

# *Форми, методи і засоби виявлення, оцінювання і прогнозування загроз інформаційній безпеці України*

---

УДК 351.86

*МІХЄЄВ Юрій Іванович  
ЛОБОДА Вероніка Вікторівна*

## **МОДЕЛЬ АВТОМАТИЗОВАНОЇ СИСТЕМИ ПОШУКУ ІНФОРМАЦІЇ ЗА ВИЗНАЧЕНОЮ ТЕМАТИКОЮ В МЕРЕЖІ ІНТЕРНЕТ**

**Постановка проблеми.** Сьогодні постійне збільшення інтернет-ресурсів призводить до ускладнення швидкого та релевантного пошуку потрібної інформації. Це пов'язано із збільшенням обсягу неструктурованих, різних за типом, формою та тематикою даних. У таких умовах необхідно передбачати використання різних методів збирання інформації окремо або в комплексі разом із спеціальними системами пошуку. Особливо це стосується інформаційно-аналітичної діяльності в інтересах спеціальних підрозділів. Для якісного забезпечення інформаційних потреб таких підрозділів інформаційно-аналітична діяльність має передбачати виконання комплексу інформаційних процесів, таких як пошук, оброблення та зберігання інформації.

На сьогодні необхідним є вирішення цілої низки питань, які стосуються функціонування інформаційно-пошукових систем. Головні з них: розробка алгоритмів і способів раціонального пошуку й подальшої обробки інформації за визначеною тематикою та як наслідок створення на їхній основі автоматизованої системи пошуку інформації (АСП) за визначеною тематикою в мережі Інтернет.

**Аналіз останніх досліджень і публікацій.** На сьогодні для пошуку інформації в мережі Інтернет використовують індивідуальні пошукові агенти, метапошукові системи, каталоги та системи інтернет-моніторингу [1; 2]. Аналіз результатів вирішення конкретного завдання щодо тематичного пошуку інформації показав, що використання зазначених засобів дещо ускладнене у зв'язку з їхньою вузькою спрямованістю. У наукових працях розглядаються різні підходи до обробки текстової інформації, алгоритми формування пошукового розпорядження й ін. [2–4]. Однак системний пошук інформації за визначеною тематикою не розглядається. Це призводить до неможливості ефективного використання запропонованих підходів в інформаційно-аналітичній діяльності в умовах забезпечення максимальної повноти інформаційного масиву, автоматизації процесу збирання даних і засобів навігації в ній із подальшим вилученням необхідних знань.

**Мета** статті – на основі проведеного аналізу різних підходів до пошуку інформації за визначеною тематикою та її обробки розроблення моделі авто-

## *Forms, methods and means of detecting, assessing and forecasting information security threats to Ukraine*

матизованої системи пошуку такої інформації. Використання розробленої системи для вирішення завдань інформаційно-аналітичної діяльності дасть змогу скоротити час, необхідний для підготовки інформаційних звітів, що забезпечить ефективніше виконання завдань із пошуку й аналізу інформації спеціальними підрозділами. Першочерговим завданням є розроблення моделі АСПІ в мережі Інтернет.

**Виклад основного матеріалу.** Будь-який пошук інформації починається з визначення інформаційної потреби, після чого оператор повинен сформувати пошуковий запит. Для здійснення пошукового запиту, по-перше, необхідно скласти його пошуковий образ, тобто його формальне подання в термінах інформаційно-пошукової мови. Після цього складається пошукове розпорядження, яке містить у собі пошуковий образ запиту та вказівки щодо логічних операцій, які виконуються при інформаційному пошуку. Далі пошукове розпорядження порівнюється із пошуковими образами документів та отримуються відомості про: адресу зберігання, короткий зміст документа. Від того, наскільки повно буде з'ясовано завдання оператором, у подальшому залежатиме ступінь релевантності (відповідності) знайденого матеріалу змісту завдання (див. рис. 1) [5].

