

УДК 004.738.5

О.В. Самчишин,
кандидат технічних наук
І.О. Орищук

МЕТОДИКА СКЛАДАННЯ РЕЙТИНГУ ЕЛЕКТРОННИХ ЗАСОБІВ МАСОВОЇ КОМУНІКАЦІЇ ПРИ ОРГАНІЗАЦІЇ ПРОЦЕСУ ЇХ КОНТЕНТ-МОНІТОРИНГУ

У статті розглянуто проблематику ранжування електронних засобів масової комунікації – джерел інформаційних повідомлень з метою оптимізації їх вибору для вирішення завдань моніторингу інформаційного простору. Запропоновано методику складання рейтингу електронних засобів масової комунікації, в основу якої покладено метод парних порівнянь та метод експертних оцінок. Такий підхід, на відміну від відомих, забезпечує урахування важливості обраних критеріїв моніторингу залежно від поставлених завдань.

Ключові слова: електронні засоби масової комунікації, метод парних порівнянь, метод експертних оцінок, ранжування джерел інформаційних повідомлень.

В статье рассмотрено проблематику ранжирования электронных средств массовой коммуникации – источников информационных сообщений с целью оптимизации их выбора для решения мониторинга информационного пространства. Предложена методика составления рейтинга электронных средств массовой коммуникации, в основу которой положено метод парных сравнений и метод экспертных оценок. Такой подход, в отличие от известных, обеспечивает учет важности выбранных критериев мониторинга в зависимости от поставленных задач.

Ключевые слова: электронные средства массовой коммуникации, метод парных сравнений, метод экспертных оценок, ранжирование источников информационных сообщений.

The paper deals with the issues of ranging the electronic means of mass communication as the sources of data messages in order to optimize their choice for solving of monitoring problems in the information space. It is suggested the rating method of electronic means of mass communication based on the paired comparison method and expert assessments. This approach, unlike the known ones, ensures the importance of selected monitoring criteria in the terms of assigned tasks.

Key words: electronic means of mass communication, paired comparison method, expert assessment, ranging of information message sources.

Постановка проблеми

Сучасна епоха інформатизації усіх сфер діяльності суспільства характеризує електронні засоби масової комунікації (е-ЗМК) як основне джерело розповсюдження новинної, оглядової та іншої інформації. Регулярний моніторинг таких матеріалів державними та приватними аналітичними структурами дозволяє не тільки аналізувати стан справ у будь-якій сфері інтересу, а й відкриває можливості прогнозування розвитку ситуації.

Моніторинг інформаційних повідомлень (далі – ІП) в е-ЗМК силовими структурами будь якої розвиненої держави спрямований, як правило, на своєчасне виявлення загроз в інформаційній сфері. При цьому основними завданнями моніторингу е-ЗМК є виявлення повідомлень заданого змісту в засобах масової інформації, блогосфері і соціальних мережах та їх джерел, аналіз тенденцій розвитку породжуваних негативними повідомленнями ситуацій за актуальними тематиками, як основи для побудови логічних послідовностей щодо виявлення можливих інформаційних компаній протиборчої сторони.

У процесі організації моніторингу е-ЗМК важлива роль відводиться функції каталогізації інформаційних повідомлень з подальшим їх накопиченням в спеціалізованій базі даних за визначеними тематиками, що становлять інтерес [1]. Однак ключову роль у процесі організації моніторингу е-ЗМК відіграє процедура оптимального вибору джерел ІП. Ця процедура є складноформалізованою, оскільки на вибір джерел ІП впливає достатньо велика кількість факторів, а тому потребує подальшого уточнення та доопрацювання

Аналіз останніх досліджень і публікацій

Аналіз найбільш відомих вітчизняних та іноземних інформаційно-аналітичних систем InfoStream, Медіалогія, Silobreaker тощо та рейтингів джерел ІП, які вони формують, показав, що перелік показників, на основі яких складаються рейтинги, дуже широкий [3–5]. Так рейтинг джерел ІП може складатися за регіональним, тематичним, мовним показником, за значимістю або впливовістю для тієї або іншої цільової аудиторії, обсягом цитування, об'єму реалізованої реклами тощо [1].

