УДК 378.147:004:34.08 DOI: 10.18372/2073-4751.86.21270

Гаврилов В. В.,
orcid.org/0009-0007-1542-9589,
e-mail: 1767320@stud.kai.edu.ua,

Рябий М. О., к.т.н.,
orcid.org/0000-0002-9651-9135,
e-mail: m.o.ryabyy@gmail.com

ОГЛЯД ЕФЕКТИВНИХ МЕТОДІВ ТА ІНСТРУМЕНТІВ ДЛЯ ПОБУДОВИ ІНФОРМАЦІЙНИХ СИСТЕМ МОНІТОРИНГУ ПОВІДОМЛЕНЬ У МАС-МЕДІА

Національний університет «Київський авіаційний інститут»

Вступ.

Останніми роками Україна стала центром уваги міжнародних медіа у зв'язку з подіями, що мають глобальне значення. Аналіз того, як ці події висвітлюються у закордонних засобах масової інформації (ЗМІ), допоможе зрозуміти міжнародну думку та позицію щодо України. Закордонні ЗМІ значно впливають на формування громадської думки та політичних рішень. Тому визначення частоти та контексту новин корисне для визначення настроїв громадян. В умовах гібридної війни, інформаційна безпека стає критично важливою. Вивчення того, як Україна представлена в закордонних медіа, допоможе виявити можливі інформаційні загрози та маніпуляції, спрямовані на дестабілізацію ситуації в країні або на міжнародній арені.

З точки зору наукових досліджень, вивчення міжнародного медіа-простору дозволить розширити знання про тенденції сучасної журналістики, що набуває особливої актуальності для дослідження.

Огляд джерел.

Побудови сучасного інформаційного простору, налагодження ефективної взаємодії та свободи слова в ЗМІ мають надзвичайно важливе значення для реалізації демократичних перетворень в Україні [1]. У праці [2, 3] автори зазначають, що український медіа простір характеризується суттєвою присутністю іноземного капіталу, значна частина засновників друкованих видань, теле- та радіоканалів мають іноземне походження, що є небезпечним для демократичного суспільства. Проте, автори не зазначають, що існує велика необхідність врахування специфіки засобу масової інформації та його спрямованості, ракурсу, уподобань аудиторії та інших параметрів, що складають формат видання чи каналу [4]. Аналіз та напрями інформаційного середовища дозволяють створити відповідний інформаційний простір з урахуванням зон «зіткнення інтересів» держави і влади потребують вироблення відповідних комунікаційних каналів [5]. Наразі актуальними є напрямки дослідження [6], які визначають рівень впливу, інструментарій, який для цього застосовується, зміни у сприйнятті медіа різного формату реципієнтами та зовнішні умови, у яких ці медіа циркулюють. Так, автори праці [7] визначають ступінь втручання держави у діяльність окремих ЗМІ та медіа-систему в цілому. Проте, автори не зазначають інтенсивність та прийоми регулярних повторень і подальшої бажаної інтерпретації повідомлень. У праці [8] досліджуються питання ефективного впливу повідомлень на велику кількість людей за короткі часові періоди, саме тому під час передвиборчих кампаній кандидати та партії максимум коштів вкладають у телевізійну рекламу. Наведений аналіз літературних джерел стосувався компаративного аналізу старих і нових медіа. Водночас, необхідним є розробка нової моделі побудови інформаційних систем з метою моніторингу повідомлень у мас-медіа.

Метою статті є розробка методики побудови інформаційної системи моніторингу мас-медіа через створенні єдиного автоматизованого середовища для оперативного виявлення, систематизації та аналізу інформаційних потоків.

Постановка задачі дослідження. Для реалізації поставленої мети сформуємо основні вихідні дані дослідження, що включають:

- новинні ресурси, такі як New York Times. Guardian;

- T - множина текстів, які знаходяться на заданих інформаційних ресурсах; необхідно за множиною текстів T провести аналіз частоти та контексту новин пов’язаних з Україною.

Для цього необхідно:

1) Отримати відповідні дані. Із T необхідно вилучити $T_{UA} \in T$, де T_{UA} - множина текстів, пов’язаних з Україною.

2) Оцифрувати дані.

Для проведення аналізу потрібно отримати числове представлення для T_{UA} .за допомогою функції:

$$\boldsymbol{f_{DIGIT}}: \boldsymbol{T} \to \boldsymbol{R}$$

$$V, t = f_{DIGIT}(T_{UA}),$$

де V - числове представлення текстів у вигляді векторів, N - кількість статей, t – відповідний час публікації статей.

