

МЕТОДИКА ПРОГНОЗУВАННЯ РОЗВИТКУ ІНФОРМАЦІЙНОГО КОНТЕНТУ В МЕРЕЖІ ІНТЕРНЕТ

У статті запропоновано методику прогнозування розвитку інформаційного контенту в мережі Інтернет на основі використання векторного поля відношень для пошуку релевантних джерел та методу найменших квадратів. Розроблена програма базується на створенні для оператора середовища прогнозування розвитку контенту в результаті опрацювання вибраного запиту в режимі реального часу. Показано, яким чином методика та програма забезпечують виконання завдання попередження розвитку тих чи інших інформаційних подій.

Постановка проблеми. Важливою на даний час є здатність приймати адекватні рішення в умовах спрогнозованого перебігу подій, які б могли завадити їх виконанню. Побудова моделей попередження розвитку інформаційного контенту в інтернет-просторі та використання їх у комп'ютерних системах є однією з найважливіших проблем сьогодення.

Головне завдання на даний час – забезпечення належного рівня безпеки держави, а одною з основних її складових є саме інформаційна безпека (ІБ). Прогнозуючи тенденції розвитку інформаційного контенту щодо оцінювання загроз, можна виявити основні ймовірні джерела виникнення конфліктних ситуацій. Саме тому актуальним є завдання створення спеціальних методик та комп'ютерних програм, які б дозволяли прогнозувати розвиток концентрації можливих загроз критичних ситуацій в інформаційній сфері для адекватного та вчасного реагування на здійснену атаку.

Огляд останніх досліджень і публікацій. Серед відомих програмних засобів, які реалізують схожі функціональні призначення, слід відзначити такі:

SPSS Statistics – комп'ютерна програма для статистичної обробки даних, один з лідерів ринку в галузі комерційних статистичних продуктів, призначених для проведення прикладних досліджень у соціальних науках [1].

Можливостями *SPSS Statistics* є: введення і зберігання даних, здатність використовувати змінні різних типів, частотність ознак, таблиці, графіки, таблиці спряженості, діаграми, первинна описова статистика, маркетингові дослідження та аналіз їх даних.

Statistica – програмний пакет для статистичного аналізу, розроблений компанією StatSoft, який реалізує функції аналізу, управління та видобутку даних, їх візуалізації із залученням статистичних методів.

Пакет має широкі графічні можливості, дозволяє виводити інформацію у вигляді різних типів графіків (включаючи наукові, ділові, тривимірні та двовимірні графіки в різних системах координат, спеціалізовані статистичні графіки тощо), усі компоненти графіків налаштовуються [2].

DataDesk – програма для аналізу візуальних даних, їх візуального дослідження та статистики. Вона забезпечує здійснення дослідницького та стандартного статистичного аналізу даних за допомогою динамічно підключених графічних дисплеїв, які оновлюють будь-яку зміну в даних одночасно [3].

Основне призначення *прикладного програмного продукту "Мезозавр"* полягає в проведенні розвідувального аналізу часових рядів. Це стосується ситуації, коли на розсуд дослідника необхідно оцінити наявну цифрову інформацію, застосовуючи різні методи обробки та аналізу даних. Пакет дозволяє здійснювати подібні дослідження з високим ступенем оперативності [3].

Зазначені вище програми лише подібні до запропонованого для реалізації програмного засобу, який інтегрує в собі пошук релевантних джерел, аналіз вибраного контенту, прогноз його розвитку, складається з математичних методик та технологічних компонентів даних у єдиний профіль для конкретної галузі застосування сфери ІБ. Головною відмінністю розробленого програмного продукту від наведених вище є локально-глобальний контроль зміни контенту і прогнозування конфліктних ситуацій залежно від зміни воєнно-політичної обстановки (ВПО).

Формулювання завдання дослідження. Метою статті є розробка методики прогнозування розвитку контенту в інтернет-просторі.

Виклад основного матеріалу дослідження. Для реалізації процесу прогнозування динаміки розвитку інформаційного контенту слід зосередитись на опрацюванні двох напрямків:

прогнозуванні зміни контенту в мережі Інтернет;

оцінюванні ВПО для відслідковування її взаємозалежності від змін контенту в Інтернеті.

Методику прогнозування зміни контенту в глобальній мережі пропонуємо реалізувати за наведеними нижче етапами:

1. За відкритими джерелами інформації (друковані видання сканованого і розпізнаного типу, масив (каталог) документів, інтернет-простір) формується сегмент початкових даних (файлів) – інформаційна стрічка, текстовий документ (певного формату), інформаційний контент сайту та ін.

2. Інформаційний сегмент формується та зберігається в базі даних (БД) у вигляді анованої інформації про характеристику джерела, час появи повідомлення та ін.

3. Надалі здійснюється селекція інформаційного масиву за визначеним контентом з метою пошуку релевантних документів (математичний апарат пошуку – метод опорних векторів; критерій відбору документів дискретно встановлюється користувачем). Результат – статистика появи контенту для кожного інформаційного джерела.

4. Формується результуюча оцінка для часу T за обраним методом обробки параметричної надмірності, наприклад, методом найменших квадратів (МНК). Результати подаються у вигляді стовпчикових діаграм.