У зв'язку зі значною кількістю е-ЗМК, можливостей мережі Інтернет з динамічного розповсюдження ІП та обмеженість технічних ресурсів інформаційних систем щодо їх аналізу встановлено, що сьогодні виникає потреба оптимізації складу і кількості джерел за умови одночасного забезпечення відповідної якості вирішення завдань моніторингу. При цьому оптимізація складу і кількості джерел може досягатися за рахунок оптимального ранжування джерел ІП – складання відповідного рейтингу.

Метою статті є підвищення ефективності організації процесу моніторингу е-ЗМК за рахунок удосконалення методики складання їх рейтингу.

Викладення основного матеріалу дослідження

Відомо, що перелік показників, за якими можливо здійснити ранжування джерел ІП, може бути достатньо різноманітним [2]. Однак їх повне врахування недоцільне і обмежується завданнями, для яких складається рейтинг джерел ІП. Тому для вирішення задачі моніторингу е-ЗМК з метою виявлення і оцінювання, наприклад, рівня негативного інформаційно-психологічного впливу на цільову аудиторію доцільно обирати тільки ті джерела, які відповідають умовам регіональності, а тематика їх публікацій – відповідно завданням моніторингу. Крім того, варто обмежитись такими показниками, як: оригінальність, продуктивність, стабільність та індекс цитування джерела [1].

Основним та таким, що визначає зміст процедури вибору джерел ІП є завдання та мета моніторингу. Тому під час вибору джерел ІП обов'язковим є визначення *ступеня відповідності джерела заданій тематиці* (K_m), яка, у свою чергу, прямо залежить від завдань моніторингу. Ступінь відповідності джерела тематиці, завданням та меті моніторингу K_m пропонується визначати методом експертних оцінок у двійковій формі, тобто:

$$K_m = \begin{cases} 1 & \text{відповідає} \\ 0 & \text{не відповідає} \end{cases}. \quad (1)$$

Так, вибір для моніторингу найбільш оригінальних джерел ІП за виразом (1) дозволяє зменшити не тільки формальне але й змістовне дублювання інформації. Наприклад, для вирішення цієї проблеми необхідно уникати вибору до списку джерел моніторингу новинних інтеграторів.

Критерій оригінальності джерела ІП (K_{opd}) пропонується визначати як співвідношення кількості повідомлень, що самостійно формуються та публікуються джерелом до загальної кількості публікацій цього джерела за одиницю часу, тобто

$$K_{opd} = \frac{n_{opd}}{N_d} \quad (2)$$

де n_{opd} – кількість оригінальних (ексклюзивних) ІП d -го джерела; N_d – загальна кількість ІП d -го джерела.

Одним з критеріїв відбору джерела є відбір за кількістю публікацій, який використовується для максимального охоплення переліку ІП. Проведені інформаційним центром “Елвісті” дослідження показують що приблизно 20 % найбільш продуктивних джерел публікують 80 % документів [6]. Але виникає протиріччя – чим більш продуктивне джерело, тим більше воно містить запозичених з інших джерел повідомлень. Це знов породжує проблему дублювання ІП та ускладнює процес їх статистичної і аналітичної обробки. Тому пропонується відбір джерел за кількістю публікацій проводити з оригінальних, а їх кількість обмежувати шляхом ранжування за кількістю публікацій та відбору з них найбільш продуктивних.

Обчислення *показника продуктивності* d -го джерела серед обраних пропонується здійснювати за формулою:

$$K_{npd} = \frac{N_d}{\sum_{d=1}^D N_d}, \quad (3)$$

де D – загальна кількість обраних для ранжування джерел.