3)Отримати оцінки емоційного контексту T_{UA} на основі V.

$$SC = f_{EST}(V)$$

$$SC = \{SC_{k,l}; 1 \le l \le N\}, 1 \le k \le 3,$$

$$SC_{\text{позитивна}} = SC_1 = \begin{pmatrix} pos_1 \\ \vdots \\ pos_N \end{pmatrix}$$

$$SC_{\text{негативна}} = SC_2 =$$

$$SC_{\text{нейтральна}} = SC_3 = \begin{pmatrix} neut_1 \\ \vdots \\ neut_N \end{pmatrix}$$

де SC - отримані оцінки, $SC_{\text{позитивна}}, SC_{\text{негативна}}, SC_{\text{нейтральна}}$ – вектори оцінок, pos_i- позитивна оцінка i -ої статі ${T_{UA}}^{(i)}$, $\neg \quad _i$ - негативна оцінка i -ої статі ${T_{UA}}^{(i)}$, $neut_i$ - нейтральна оцінка i -ої статі ${T_{UA}}^{(i)}$.

4) Провести аналіз оцінок

Оскільки для кожного набору значень оцінок $SC_{all}^{(i)} =$ існує відповідна дата публікації, можемо провести тренд аналіз, для ідентифікації та побудови тренду$f_{mp}(t)$на основі часових рядів та визначення ймовірно тей станів системи на основі марковських процесів.

5) На основі результатів аналізу надати рекомендації щодо реагування на інформаційний стан України у відповідних джерелах інформації.

Основна частина дослідження.

Необхідно отримати числове представлення V на основі текстів T_{UA} . Це завдання є частиною теорії обробки природного мовлення, для отримання оцінок SC використовують великі мовні моделі (англ. Large language models, LLM). Оскільки отримані оцінки залежні від часу, вводяться основні поняття та методи обробки часових рядів та марковських процесів. Розглянемо більш детально ці підходи.

Часові ряди

Випадковий процес – це функція від двох величини часу t та елементарної події ω.

У математиці випадковим процесом називається функція двох аргументів, значення якої – випадкові величини:

$$\zeta(t) = \varphi(\omega, t), \omega = 0,1, \dots, N, t = 0,1, \dots, l,$$

де N +1– кількість елементарних подій (воно може бути і нескінченним); t – параметр. Для кожного значення параметра t функція $\varphi(\omega, t)$ є функцією тільки від ω і, отже, становить випадкову величину. Для кожного фіксованого значення аргументу ω (тобто для кожної елементарної події) $\varphi(\omega, t)$ залежить тільки від t і є, таким чином, просто функцією одного дійсного аргументу. Кожна така функція є реалізацією випадкового процесу [9]. Часовий ряд – це частинний випадок випадкового процесу. Для часових рядів, на відміну від даних про часовий переріз, сама послідовність спостережень несе в собі важливу інформацію [10].

Характеристики випадкового процесу

1. Функція розподілу випадкового процесу.

$F(t,x) = P\{\xi(t) < x\}.$

Функція розподілу може змінюватися від конкретного t . Тому для загального випадку:

2. N - вимірна функція розподілу.

$F(t_1, \dots, t_N, x_1, \dots x_N) = P\{\xi(t_1) \leq x_1, \dots, \xi(t_N) \leq x_N\}.$

Але, здебільшого, під час практичного дослідження відмовляються від законів розподілу випадкового процесу на користь його **основних характеристик**:

Надалі припускаємо, що переріз випадкового процесу – є неперервна випадкова величина з функцією щільності $f(t,x)$.

3. На рис. 1 “жирна” лінія позначає функцію, навколо якої можливі різні реалізації. Ця функція називається математичним сподіванням випадкового процесу. Для отримання функції необхідний математичний опис випадкового процесу, до того ж для кожного виду випадкового процесу він може бути іншими.

Математичне сподівання. – функція $m(t)$, яка при будь якому значені t дорівнює математичному сподіванню його перерізу [10].