5. З результуючої оцінки для дискретних моментів часу x_i формуються відповідні масиви (4). Вони є вихідними даними для побудови трендової моделі.

6. Отриманий після згладжування тренд накладається на стовпчикову діаграму відібраних релевантних джерел.

7. Надалі прогнозується ситуація на заданий час за різними моделями, методами.

8. Оцінювання ВПО здійснюється як паралельна процедура з процесом прогнозу зміни контенту.

З урахуванням зазначеного впливає послідовність виконання даного завдання:

пошук у вікні запиту контенту потрібної нам спрямованості;

відбір релевантних джерел;

усереднення відібраних релевантних джерел;

побудова тренду з відображеним прогнозом за межами спостереження за контентом;

оцінювання ВПО;

кінцеве накладання прогнозу на ВПО, що дозволяє спостерігати синергетичний ефект динаміки контенту.

На рис. 1 наведено структурну схему методики прогнозування динаміки розвитку інформаційного контенту в мережі Інтернет, яка розкриває вказану вище послідовність виконання завдання.

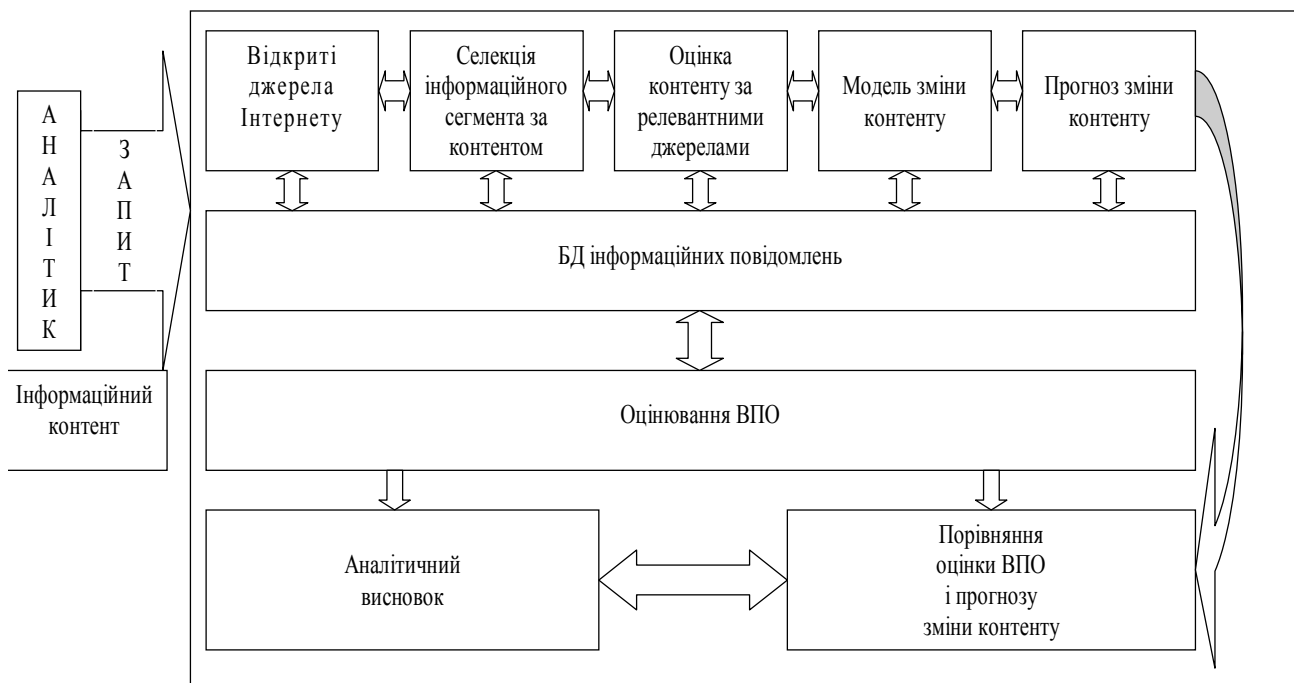


Рис. 1. Структурна схема методики прогнозування динаміки розвитку інформаційного контенту в інтернет-просторі

Практика реалізації даної методики передбачає необхідність математичного та операційного наповнення кожного блоку.

Селекцію з відкритих джерел інформаційного сегмента за контентом здійснюють для певного фіксованого часу появи інформаційного контенту за допомогою методу опорних векторів [4].

За даним методом документ відображається як вектор у багатовимірному просторі, кожний вимір якого відповідає певній характеристиці лексем зі словника аналізованих текстових масивів. Текстовий масив можна зобразити у вигляді матриці слів, у якій стовпці визначають документ, а рядки – частоти термів у ньому. Тоді кожен стовпець використовують для формування вектора відповідного документа. Мірою відстані між двома документами може бути кут між їх векторами в утвореному векторному просторі.

Для визначення належності до тематики скористаємось скалярним добутком векторів, звідки знайдемо кут між тематичним і текстовим векторами:

$$\cos(\alpha) = \frac{tf_1 \cdot tf_2 + idf_1 \cdot idf_2}{\sqrt{tf_1^2 + idf_1^2} \cdot \sqrt{tf_2^2 + idf_2^2}}, \quad (1)$$

де tf_1 та idf_1 – координати опорного вектора тематики;

tf_2 та idf_2 – координати вектора текстового документа.