З практики [2] відомо, що при організації моніторингу є-ЗМК серед множини проблем підбору і аналізу джерел ІП суттєве значення має врахування показника стабільності джерела, тому *тематична стабільність* джерела може бути визначена як кореляція наборів тематичних рубрик, яким відповідають ІП з цього джерела в різні періоди часу [7]. Для обчислення рівня відхилення (нестабільності) K_{cm} джерела ІП пропонується скористатися критерієм, що заснований на лінійній метриці, а саме

$$K_{cm} = \frac{1}{Q} \sum_{i=1}^Q \frac{S_i}{R_i}, \quad (4)$$

де Q – кількість рубрик джерела; S_i – середньоквадратичне відхилення за i -ю рубрикою за визначений період часу; R_i – діапазон значень ІП за i -ю рубрикою.

З виразу (4) видно, що значення K_{cm} враховує не тільки кількість тематичних відхилень, але й відхилення за кількістю попадань у рубрики, а саме фактичну кількість ІП від джерел, що належать до цієї рубрики. У [7] показано, що виключну стабільність мають, як правило, джерела, що містять не більше 6 рубрик.

Прикладом стабільних джерел є великі інформаційні агенції, що регулярно поставляють користувачам приблизно однакові об'єми інформації протягом тривалого часу.

Прикладом нестабільних джерел можуть служити джерела, що активно діють протягом декількох діб, а потім припиняють свою роботу, публікують інформацію хаотично в часі як за кількістю, так й об'ємом повідомлень.

Нестабільні джерела також можуть представляти інтерес під час виконання специфічних задач моніторингу, але вони не формують основні тенденції в інформаційному просторі. Їх доцільно використовувати при несистематичному аналізі інформаційного простору.

На сьогодні найбільш розповсюдженим критерієм ранжування як окремих документів, так і окремих веб-сайтів є кількість посилань на них інших джерел – індекс цитування.

Індекс цитування є-ЗМК – це доля посилань на матеріали цього джерела в загальній кількості посилань на ресурси визначеного переліку джерел.

Для оцінки рівня цитування окремих документів і веб-ресурсів у мережевих пошукових системах використовують відому модель А. Бредера, критерій PageRank, SALSA, h-індекс та інші [8]. Однак при оцінюванні рівня цитування новинних веб-ресурсів як джерел ІП необхідно враховувати ряд умов, які передбачають некоректним використання зазначених вище моделей. Зокрема через підвищену динаміку новинних потоків, відсутність прямих гіперпосилань на деякі Web-ресурси, необхідність врахування у новинних потоках не тільки гіперпосилання, а й контекстних посилань на об'єкти відкритої і закритої частини Web-простору, відсутність врахування дублювання інформації.

При дослідженні цитування документів джерела ІП іншими веб-ресурсами, проведеного інформаційним центром “Елвісті” [6] було вперше запропоновано підхід, у якому для кожного ІП, яке належить до визначеного джерела – веб-сайту, виявляються вихідні посилання на інші джерела та отримується розподілення новинних джерел за кількістю веб-сайтів, що мають на них посилання.

Приклад ранжування списку новинних джерел подано в таблиці 1.

Таблиця 1

Список новинних джерел

Web-сайт	Кількість web-сайтів, що посилаються	Ранг джерела
ІА “Інтерфакс”	1051	1
“РосБізнесКонсалтинг”	983	2
“Reuters”	882	3
ІТАР-ТАСС	787	4
РІА “Новости”	773	5
УНІАН	675	6
Радіо “Свобода”	662	7
НТВ	631	8
“Коммерсант”	623	9
ВВС	598	10

Обчислення значення показника цитування d -го джерела пропонується здійснювати як відношення кількості посилань на інформаційні повідомлення цього джерела до загальній кількості посилань на ресурси визначеного переліку джерел згідно з критерієм

$$K_{\text{цд}} = \frac{L_d}{\sum_{d=1}^D L_d}, \quad (5)$$

де L_d – кількість посилань на повідомлення d -го джерела.