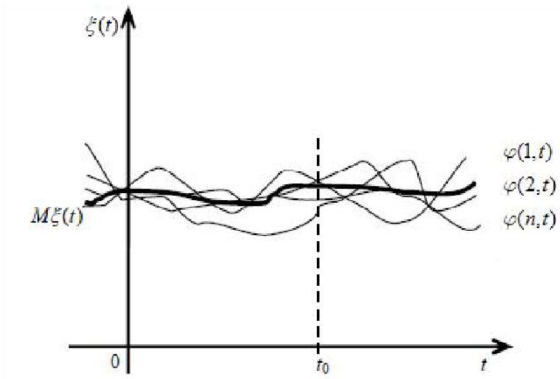

Рис. 1. Реалізація випадкового процесу $\zeta(t) = \varphi(\omega, t)$ [9]

$$m(t) = E\{\xi(t)\} = \int_{-} +xf(t,x)dx$$

4. Дисперсія – функція $D(t)$, яка при будь якому значені t дорівнює дисперсії перерізу. Дисперсія характеризує розсіювання можливих реалізацій випадкового процесу відносно середнього [10].

$$D(t) = D\left\{\left(\xi(t) - m(t)\right)^2\right\} = \int_{-} +\big(x - m(t)\big)^2 f(t,x)dx.$$

5. Коваріаційна функція – це функція від двох аргументів t_1 та t_2 , яка дорівнює коваріаційному моменту відповідних перерізів процесу [10].

$$K(t_1,t_2) = cov(t_1,t_2) = E\big\{\big(\xi(t_1) - m(t_1)\big)\big(\xi(t_2) - m(t_2)\big)\big\} = \int_{-} + \int_{-} +\big(x_1 - m(t_1)\big)\big(x_2 - m(t_2)\big) f(t_1,t_2,x_1,x_2)dx$$

6. Коваріаційна функція характеризує ступінь лінійної залежності між перерізами, а також розсіювання цих перерізів відносно математичного сподівання $m(t)$. При $t_1 = t_2 = t$ дорівнює дисперсії [10].

$$K(t,t) = E\left\{\left(\xi(t) - m(t)\right)^2\right\} = D(t).$$

У разі збільшення відстані між перерізами t_1 та t_2 – зменшується залежність між їх випадковими величинами (окрім деяких часткових випадків).

7. Кореляційна функція – нормована коваріаційна функція. Характеризує ступінь лінійної залежності випадкових величин [10].

$$R(t_1, t_2) = \frac{K(t_1, t_2)}{\sigma(t_1)\sigma(t_2)}$$

Властивості:

1) $R(t_1, t_2) = 1$

2) Симетрична відносно своїх аргументів. $R(t_1, t_2) = R(t_2, t_1)$

3) За модулем не перевищує одиницю. $|R(t_1, t_2)| \leq 1.$

Стаціонарний випадковий процес – процес, характеристики такого процесу незалежні від початкового відліку часу. Середня амплітуда та частота суттєво не змінюються. Такий процес відбувається в приблизно однорідних умовах та має вигляд неперервних випадкових коливань навколо деякого середнього значення. Кожен стаціонарний процес можна розглядати як такий, що продовжується у часі нескінченно довго. **Нестаціонарні** процеси характеризуються наявністю певної тенденції розвитку в часі. Такі процеси залежать від початкового відліку та від часу [10]. Не всі нестаціонарні процеси нестаціонарні протягом всього протягом всього розвитку. Стаціонарність у вузькому або широкому сенсі залежить від того, які характеристики процесу є залежними від початку відліку часу:

- У вузькому розумінні – якщо його N -вимірна функція розподілу не змінюється від зсуву усіх його спостережень на величину τ

- У широкому розумінні – якщо математичне сподівання постійне, а коваріаційна функція залежить не від положення першого аргументу, а від проміжку між першим та другим аргументами:

$$m(t) = const$$

$$K(t_1, t_1 + \tau) = \gamma(\tau)$$

Функція$\gamma(\tau)$від проміжку часу – називається автоковаріаційною, а проміжок часуτ називають зміщенням або запізненням. Стаціонарний процес у широкому розумінні є підмножиною процесу у вузькому розумінні. Якщо процес є нестаціонарним лише із-за змінного мат. сподівання, він може бути зведений до стаціонарного.

