources usage study in the digital reality, in particularly of 'access hierarchy' and 'information hegemony' defining. It may be additionally significant for some practical tasks achieving: news monitoring, professional standards compliance analysis. Research objective is to observe news sources exploiting in the regional mass media with computer analysis, define leading people sources and evaluate balance of the news.

Methodology. Two local news sites were chosen for the analysis: Dymkaya and Odesa Life. At first a sample of texts was formed and regular expressions for the program were tested. References to the quotes, 'communicative' verbs and other expressions were coded. Content analysis was used, news sources were difened as elements of the analysis.

Results. The program and manual processing results comparison shown an insignificant percent of mistakes (7,8% and 11% respectively). The reasons of the program's mistakes were: incorrect identification of the news sources in analytics, blogs and columns, included in news feeds; larger vocabulary of the communicative verbs vocabulary; double references. In some cases the program shown pronouns and nouns (not just surnames). The latter problem may be solved with the list of spot words implementing. With the help of the program the leading news sources of Dumskaya site were defined: own correspondent, the president of Ukraine, the prime minister and other officials. All-Ukrainian news sources prevailed, while local authorities, activists, experts were not so popular. In the case of Odesa Life there was more balanced approach: local authorities and activists were represented.

Conclusions. Thus, it's quite possible to design the automatic approach to news sources study. The computer program was proposed and tested for this aim. In the majority of cases the program's mistakes may be explained with the professional standards violation by the journalists (lack of accuracy in the references, analytics and columns placement in the news feeds etc). The limitations of the program: the need of resetting for every new site, because of some stylistic peculiarities and editorial policy. The analysis has shown the level of balance in the regional mass media: the texts form independent media outlet were more balanced.

Key words: *news sources, computer analysis, regional mass media.*

Стаття надійшла до редакції 28.06.2021.