Документ для розгляду відбирають за мінімальним значенням величини косинуса кута між координатами опорного вектора тематики та координатами вектора текстового документа (1).

Для оцінювання контенту за релевантними джерелами потрібно обробити надмірнісний інформаційний масив, оскільки працюємо саме з великим обсягом даних. Цю задачу запропоновано розв'язати за допомогою методу середньозважених, суть якого викладено нижче. Множина повторень у вибраній тематиці G характеризується різними джерелами по-різному, суть даного методу полягає в усередненні повторень декількох джерел в один показник I, II, III і т. д. (рис. 2).

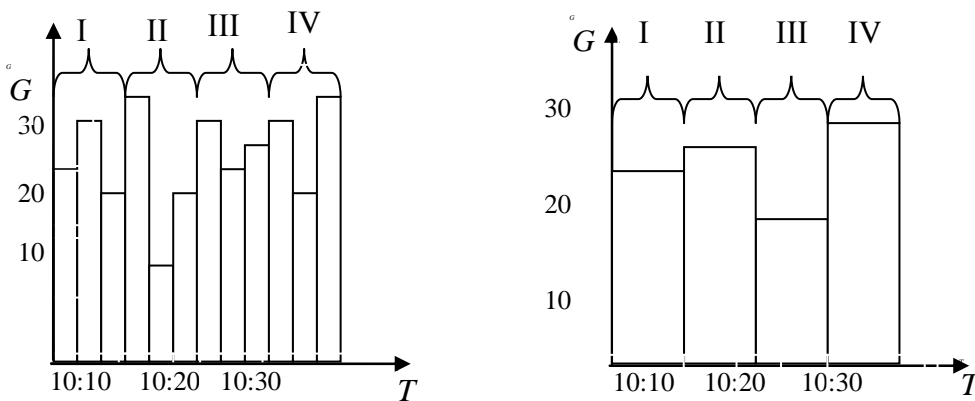


Рис. 2. Графіки опису роботи методу середньозважених

Нехай зроблено виміри деякої величини з неоднаковою точністю, необхідно визначити її середнє значення за результатами декількох вимірів. У даному випадку потрібно враховувати точності окремих вимірів за допомогою відносних ваг вимірів. Під відносною вагою виміру розуміється величина η , яка обернено пропорційна дисперсії кожного виміру σ :

$$\eta = \frac{1}{\sigma_i^2}. \quad (2)$$

Середнє значення ζ при нерівноточних показниках довіри до джерела називається середньозваженим і визначається в такий спосіб:

$$\zeta = \frac{\sum_{i=0}^n x_i \eta_i}{\sum_{i=0}^n \eta_i}, \quad (3)$$

де x_i – вимір i -го результату.

Згладжене значення швидкості після чергового аналізу визначається за двома нерівноточними його значеннями. Одним з них є згладжене значення за результатами попередніх досліджень, а інше – обрахованим значенням у новому аналізі.

За умов постійності усереднюючого параметра і за великої кількості спостережень є теоретична можливість отримати значення даного параметра з будь-якою мірою точності. Висока ж точність оцінювання параметрів забезпечує обрахунки згладжених і екстраполярих значень також з високою точністю.

Іншою важливою перевагою даного методу є малий об'єм збережуваної інформації за кожною ціллю, оскільки для нових розрахунків необхідно зберігати тільки усереднене значення параметрів з відповідними дисперсіями.

Усереднювання за методом середньозважених виправдовує себе тільки тоді, коли параметри є постійними. Проте умова постійності параметрів, особливо на великих інтервалах спостереження, виконується не завжди, оскільки ціль може маневрувати. У разі маневру цілі усереднювання параметрів призводить не до поліпшення, а до погіршення результатів. Таким чином, як і раніше, потрібно розрізняти два випадки: або ціль рухається рівномірно і прямолінійно, або маневрує за курсом чи швидкістю. За наявності маневру, що пов'язано зі зміною того або іншого параметра, усереднювання за ним повинне бути зупинено. Якщо при цьому інший параметр залишається незмінним, то доцільно продовжувати його усереднювання.

Здійснивши селекцію інформаційного масиву і розподіливши релевантні документи за датою появи і довірою до джерела, потрібно обробити його для прогнозування за МНК, суть роботи якого викладено нижче.

Застосування МНК вдало підходить для реалізації *моделі зміни контенту та моделі прогнозу зміни контенту*. Розглянемо основну суть даного методу.

Початковою інформацією для МНК є вибірка експериментальних даних, подамо їх у вигляді табл. 1.

Таблиця 1

t	t_1	t_2	t_3	...	t_n
G	G_1	G_2	G_3	...	G_n

У табл. 1 $t_i, i=1...n$, – дискретні моменти часу (значення аргументу) вимірювання значень функції $G_i, i=1...n$. Задача обробки вибірки вимірів (табл. 1) з використанням МНК полягає у визначенні параметрів апроксимуючої функції $G = f(t)$, яка найбільш точно відтворювала б перебіг реального процесу.