Модель та оцінка характеристик часового ряду

Часовий ряд – це дані, які фіксуються в часі та які розташовані в хронологічному порядку. Принципова відмінність часового ряду від вибірки полягає в тому, що члени ряду не є статистично незалежними, а також можуть бути не однаково розподіленими, тобто:

$$P\{\xi(t_1) \leq x\} \neq P\{\xi(t_2) \leq x\}, t_1 \neq t_2.$$

Виділяють чотири основні фактори, що можуть формувати значення часового ряду [10]:

1) Довготривала невипадкова монотонна функція, що формує тенденцію у зміні$x(t)$. Таку функцію називають трендом $f_{mp}(t)$. .

2) Сезонні фактори, які формують періодичні коливання у певні річні сезони. У аналітичному представлені описуються тригонометричними формулами.

3) Циклічні фактори, які формуються довготривалими циклами економічної, демографічної і т.п. природи. Описуються невипадковою функцією.

4) Випадкові фактори, які формують стохастичну природу елементів часового ряду.

Обов'язковим є лише четвертий фактор. Але може бути декілька факторів одного типу. Загальний вигляд часового ряду:

$$x(t) = \chi(A)f_{mp}(t) + \chi(B)\varphi(t) + \chi(C)\psi(t) + \varepsilon(t), t = 1,2, \dots, N$$

якщо, фактор «D» присутній, в іншому

$$\chi(D) = \begin{cases} 1, \\ 0, \\ D = \{A, B, C\}. \end{cases}$$

Дві задачі, які виникають при роботі з часовими рядами:

1) При наявності траєкторії часового ряду визначити невипадкові функції, оцінити їх параметри та підібрати модель, яка адекватно описує поведінку випадкових залишків.

2) На основі першої задачі здійснити прогноз значень часового ряду.

Модель випадкового ряду оцінюють із припущення про стаціонарність часового ряду.

Первинний аналіз часових рядів

Постановка задачі:

Часовий ряд представлений однією реалізацією випадкової величини, та результати спостереження подані масивом даних $X_N = \{x(t_i), i = \overline{1, N}\}$, де кожне спостереження це функція від t i -го. Спостереження над випадковим процесом проводяться рівномірно з заданим кроком $h = t_{i+1} - t_i$. У простішому вигляді $t_i = i$та записувати масив спостереження як звичайну вибірку $X_N = \{x_i, i = \overline{1, N}\}$.

Тренд – це функція, що показує глобальні зміни у досліджуваному процесі. Тренд функціонально визначається математичним сподіванням. Коли математичне сподівання є константою $H_0: m(t) = m = const$– тренд вважається відсутнім [10].

Для проведення якісного спектрального аналізу випадкових процесів необхідно при попередній обробці даних виявляти і вилучати тренд.

Природою тренда можуть бути:

1) детермінована складова фізичного процесу, який має деградацію (наприклад, процеси зносу, старіння, виходу з використання тощо);

2) похибка, що накопичується за рахунок складової (наприклад, похибка калібрування при вимірюваннях, яка має лінійний тренд);

3) наявність низькочастотних шумів при фіксації даних у часі;

4) сезонні коливання;

5) інші фактори.

Критерії для ідентифікації тренду

Розглянемо алгоритми визначення тренду, які не включають параметрів часового ряду. Наявність тренду означає нестаціонарність часового ряду, що означає статистичну незалежність спостереженьx_l[10].

• **Критерій знаків:**

Знаходимо кількість зростань значення спостережень. Зростання порівнюємо між поточним спостереженням та наступним.

$$y_i = \begin{cases} 1, & x_{i+1} > x_i, \\ 0, & x_{i+1} \le x_i \end{cases}$$

$i = \overline{1, n-1}$.

Кількість зростань:

$$c = \sum_{i=1}^{N-1} y_i,$$

статистики для цього критерію:

$$E\{c\} = (N-1)E\{y_i\} = \frac{1}{2}(N-1),$$

$$E\{c^2\} = \frac{1}{2}(N-1) + \frac{1}{3}(N-2) + \frac{1}{4}(N-2)(N-3),$$

$$D\{c\} = \frac{1}{2}(N-1) + \frac{1}{3}(N-2) + \frac{1}{4}(N-2)(N-3) - \frac{1}{4}(N-1)^2 = \frac{1}{12}(N+1).$$

Оскільки розподіл для величини c швидко збігається до нормального – порівнюємо статистику

$$C = \frac{c - E\{c\}}{\sqrt{D\{c\}}}.$$

із квантилем стандартного нормального розподілу $u_{1-a/2}$. Процес є стаціонарним за умови:$|C| \leq u_{1-a/2}$. Процес має тенденцію до зростання за умови: $|C| < u_{1-a/2}$, відповідно процес має тенденцію до спадання за умови:$|C| u_{1-a/2}$.