З аналізу найбільш популярних новинних е-ЗМК (див. табл. 1) встановлено, що числове значення запропонованих показників залежно від джерела може змінюватись у досить великому діапазоні. Діапазон залежить від кількості ІП на даному сайті та періоду визначеному для моніторингу (від декількох ІП до десятків тисяч). Тому в процесі організації моніторингу е-ЗМК необхідно вводити інтервальну шкалу (табл. 2) нормованих коефіцієнтів однакових показників різних джерел з метою додержання метричності.

Таблиця 2

Інтервальна шкала нормованих коефіцієнтів
для побудови рейтингу е-ЗМК

Назва показника	Визначення коефіцієнта	Інтервал зміни значень показника за K_i критерієм ранжування	Ранг показника
Відповідність джерела тематиці	$K_m = \begin{cases} 1 & \text{відповідає} \\ 0 & \text{не відповідає} \end{cases}$		
Показник оригінальності джерела	$K_{\text{опд}} = \frac{n_{\text{опд}}}{N_d}$	0-0,01 0,01-0,1 0,1-0,2 0,2-0,3 0,3-0,4 0,5-0,6 0,6-0,7 0,7-0,8 0,8-0,9 0,9-1	1 2 3 4 5 6 7 8 9 10
Показник продуктивності джерела	$K_{\text{нпд}} = \frac{N_d}{\sum_{d=1}^D N_d}$	0-0,01 0,01-0,1 0,1-0,2 0,2-0,3 0,3-0,4 0,5-0,6 0,6-0,7 0,7-0,8 0,8-0,9 0,9-1	1 2 3 4 5 6 7 8 9 10

Назва показника	Визначення коефіцієнта	Інтервал зміни значень показника за K_i критерієм ранжування	Ранг показника
Показник стабільності джерела	$K_{cm} = \frac{1}{Q} \sum_{i=1}^Q \frac{S_i}{R_i}$	0-0,01	1
		0,01-0,1	2
		0,1-0,2	3
		0,2-0,3	4
		0,3-0,4	5
		0,5-0,6	6
		0,6-0,7	7
		0,7-0,8	8
		0,8-0,9	9
		0,9-1	10
Індекс цитування джерела	$K_{ud} = \frac{L_d}{\sum_{d=1}^D L_d}$	0-0,01	1
		0,01-0,1	2
		0,1-0,2	3
		0,2-0,3	4
		0,3-0,4	5
		0,5-0,6	6
		0,6-0,7	7
		0,7-0,8	8
		0,8-0,9	9
		0,9-1	10

Залежно від завдань моніторингу завжди вважається за необхідне визначати і враховувати важливість коефіцієнтів однакових показників різних е-ЗМК. При цьому важливість кожного коефіцієнта буде також різною. Так як оцінювання показників за обраним критерієм проводиться для достатньо великої кількості джерел (обмежених тільки технічними ресурсами посту моніторингу), то для досягнення заданої точності оцінювання пропонується скористатися методом парних порівнянь [9].

Вектор важливості визначених критеріїв $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)$ характеризується важливістю відповідного коефіцієнта. Тобто $\alpha_i \geq \alpha_j$, якщо критерій f_i має перевагу над критерієм f_j . При цьому

$$\sum_{k=1}^K \alpha_k = 1, \quad \alpha_k \geq 0. \quad (6)$$

Складання матриці парних порівнянь проводиться за умови, що всі діагональні елементи дорівнюють одиниці, а значення решти елементів $K = [K_{ij}]$ групою експертів визначається за таких умов:

2 – якщо критерій i важливіший ніж критерій j ;

1 – якщо критерій i і j мають однакову важливість;

0 – якщо критерій i менш важливіший ніж критерій j .