У (4) подано вигляд матричного алгоритму МНК:

$$\hat{C} = (F^T R^{-1} F)^{-1} F^T Y^- ; \tag{4}$$

$$\bar{T} = F \hat{C} .$$

Застосування алгоритму МНК (4) дозволяє одержати поліноміальну модель досліджуваного процесу, узгоджену з експериментальними даними (табл. 1), а також згладжені значення вимірних координат, які задовольняють критерій оптимальності.

Матриці, що входять у вирази (4), мають такі розмірності, назви та загальний вигляд:

$C[m \times 1]$ – вектор коефіцієнтів згладжувального полінома;

$F[n \times m]$ – матриця значень базисних функцій (матриця плану);

$Y[n \times 1]$ – вектор вимірних параметрів;

$T[n \times 1]$ – вектор згладжених параметрів.

Вказані матриці в загальному випадку мають такий вигляд:

$$\begin{aligned} \hat{C} &= (C_0 C_1 \dots C_n); \\ F &= \begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \dots & \varphi_m(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) & \dots & \varphi_m(x_2) \\ \dots & \dots & \dots & \dots \\ \varphi_0(x_n) & \varphi_1(x_n) & \dots & \varphi_m(x_n) \end{pmatrix}; \\ \bar{Y} &= (y_1 y_2 \dots y_n)^T; \\ \bar{T} &= (t_1 t_2 \dots t_n)^T. \end{aligned} \quad (5)$$

Аналіз алгоритму МНК показує, що вектор вимірних координат G є вхідною інформацією, вектори C і T підлягають визначенню. Далі необхідно сформувати матрицю базисних функцій F .

Розглянемо докладно процес формування матриці базисних функцій.

Вихідною інформацією для формування матриці F є, перш за все, порядок експериментальної кривої k , що підлягає згладжуванню. Оперуючи вихідною інформацією, для формування матриці F необхідно розв'язати такі задачі:

- 1) вибрати порядок згладжування полінома;
- 2) поставити у відповідність вимірним дискретним значенням досліджуваного процесу значення аргументу апроксимуючої функції, параметри якої підлягають визначенню (сформувати сітку вимірів);
- 3) визначити базисні функції згладжувального полінома.

Якість реалізації кожного із зазначених вище етапів формування базисних функцій визначає ефективність згладжування вимірювальної інформації з використанням МНК. Розглянемо зміст цих етапів.

Вибір порядку згладжувального полінома

Для розв'язання даної задачі сформуємо базисні функції, спираючись на степеневий ряд, наприклад, до третього порядку:

$$1, t, t^2, t^3 \dots; \quad (6)$$

тоді згладжувальний поліном матиме таку структуру:

$$T(x) = C_0 + C_1 x + C_2 x^2 + C_3 x^3 \dots \quad (7)$$

Саме даний метод було обрано для проведення згладження трендової моделі та здійснення прогнозу на часовому інтервалі T у вигляді, наведеному на рис. 3.

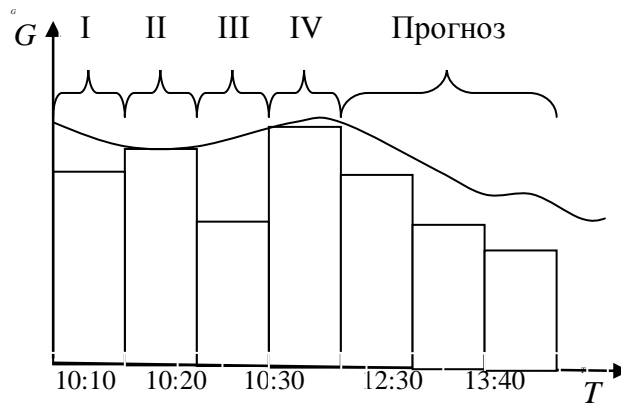


Рис. 3. Згладжений тренд за МНК

Прогноз спостерігається поза межами оцінювання контенту за МНК, а реалізується відповідно до наступних значень апроксимуючої поверхні, які досить просто встановити відповідно до вибраного порядку згладжувального полінома.

З викладеного вище впливають основні прогностичні властивості:

надання можливості налагодження роботи в разі кризових ситуацій для виконання поставленого завдання;

приймання ризикових і нестандартних рішень у випадку відхилення розвитку ситуації від допустимого ходу подій;

можливість контролювати і координувати перебіг подій.

Оцінювання ВПО

Основна ідея полягає в поясненні динаміки зміни інформаційного контенту в мережі Інтернет на фоні ВПО та інтеграції цих результатів з метою підвищення ефективності інформаційних зведень за цільовим призначенням. Надзвичайною властивістю людського інтелекту є здатність приймати правильні рішення в умовах неповної та нечіткої інформації. Побудова моделей, наближених до розуму людини, і використання їх у комп'ютерних системах є на сьогодні однією з найважливіших проблем науки.

У світі розвиток нечіткої логіки йде шляхом створення систем, потрібних великому бізнесу і військовим. Нечітка логіка застосовується під час аналізу нових ринків, біржової гри, оцінювання політичних рейтингів, вибору оптимальної цінової стратегії тощо [6]. У даній ситуації ці знання були застосовані для оцінювання ВПО.