• **Критерій Манна:**

Обчислюють наступну величину

$$Mn\ Mn_{i,j} = \begin{cases} 1, & x_i < x_{j,} \\ 1/2\,, & x_i = x_j, \qquad i < j \\ 0, & x_i > x_j, \end{cases}$$

$$Mn = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Mn_{i,j}$$

Величина *Mn* має розподіл близький до нормального. З наступними статистиками:

$$E\{Mn\} = \frac{1}{4}N(N-1),$$

$$D\{Mn\} = \frac{1}{72}(2N+5)(N-1)N.$$

Після чого визначають статистику

$$u = \frac{Mn + 0.5 - E\{Mn\}}{\sqrt{D\{Mn\}}}.$$

Для заданого критичного рівня a, якщо $|u| \leq u_{1-a/2}$, то говорять, що досліджуваний процес стаціонарний. Інакше існує тенденція зміни процесу. Коли $u > u_{1-a/2}$,є тенденція до збільшення, а за $u < u_{a/2}$ - до зменшення.

• **Критерій серій:**

Передбачає побудову бінарного ряду спостережень у , що приймають значення:

$$y_i = \begin{cases} 1, & x_i \geq x'_m, \\ -1, & x_i < x'_m, \end{cases}$$

де x'_m, - медіана відсортованого ряду$\{x'_i; i = \overline{1,N}\}$

$$x'_m = \begin{cases} x'_{(N+1)/2}, & \text{коли} N - \text{непарне}, \\ \frac{1}{2}\left(x'_{N/2} + x'_{N/2+1}\right), & \text{коли} N - \text{парне}. \end{cases}$$

Сформований ряд $\{y_i; i = \overline{1,N}\}$характеризується послідовностями серій. Серія є сукупність розташованих підряд «1» або «-1». Позначимо$v(N)$ - загальну кількість серій в ряді, а$d(N)$- довжину найбільшої серії. Тоді на рівні значущості $a = 0{,}05$гіпотеза про стаціонарність процесу приймається, якщо одночасно виконуються умови

$$v(N) > \left[\frac{1}{2}\left(N + 1 - 1.96\sqrt{N-1}\right)\right],$$

$$d(N) < [3{,}3 \cdot lg(N+1)],$$

де $[\cdot]$ - де позначає цілу частину.

У противному разі робиться висновок про наявність залежності між спостереженнями й існування тренду.

Критерій «зростаючих» і «спадаючих» серій:

Базується на дослідженні ряду з «1» і «-1», утвореного за правилом

$$y_i = \begin{cases} 1, & x_{i+1} - x_i \geq 0, \\ -1, & x_{i+1} - x_i < 0. \end{cases}$$

Статистики $v(N)$ і $d(N)$ обчислюються аналогічно попередній процедурі, але мають задовольняти умовам

$$v(N) > \left[\frac{1}{3}(2N-1) - 1.96\sqrt{\frac{1}{90}(16N-29)}\right],$$

$$d(N) > d_o(N),$$

$$\text{де } d_0(N) = \begin{cases} 5, & N \leq 26, \\ 6, & 26 < N \leq 153, \\ 7, & N > 153. \end{cases}$$

• **Критерій Аббе:**

Критерій квадратів послідовних різниць і вимагає підрахування значення

$$\gamma = \frac{q^2}{2s^2},$$

де

$$q^2 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (x_i - x_{i+1})^2\,;$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2.$$

За $N \to \infty$ статистика

$$u = (\gamma - 1)\sqrt{\frac{N^2 - 1}{N - 2}}$$

асимптотично має нормальний розподіл. Тому за виконання | u |□ ^{u}1□□/2

гіпотеза про стохастичну незалежність результатів спостережень приймається.

Марковські процеси

Процес – послідовна зміна явищ, станів об’єкта в часі. [9, 10]

Випадковий процес – процес, протікання якого залежить від випадку та для якого визначена ймовірність того, чи іншого протікання.

Марковський процес – це випадковий процес, в якому ймовірність будьякого стану у будь-який момент часу не залежить від розвитку процесу на попередніх станах [11].

Найдослідженішими є такі типи випадкових процесів:

- процеси з незалежними приростами;
- процеси з некорельованими значеннями;
- нормальні випадкові процеси;
- марківські процеси.