Підраховується рівень важливості кожного коефіцієнта K_i як

$$K_i = \sum_{j=1}^n K_{ij}. \quad (7)$$

Визначається загальний рівень важливості коефіцієнтів усіх критеріїв K_c , тобто

$$K_c = \sum_{i=1}^n K_i. \quad (8)$$

Розраховується важливість визначених коефіцієнтів згідно з виразом

$$\alpha_i = \frac{K_i}{K_c}. \quad (9)$$

Таким чином, визначивши кількісні значення нормованих коефіцієнтів обраних показників та розрахувавши їх важливість, можна провести оптимальне ранжування джерел ІІ (скласти рейтинг е-ЗМК) відповідно до поставлених завдань, використавши при цьому метод вагових коефіцієнтів за допомогою адитивної згортки.

У цьому випадку будується цільова функція вигляду

$$f(X) = K_m \sum_{i=2}^5 \alpha_k K_i(X). \quad (10)$$

й вирішується завдання побудови рейтингу е-ЗМК за оптимізацією скалярного критерію, тобто

$$z = f(X) \rightarrow \max, \text{ за умови } X \in D, \quad (11)$$

де D – кількість обраних для ранжування джерел, X – кількість е-ЗМК вибраних для вирішення конкретної задачі моніторингу.

Наведемо практичний приклад застосування запропонованої методики (1) – (11) складання рейтингу е-ЗМК при організації моніторингу.

Вхідні дані. При постановці завдання на моніторинг визначено деяку тематику, що становить інтерес. Кількість джерел, за якими можна проводити моніторинг, дорівнює 40 новинним сайтам. Після розрахунку відповідності джерел тематиці поставленого завдання K_m , отримано 5 е-ЗМК, які необхідно розташувати за рейтингом відповідно до завдань моніторингу. За виразами (2) – (5) для кожного джерела розраховано визначені коефіцієнти та значення показників K_i , тобто

1-джерело – $K_{opd} = 2, K_{npd} = 3, K_{cm} = 6, K_{ud} = 4,$

2-джерело – $K_{opd} = 7, K_{npd} = 5, K_{cm} = 2, K_{ud} = 4,$

3-джерело – $K_{opd} = 3, K_{npd} = 8, K_{cm} = 4, K_{ud} = 6,$

4-джерело – $K_{opd} = 9, K_{npd} = 1, K_{cm} = 6, K_{ud} = 3,$

5-джерело – $K_{opd} = 4, K_{npd} = 3, K_{cm} = 5, K_{ud} = 1.$

За допомогою групи з 5-ти експертів розраховано важливості визначених критеріїв методом парних порівнянь. Дані наведено в табл. 3.

Таблиця 3

Рівень важливості коефіцієнтів K_i для першого е-ЗМК

	K_{opd}	K_{npd}	K_{cm}	K_{yd}	K_i
K_{opd}	1	0	2	2	5
K_{npd}	2	1	2	2	7
K_{cm}	0	0	1	0	1
K_{yd}	0	0	2	1	3

Тоді значення важливості розглянутих коефіцієнтів за формулою (9) дорівнює:

$$\alpha_{K_{opd}} = \frac{5}{16} = 0,3125; \quad \alpha_{K_{npd}} = \frac{7}{16} = 0,4375; \quad \alpha_{K_{cm}} = \frac{1}{16} = 0,0625; \quad \alpha_{K_{yd}} = \frac{3}{16} = 0,1875.$$

Далі обчислено адитивну згортку на основі методу вагових коефіцієнтів для кожного джерела за формулою (10):

$$\begin{aligned} f(1) &= 1 \times (0,3125 \times 2 + 0,4375 \times 3 + 0,0625 \times 6 + 0,1875 \times 4) = 3,06; \\ f(2) &= 1 \times (0,3125 \times 7 + 0,4375 \times 5 + 0,0625 \times 2 + 0,1875 \times 4) = 5,25 \\ f(3) &= 1 \times (0,3125 \times 3 + 0,4375 \times 8 + 0,0625 \times 4 + 0,1875 \times 6) = 5,81 \\ f(4) &= 1 \times (0,3125 \times 9 + 0,4375 \times 1 + 0,0625 \times 6 + 0,1875 \times 3) = 4,19 \\ f(5) &= 1 \times (0,3125 \times 4 + 0,4375 \times 3 + 0,0625 \times 5 + 0,1875 \times 1) = 3,13 \end{aligned}$$

З приведених даних випливає те, що для визначеного завдання на моніторинг джерела ІІІ розподілені за таким рейтингом (табл. 4).