Згідно з [7] ВПО – це конкретний стан воєнно-політичних відносин між воєнно-політичними силами (ВПС) (державами, коаліціями, націями, партіями, різного роду союзами та ін.), пов'язаний зі створенням і використанням військової потужності держав у політичних цілях. У сьогоднішній ситуації ВПО притаманний більш глибокий зміст, розширився спектр вирішуваних при цьому завдань, збільшилась кількість складових, що його визначають.

З визначення ВПО випливає, що ВПС є головними учасниками воєнно-політичних відносин, носіями воєнно-політичних ідей та інтересів. Тому для оцінювання та аналізу тенденцій розвитку ВПО є актуальним визначення класів, подібних ВПС, що дасть змогу виявляти найбільш небезпечні серед них, які готові реалізовувати свої цілі за допомогою військової сили, а також визначати ситуації у відносинах між ВПС, що становлять загрозу національним інтересам України.

Виходячи з цього, виникає потреба в створенні спеціальних систем, які б дозволяли правильно визначати найбільш небезпечні ВПС в умовах неповної та нечіткої інформації.

Прогнозуючи тенденції розвитку ВПО в державі щодо оцінювання воєнної загрози, можна виявити основні ймовірні джерела виникнення конфліктних ситуацій, які здатні призвести до збройних конфліктів.

З використанням даних джерел необхідно оцінити ВПО, враховуючи шість основних її станів: спокійна (звичайна), виникнення напруженості, загострена, кризова, збройний конфлікт, війна.

Кожний стан ВПО будемо оцінювати за ймовірними джерелами виникнення конфліктних ситуацій.

Таким чином, визначимо, що $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ – множина станів ВПО; $P = \{P_1, P_2, \dots, P_n\}$, де $n = \overline{1, m}$, – множина показників (імовірних джерел виникнення конфліктних ситуацій).

Необхідно за допомогою проведення процедури експертного оцінювання визначити стан ВПО, виходячи із визначених вище показників.

Реалізація задачі оцінювання ВПО

Оцінювати зазначені показники будемо за допомогою шкали порівняльних оцінок Сааті (табл. 2) [9].

Розподілимо альтернативи таким чином: a_1 – спокійна (звичайна) ситуація; a_2 – виникнення напруженості; a_3 – загострена; a_4 – кризова; a_5 – збройний конфлікт; a_6 – війна.

Таблиця 2

Шкала оцінок важливості

Відносна важливість критеріїв P_i і P_j	Оцінка (ступінь важливості)	Пояснення
Рівна важливість	1	Рівний внесок двох видів діяльності задля мети
Помірна перевага одного над іншим	3	Досвід і судження надають незначну перевагу одного виду діяльності над іншим
Суттєва або сильна перевага	5	Досвід і судження надають більшу перевагу одному виду діяльності над іншим
Значна перевага	7	Одному виду діяльності надається настільки сильна перевага, що вона стає практично значущою
Дуже сильна перевага	9	Очевидна перевага одного виду діяльності над іншим підтверджується найбільш сильно
Проміжні рішення між двома суміжними судженнями	2, 4, 6, 8	Застосовуються в компромісному випадку

Оцінювання проведемо за допомогою методу А. П. Ротштейна [8], суть якого викладена нижче.

За табл. 2 формується матриця парних порівнянь (матриця рангів):

$$A = \begin{pmatrix} 1 & \frac{r_2}{r_1} & \frac{r_3}{r_1} & \dots & \frac{r_n}{r_1} \\ \frac{r_1}{r_2} & 1 & \frac{r_3}{r_2} & \dots & \frac{r_n}{r_2} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{r_1}{r_n} & \frac{r_2}{r_n} & \frac{r_3}{r_n} & \dots & 1 \end{pmatrix} \quad (8)$$

Для цієї матриці характерні такі властивості:

вона діагональна, у зв'язку з чим $a_{ji} = 1, i = \overline{1, n}$;

елементи, симетричні відносно головної діагоналі, пов'язані залежністю $a_{ij} = 1 / a_{ji}$;

вона транзитивна, тобто $a_{ik} \cdot a_{kj} = a_{ij}$, оскільки $\frac{r_i}{r_k} \cdot \frac{r_k}{r_j} = \frac{r_i}{r_j}$.

Наявність цих властивостей зумовлює те, що при відомих елементах одного рядка матриці A легко визначити елементи всіх інших рядків. Якщо відомий r -й рядок, тобто елементи $a_{kj}, k, j = \overline{1, n}$, то будь-який елемент a_{ij} знаходиться таким чином:

$$a_{ij} = a_{kj} / a_{ki}, i, j, k = \overline{1, n}.$$

Ступені належності визначаються в такий спосіб [8]:

$$\begin{cases} \mu_1 = \left(1 + \frac{r_2}{r_1} + \frac{r_3}{r_1} + \dots + \frac{r_n}{r_1} \right)^{-1} \\ \mu_2 = \left(\frac{r_1}{r_2} + 1 + \frac{r_3}{r_2} + \dots + \frac{r_n}{r_2} \right)^{-1} \\ \dots \\ \mu_n = \left(\frac{r_1}{r_n} + \frac{r_2}{r_n} + \frac{r_3}{r_n} + \dots + 1 \right)^{-1} \end{cases} \quad (9)$$

Формули (9) дають можливість розраховувати ступені належності $\mu_s(u_i)$ елементів $u_i \in U$ до нечіткого терму \tilde{S} двома незалежними шляхами: за абсолютними оцінками рівнів r_i (табл. 2); за відносними оцінками рангів $\frac{r_i}{r_j} = a_{ij}, i, j = \overline{1, n}$, які утворюють матрицю (8).