У теорії масового обслуговування внаслідок специфічності характеру потоку запитів і часу обслуговування більшість задач вирішується із залученням теорії марківських процесів. Її перевагами є досконалий математичний апарат, можливість досліджувати як нестаціонарні, так і стаціонарні процеси, наочна інтерпретація результатів досліджень

Випадковий процес називається **марковським** (або процесом без післядії), якщо для кожного моменту часу t ймовірність будь-якого стану системи в майбутньому обумовлюється тільки її станом на сьогодні і не залежить від того, яким шляхом система набула цього стану. Параметр часу t може розглядатися і як дискретний, і як безперервний, унаслідок чого використовуються різні методи дослідження. [9]

Марковські процеси поділяють на процеси [11]:

- **З дискретними станами** – коли виокремлюються стани, які можна розрізнити та присвоїти порядковий номер. Система з дискретними станами характеризується миттєвою або стрибкоподібною зміною станів.
- **Неперервними станами** – характеризується плавним переходом у нову якість.

Марковські процеси також поділяють на процеси з:

- Дискретним часом – перехід в стан визначений в конкретний момент часу.
- Неперервним часом – перехід із стану в стан можливий у будь-який момент, наперед невідомий, часу.

При аналізі процесів з дискретними станами використовують **граф станів**. Такий граф відображає можливі стани досліджуваної системи та взаємозв’язки між ними (рис.2).

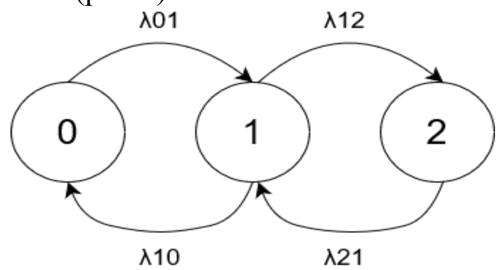

Рис. 2. Граф з трьома станами.

Стан в якому перебуває після k кроків позначають:

$$S_i^{(k)}, i = \overline{1, \infty}$$

де i – номер стану.

Випадкову послідовність подій або ж станів в яких опинилась система на кожному кроці процесу називають **марковським ланцюгом**, якщо для кожного кроку ймовірність переходу із довільного стану S_i в інший довільний стан S_j не залежить від того, як і коли система опинилась в стані S_i.

$$S_1^{(0)}, S_2^{(1)}, S_4^{(2)}, S_1^{(3)}, \ldots.$$

Для довільного кроку існують ймовірності переходу системи із одного стану в інший, причому деякі з таких ймовірностей дорівнюють нулю, якщо перехід зі стану в деякий інший неможливий.

Марковський ланцюг називають однорідним якщо ймовірність не залежить від номеру кроку.

Для однорідного ланцюга вводять матрицю ймовірностей переходу. Кожен елемент якої – ймовірність переходу та позначається [11]:

$$P_{ij} = P\left\{S_j^{(k)} / S_i^{(k-1)}\right\}.$$

Сума елементів в рядку дорівнює одиниці.

Ймовірність переходу визначається за наступною рекурентною формулою:

$$p_i(k) = \sum_{j=1}^{n} p_j(k-1) P_{ji}^{(k)}, i = \overline{1,n}.$$

Рівняння і початкові умови повністю описує ланцюг Маркова.

Якщо перехідні ймовірності не залежать від часу $\left(p_{ij}(t) = p_{ij}\right)$, то такий ланцюг називається однорідним (в іншому разі неоднорідним). Здебільшого застосовуються однорідні ланцюги Маркова. [12]

Інтенсивності переходу використовуються між станами замість ймовірностей переходу при наявності неперервного часу. [11]

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t},$$

де Δt - проміжок часу після моменту часу t .

$P_{ij}(\Delta t)$- ймовірність того, що протягом часу Δt система перейде зі стану S_i в стан S_j.

Якщо інтенсивність зміни станів системи не залежить від часу марковський процес називають **однорідним**, якщо ж маємо справу з функціональними залежностями то процес називають **неоднорідним**.

У випадку однорідного процесу визначення ймовірностей станів здійснюється за використанням рівнянь Колмогорова – системи звичайних диференційних рівнянь. В лівій частині кожного рівняння стоїть похідна ймовірності стану, а права частина містить кількість членів, що пов’язана з кількістю переходів в даний стан та переходів із цього стану. Мінус – якщо перехід відбувається із стану і з плюсом якщо.