Таблиця 4

Рейтинг електронних засобів масової комунікації при організації контент-моніторингу

Позиція в рейтингу	Номер е-ЗМК	Значення цільової функції
1	3-джерело	5,81
2	2-джерело	5,25
3	4-джерело	4,19
4	5-джерело	3,12
5	1-джерело	3,06

Отже, запропонована методика складання рейтингу електронних засобів масової комунікації забезпечує оптимізацію складу і кількості джерел шляхом

складання рейтингу за найбільш впливовими показниками відповідно до цілей і задач моніторингу, що забезпечує підвищення ефективності процесу організації моніторингу е-ЗМК, зменшення залучених людських та технічних ресурсів для вирішення поставлених завдань.

Висновки та перспективи подальших досліджень

Таким чином, в статті показано, що з метою визначення пріоритетності вибору джерел для організації їх моніторингу необхідно здійснювати їх ранжування. Доведено, що для побудови їх рейтингів доцільно скористатися рядом критеріїв, що розраховують відповідні показники залежно від задач моніторингу та характеристик джерел інформаційних повідомлень. Це дозволяє оптимізувати вибір джерел відповідно до задач моніторингу, виявляти першоджерела інформаційних повідомлень, значно скоротити витрати часу і засобів шляхом ігнорування або виключення з аналізу найменш інформативних джерел.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Грищук Р. В. Основи кібернетичної безпеки : монографія / Р. В. Грищук, Ю. Г. Даник ; за заг. ред. проф. Ю. Г. Даника. – Житомир : ЖНАЕУ, 2016. – 634 с.
2. Грищук Р. В. Особливості організації та ведення моніторингу електронних засобів масової комунікації / Р. В. Грищук, О. В. Манько, І. О. Орищук // Інформаційна Безпека. Східноукраїнський національний університет ім. Володимира Даля. – 2014. – № 3–4 (15–16). – С. 10–14.
3. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис: научно-методическое пособие / А.Н. Григорьев, Д.В. Ландэ, С.А. Бороденков та ін. – К., ООО “Старт-98”, 2007. – 40 с. [Електронний ресурс]. – Режим доступу : <http://infostream.ua/news36/>.
4. Информационно-аналитическая система Медиология [Електронний ресурс]. – Режим доступу : <http://www.mlg.ru/company/technologies/>.
5. Silobreaker [Електронний ресурс]. – Режим доступу : <http://www.silobreaker.com/>.
6. Ранжирование источников информации в системе мониторинга новостей InfoStream / Д.В. Ландэ, С.М. Брайчевский, А.Т. Дармохвал, А.Ю.Морозов // Труды 10-ой Всероссийской научной конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”- RCDL’2008, Дубна, Россия, 2008. – С. 213–219.
7. Ландэ Д.В. Стабильность источников как один из параметров информационных потоков / Д.В. Ландэ, А.Н. Григорьев, С.М. Брайчевский // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции “Диалог”. Вып.7 (14). – М. : РГГУ, 2008. – 649 с.
8. Основы моделирования и оценки электронных информационных потоков: монография / Д.В. Ландэ, В.Н. Фурашев, С.М. Брайчевский, А.Н. Григорьев. – К. : Инжиниринг, 2006. – 176 с.
9. Колomoец Ф.Г. Основы системного анализа и теории принятия решений: пособие для исследователей, управленцев и студентов вузов / Ф.Г. Колomoец. – Мн. : Тесей, 2006. – 300 с.

Отримано 27.10.2016

Рецензент Рибальський О.В., д.т.н.