При проведенні розрахунків ступенів належності за формулами (9) необхідно шкалу формувати (оцінювати альтернативи) як r_j над r_i . Наприклад: 1 – за відсутності переваги r_j над r_i . У такому випадку результати не будуть суперечити тій же процедурі за методом аналізу ієрархій [9].

Таким чином, нехай є множина з m альтернатив, які визначають множину станів ВПО $A = \{a_1, a_2, \dots, a_m\}$.

Тоді для критерію P , який визначає множину показників (імовірних джерел виникнення конфліктних ситуацій), може бути розглянута нечітка множина $P = \{\mu_c(a_1)/a_1, \mu_c(a_2)/a_2, \dots, \mu_c(a_m)/a_m\}$, де $\mu_c(a_i) \in [0,1]$ – оцінка альтернативи a_i за критерієм P , що характеризує ступінь відповідності альтернативи поняттю, обумовленому критерієм.

Якщо є n критеріїв: P_1, P_2, \dots, P_n , – то кращою вважається альтернатива, що задовольняє і критерій P_1 , і P_2, \dots, P_n . Тоді правило для вибору найкращої альтернативи може бути записане у вигляді перетину відповідних нечітких множин: $D = P_1 \cap P_2 \cap \dots \cap P_n$.

Операції перетину нечітких множин відповідає операція \min , виконувана над їх функціями належності [8]:

$$\mu_D(a_j) = \min_{i=1,n} \mu_{P_i}(a_j), \quad j = 1, m. \quad (10)$$

За кращу обирається альтернатива a^* , що має найбільше значення функції належності:

$$\mu_D(a^*) = \max_{i=1,m} \mu_D(a_j). \quad (11)$$

Будемо вважати за обмеження, що всі критерії P однакової важливості.

Узгодженість думок експертів оцінюємо, розрахувавши величину коефіцієнта конкордації:

$$W = \frac{12S}{n^2(m^3 - m)}, \quad (12)$$

де S – сума квадратів відхилень усіх оцінок рангів кожного об'єкта експертизи від середнього значення;

n – кількість експертів;

m – кількість об'єктів експертизи (альтернатив).

Коефіцієнт конкордації змінюється в діапазоні $0 \leq W \leq 1$, де 0 – повна неузгодженість, 1 – повна узгодженість.

Від якості конкретних експертних оцінок залежить надійність рішень, що приймаються на їх основі, а вирішення завдання підвищення точності й надійності оцінок – від якісного, добре підбраного складу експертів, ретельності розробки процедури експертизи, рівня проведення оцінювання й опрацювання його результатів, від зменшення кількості факторів, що піддаються вимірюванню.

Узагальнені результати проведеної експертизи зводять до відповідної таблиці. Але остаточне рішення експерти можуть не прийняти, хоча коефіцієнт конкордації дозволяє ці рішення врахувати.

У даному випадку найпростішим було б визнати найкращою альтернативою ту, яку прийняли більшість експертів (поставили на перше місце). Наприклад, це буде альтернатива a_2 – виникнення напруженості (загострена).

Оскільки ми ввели обмеження про те, що компетентність експертів однакова, то можна застосувати принцип Кондорсе для обрання найкращої альтернативи, який полягає в тому, що кожен експерт враховує альтернативи [9]. На основі цього для кожної пари альтернатив (A_i, A_j) розраховується m_{ij} – кількість експертів, які визнали альтернативу A_i кращою, ніж альтернатива A_j .

Завданням дослідження було підтвердити актуальність розробки алгоритму прогнозування динаміки розвитку інформаційного контенту в мережі Інтернет. Для цього було вибрано інформацію, яка накопичилась за рахунок чергувань щодо моніторингу Інтернету та телевізійного простору, оскільки досліджувані характеристики задовольняють тематику, яку на даний час найактуальніше можна дослідити і досить перспективно визначити прогноз розвитку ситуації (події на сході України, у Криму). Паралельно проводилося оцінювання обстановки для визначення впливу ВПО, яка склалась у державі, на прогноз розвитку подій. Для початку роботи вибрано контент для дослідження (рис. 4), потім вилучено з нього надмірності методом середньозважених.