При розгляді процесів, що протікають у системі з дискретними станами та неперервним часом можна вважати процес переходів системи зі стану в стан як такий, що зумовлений дією деяких потоків подій.

Множина моментів надходження в систему вимог називається вхідним потоком. Формально під випадковим потоком однорідних вимог розуміємо наступний стохастичний об’єкт [13].

Нехай (Ω, F, P) - ймовірнісний простір. Випадковим потоком однорідних подій називається випадкова послідовність

$$t_n = t_n(\omega), n = 1,2, \ldots, t_1 \le t_2 \le \cdots \le t_n \le \cdots$$

з такими властивостями:

а) для $\forall x$ та $\forall n \ge 1$ визначена ймовірність $P(t_n(\omega) < x)$, тобто $t_n(\omega)$ випадкова величина;

б) $P(t_n(\omega) < x) \to 0$ при $n \to \infty$ для $\forall x > 0$.

Властивості а), б) виключають нереальні випадки, коли за скінченний час може з’явитися нескінченно багато подій, чи коли ймовірність появи n –ої події раніше моменту х не визначена.

Аналогічно можна визначити скінченний потік однорідних вимог

$$\{t_n = t_n(\omega), 1 \le n \le N(\omega)\}.$$

Однак у більшості задач має $N(\omega) = \infty$ з ймовірністю 1. [14]

Потік подій називають **стаціонарним**, якщо ймовірність потрапляння певної кількості подій на відрізок часу довжиною □ залежить лише від довжини відрізку та не залежить від його розташування на вісі часу.

Якщо для будь-яких відрізків часу, що не перетинаються, кількість подій, що потрапили на один з них не залежить від кількості подій, що потрапили на інший, то такий потік подій називають **потоком без післядії.**

Потік подій називають **ординарним**, якщо ймовірність надходження двох або більше вимог одночасно, або, іншими словами, ймовірність потрапляння двох чи більше подій на елементарний відрізок часу, прямує до нуля.

Лема 1. Нехай потік однорідних вимог є стаціонарним та в ньому відсутня післядія. $P_0(t) = e^{-\lambda t}$ [11]

Теорема 1. Найпростіший потік є процесом з незалежними приростами, у якого прирости розподілені за законом Пуассона. [11]

Теорема 2. Найпростіший нестаціонарний потік представляє собою неоднорідний пуассонівський процес з ведучою функцією $\Lambda(t) = \int_0^t \lambda(u) du$,

де $\lambda(t)$ - змінний у часі параметр нестаціонарного потоку.[11]

Забезпечуючи зв'язки між елементами системи вищого порядку, процес комунікативної підсистеми становить семіотичну підсистему. Одночасно дана семіотична підсистема забезпечує зв'язки даної системи з її середовищем. При цьому ці засоби зв'язку можуть розглядатися як складова і системи, і її середовища. Це обумовлює особливий склад елементів семіотичної системи (підсистеми), їх характеристики, принципи взаємозв'язку, взаємодії, функціонування.

Висновки

Отже, вивчення основних аспектів управління контекстом та комунікаціями в інформаційному просторі, слід зауважити, що важливою складовою є дослідження повідомлень, оскільки вони стосуються низки соціальних та суспільних проблем. При цьому особливої уваги потребує аналіз елементів контексту, таких як ментальні концепти суб'єктів комунікації та соціальні моделі інтерпретації знаків, що використовуються під час комунікацій.

У цій статті було формалізовано постановку задачі дослідження, проаналізовано підходи до обробки та характеристики часових рядів, ідентифікації тренду (за критеріями Мана, знаків, серії, а також за критеріями спадаючої та зростаючих серій), побудови тренду, використання теорії марковських процесів. Подальші дослідження будуть пов'язані з розробленням і дослідженням ефективних методів щодо розпізнавання мови та обробки текстових значень з використанням цифрових технологій та автоматичних систем, а також архітектури нейронної мережі Transformer.

Література

1. Джига Т. В. Сучасний стан та основні проблеми взаємодії органів державної влади України із засобами масової інформації. Електронне урядування. 2010. № 1. С. 97-104.
2. Ковалевський В. Сучасний стан та тенденції розвитку медіа-сфери України. Політичний менеджмент. 2009. № 5. С. 109-119.
3. Матвєєнко І. В. Теоретичні аспекти впливів інформаційно-комунікативних технологій на громадянське суспільство. Науковий вісник Академії муніципального управління . 2010. №2. С.319-325.
4. Комунікація: демократичні стандарти в роботі органів державної влади / За заг. ред. Н. К. Дніпренко. К.:ТОВ «Вістка», 2008. 164 с. С.136
5. Чукут С .А., Загвойська О. В., Драгомирецька Н. М., Дрешпак В. М. Налагодження механізмів взаємодії уряду з неурядовими громадськими організаціями. К.: НАДУ, 2010. 64 с.
6. Русиняк А. Традиційні ЗМІ як актори процесу політичного конструювання суспільства. Вісник Прикарпатського університету. Політологія. 2020. Вип. 14. С. 74-84
7. Галлін Д., Манчіні П. Сучасні медіа-системи: три моделі відносин ЗМІ та політики. Київ: Наука, 2008. 320 с.
8. Царенко О. Політична культура як чинник впливу на політичну поведінку громадян. Панорама політологічних студій: Науковий вісник Рівненського державного гуманітарного університету. 2013. № 10. URL: http://nbuv.gov.ua/UJRN/Pps_2013_10_24
9. Литвинов А. Л. Теорія систем масового обслуговування. Харків, 2018. 142 с.
10. Приставка П. О. Статистичний аналіз даних. Київ, 2021. 154 с.
11. Аналіз часових рядів / А. Т. Яровий, Є. М. Страхов. Одеса: Освіта України, 2019. 109 с.
12. Приставка П. О. Методи штучного інтелекту. Київ, 2022. 63 с.
13. The math behind gated recurrent units. Towards data science. Режим доступу до ресурсу: https://towardsdatascience.com/the-mathbehind-gated-recurrent-units-854d88aded65
14. Бідюк П. І. Аналіз часових рядів: навчальний посібник. НТУУ «КПІ». Київ : НТУУ «КПІ», 2013. 600 с.

Гаврилов В. В., Рябий М. О.,

ОГЛЯД ЕФЕКТИВНИХ МЕТОДІВ ТА ІНСТРУМЕНТІВ ДЛЯ ПОБУДОВИ ІНФОРМАЦІЙНИХ СИСТЕМ МОНІТОРИНГУ ПОВІДОМЛЕНЬ У МАС-МЕДІА

В статті проведено аналіз основних методів та засобів для побудови інформаційних систем моніторингу мас-медіа. Розглянуто основні особливості системи, здатних обробляти великі масиви данних та провести аналіз частоти, контексту новин. Автором запропонована методика для збору та аналізу інформації з різнорідних джерел, а також проведена їх оцінка за допомогою методів обробки природної мови. Проаналізовано алгоритми розпізнавання іменованих сутностей, моделювання та автоматичного визначення повідомлень з використанням часових рядів, марковських процесів та критеріїв їх оцінки. Результати дослідження дозволяють підвищувати рівень точності та ідентифікації інформаційних систем через оперативне виявлення ризиків. Практичне значення статті полягає в використанні запропонованої методики як ефективного інструменту управління інформаційними системами, медіапростором, а також цифровими комунікаціями.

Ключові слова: *інформаційні системи, моніторинг мас-медіа, моніторинг повідомлень, часові ряди, маркаровські процеси, контент-аналіз, медіа-аналітика.*

Vladyslav Havrylov, Myroslav Riabyi

OVERVIEW OF EFFECTIVE METHODS AND TOOLS FOR BUILDING INFORMATION SYSTEMS FOR MONITORING MESSAGES IN THE MASS MEDIA

The article analyzes the main methods and tools for building information systems for monitoring mass media. The main features of the system are considered, capable of processing large data sets and analyzing the frequency and context of news. The author proposes a methodology for collecting and analyzing information from diverse sources, and also evaluates them using natural language processing methods. The algorithms for recognizing named entities, modeling and automatic message identification using time series, Markov processes and criteria for their evaluation are analyzed. The research results allow to increase the level of accuracy and identification of information systems through the operational detection of risks. The practical significance of the article lies in using the proposed methodology as an effective tool for managing information systems, media space, and digital communications.

Keywords: *information systems, mass media monitoring, message monitoring, time series, Markarovsky processes, content analysis, media analytics.*

Стаття подана до редакції: 04/05/2026
Стаття прийнята до опублікування: 12/05/2026
Стаття опублікована: 30/05/2026