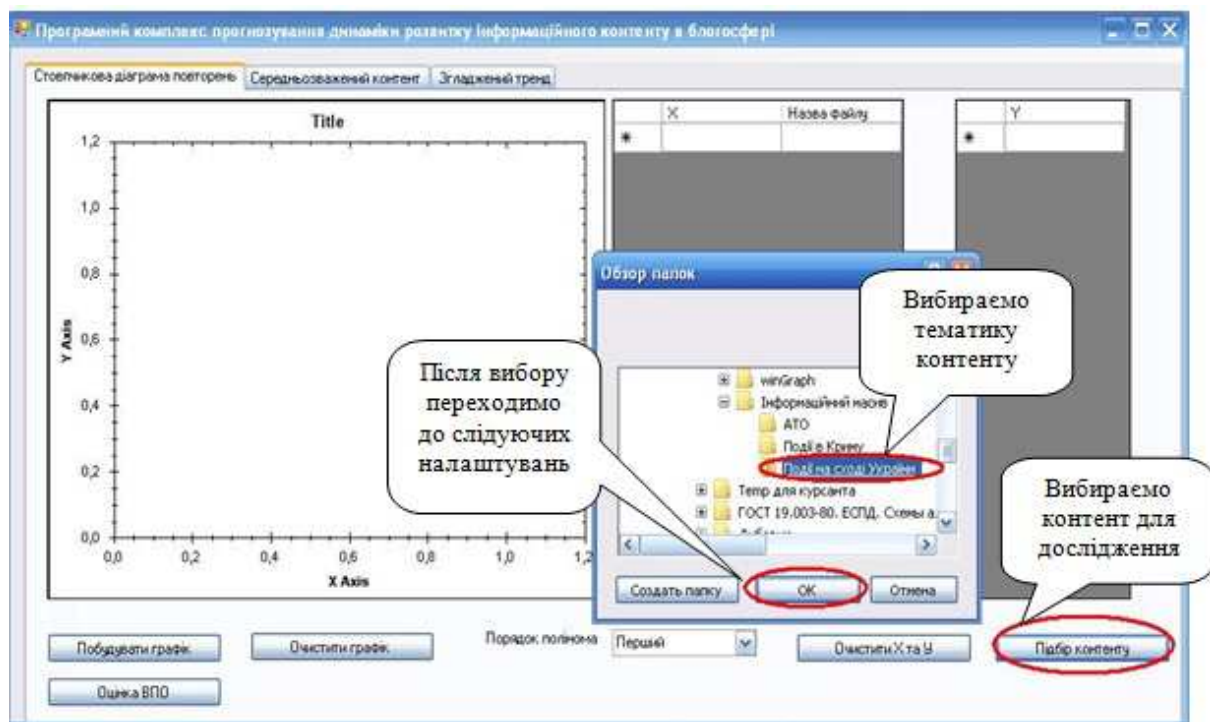


Рис. 4. Вікно програми для початку роботи

На кінцевому етапі побудовано згладжений тренд з прогнозом розвитку події відібраного контенту (рис. 5).

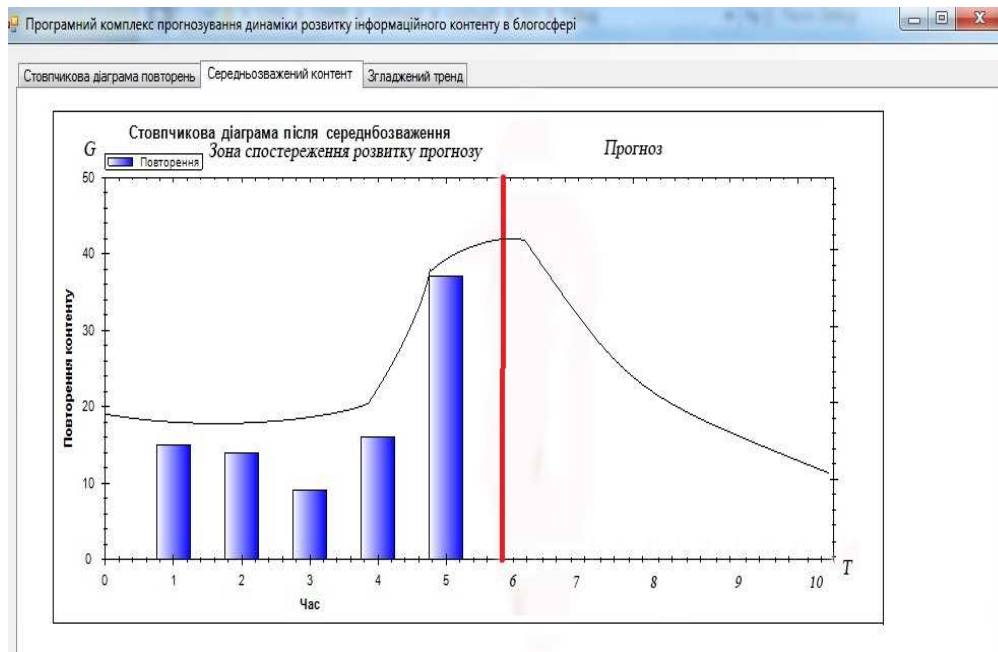


Рис. 5. Згладжений тренд з прогнозом

В останню чергу проводиться опитування експертів щодо ВПО в державі, з якого випливає кінцеве рішення (рис. 6), що порівнюється аналітиками з виглядом апроксимуючої поверхні прогнозованих подій у державі, це значно спрощує і мобілізує прийняття рішення вищим військовим керівництвом. Розроблений програмний додаток можна використовувати в підрозділах, які забезпечують інформаційно-психологічну безпеку в державі.

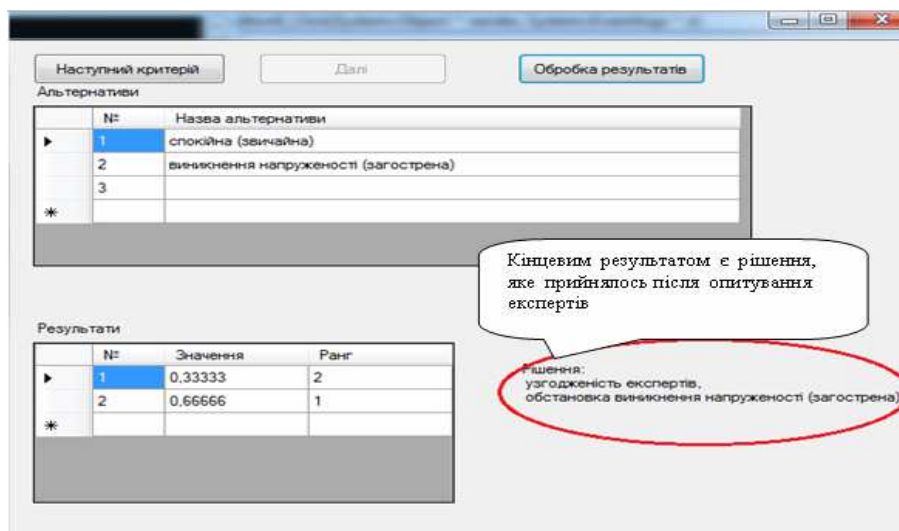


Рис. 6. Кінцеве рішення експертів по оцінюванні ВПО

Висновки. У статті розроблено методика прогнозування розвитку контенту в інтернет-просторі. Було відібрано математичний апарат для розроблення програми. При цьому використано метод опорних векторів для здійснення селекції інформаційного масиву, МНК – для згладження тренду та здійснення прогнозу розвитку подій спостереженням екстраполяції.

Відкритий програмний код дозволить швидко налагоджувати та нарощувати потужності програмного комплексу. Ефективність інформаційно-аналітичних систем

такого класу була підвищена за рахунок можливості оцінювання ВПО в державі, що показує її вплив на думку громадськості та подальший розвиток контенту вибраної тематики.

СПИСОК ЛІТЕРАТУРИ

1. SPSS [Електронний ресурс]. – Режим доступу : <http://cwppb.org.ua/articles/0032.pdf>.
2. Нейронные сети. STATISTICA Neural Networks: Методология и технологии современного анализа данных / Под редакцией В. П. Боровикова. – 2-е изд., перераб. и доп. – М. : Горячая линия–Телеком, 2008. – 288 с.
3. Горбуновін Є. А. DataDesk в інформаційних технологіях [Електронний ресурс] / Є. А. Горбуновін, В. В. Бакланчик. – Режим доступу : <http://lib4all.ru/base/B2222/B2222Part27-110.php>.
4. Лифшиц Ю. Метод опорных векторов / Ю. Лифшиц // Сучасні підходи до спрощення обрахунків. – 2011. – № 1-2 (10-11). – С. 138–141.
5. Ковбасюк С. В. Метод найменших квадратів та його практичне застосування : монографія / С. В. Ковбасюк, О. О. Писарчук, М. Ю. Ракушев. – Житомир : ЖВІ НАУ, 2008. – 228 с.
6. Нечеткие множества и теория возможностей. Последние достижения / Под ред. Р. Ягера. – М. : Наука, 1986. – 312 с.
7. Горбунов Є. А. Описово-синергетичний метод оцінки та прогнозування агресивності суб'єктів міжнародної політики [Електронний ресурс] / Є. А. Горбунов, В. В. Баклан. – Режим доступу : <http://cwppb.org.ua/articles/0059.pdf>.
8. Ротштейн А. П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети / А. П. Ротштейн. – Винница : УНИВЕРСУМ, 1999. – 320 с.
9. Хомяков В. І. Менеджмент підприємства. / В. І. Хомяков. – К. : Кондор, 2005. – 434 с.

Подано 19.02.2015

А. А. Писарчук, Д. В. Порада

МЕТОДИКА ПРОГНОЗИРОВАНИЯ РАЗВИТИЯ ИНФОРМАЦИОННОГО КОНТЕНТА В СЕТИ ИНТЕРНЕТ

В статье предложена методика прогнозирования развития информационного контента в сети Интернет на основе применения векторного поля отношений для поиска релевантных источников и метода наименьших квадратов. Разработанная программа базируется на создании для оператора среды прогнозирования развития контента в результате обработки выбранного запроса в режиме реального времени. Показано, каким образом методика и программа обеспечивают выполнение задания предупреждения развития тех или иных информационных событий.

О. О. Pisarchuk, D. V. Porada

THE METHODIC OF PREDICTION OF INFORMATIONAL CONTENT DEVELOPMENT IN THE INTERNET

The article presents the methodic of prediction of informational content development in the Internet having basis on usage of vector field of relevations for the most relevant source search and lowest squares method. The system is based on creating a comfortable for operator system of prediction of content development as a result of query processing in real-time. It is shown how the system helps to prevent the development of certain informational events.