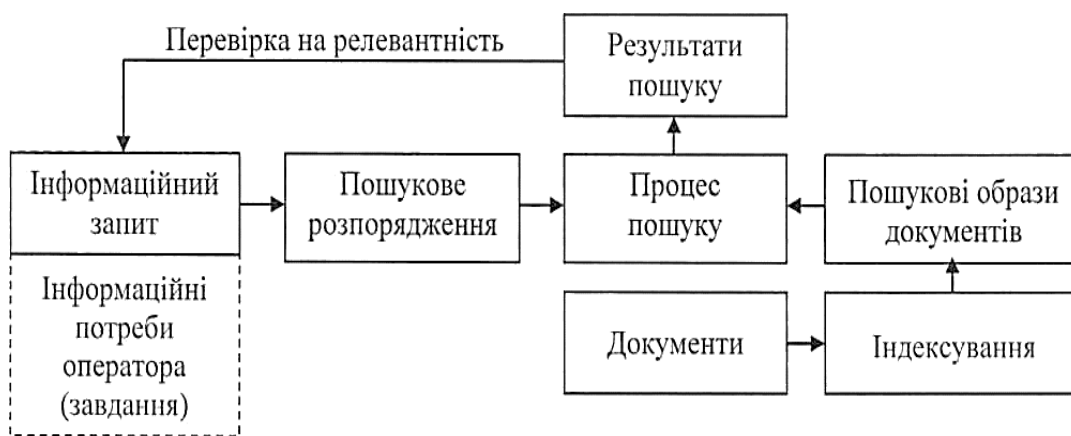


Рисунок 1 – Схема функціонування інформаційно-пошукової системи

Такий підхід до пошуку інформації відповідає векторній (лінійній) моделі пошуку, яка сьогодні застосовується в більшості відомих пошукових машин завдяки своїй простоті [4]. Для зображення цієї моделі у формалізованому вигляді використовуються такі поняття, як: словник – упорядкована множина термінів потужності  $D$ ; документ (пошуковий образ документа) – двійковий вектор розмірності  $D$ ; база – матриця  $L = Nx D$ , рядки якої відповідають  $N$  документам. Якщо термін входить у документ, то у відповідному розряді вектора стоїть 1, інакше – 0. У такому випадку процедура обробки запиту набуває такого вигляду:  $Lq = r$ , де  $q$  – вектор запиту, а  $r$  – відгук системи на

## **Форми, методи і засоби виявлення, оцінювання і прогнозування загроз інформаційній безпеці України**

запит. Модель ускладнюється шляхом використання вагових коефіцієнтів у термінах документа та запиту, які відображають їхню значущість. Використання вагових коефіцієнтів дає можливість визначити міру близькості «документ – запит»  $s$  як косинус кута між векторами запиту та документа:

$$s(q, d) = \frac{\sum_i q_i d_i}{\sqrt{\sum_i q_i^2 \sum_i d_i^2}}, \quad (1)$$

де  $0 \leq s(q, d) \leq 1$ ;

$q = (q_1, \dots, q_n)$  – запит,  $q_i$  – ваговий коефіцієнт  $i$ -го терміна запиту;

$d = (d_1, \dots, d_n)$  – документ,  $d_i$  – ваговий коефіцієнт  $i$ -го терміна документа.

Більш гнучку реакцію на запит оператора можливо забезпечити шляхом інтерактивного пошуку. Суть цієї моделі полягає у зворотному зв'язку за релевантністю. Оператору надаються початкові результати обробки запиту, які він оцінює щодо релевантності. Після цього запит корегується та пошук продовжується. Крім того, АСПІ повинна підтримувати архів запитів, включати тезаурус, засоби перевірки орфографії та синтаксичного розбору мовного запиту. Для підвищення повноти пошуку можна використати тезаурус. У цьому разі до слів запиту додаються близькі їм слова з тезаурусу [3]. Це дає змогу знаходити релевантні текстові фрагменти, які взагалі не містять слів запиту. Разом із тим, ступінь відповідності знайденого матеріалу інформаційному запиту зменшується за рахунок наявності в інтернеті різних видів замаскованої інформації, зміст якої не відповідає головному заголовку. Метою такої діяльності є підвищення рейтингу власників сайтів за рахунок збільшення звернень уваги користувачів на сенсаційну інформацію. У результаті такого пошуку оператор отримує великі обсяги зайвої інформації, яка не відповідає заданій тематиці. Можливим шляхом вирішення цієї проблеми є створення бази даних синонімів, що дасть можливість частково автоматизувати процес формування інформаційного запиту. Крім того, АСПІ повинна самонавчатися. Зробити це можливо завдяки аналізу результату пошуку (перевірки відповідності знайденого матеріалу фразі, яка була задана на пошук). Оператору має бути надана можливість обрати варіант пошукової фрази, яка формується з отриманого завдання-речення.

Проведений аналіз існуючих АСПІ показав, що для ефективного пошуку тематичної інформації в мережі Інтернет запит повинен формуватися з урахуванням [6]:

- можливості використання логічних зв'язок;
- обмеження у відстані між словами запиту в документі;
- пошуку за визначеними полями документа;
- можливості призначення вагових коефіцієнтів словам у запиті;
- проведення сортування за датою, розміром документа.

## *Forms, methods and means of detecting, assessing and forecasting information security threats to Ukraine*

Ранжування знайдених документів повинне відбуватися на основі міри близькості «запит – документ». Один з ефективних методів ранжування полягає у перевірці частоти появи слова запиту у відповідних полях HTML-документа: <Title>, <H1...H6>, <ADDRESS>, <B>, <STRONG>. Чим ближче розташовано слово до початку документа, тим більший його ваговий коефіцієнт. Для встановлення вагових коефіцієнтів окремих слів оцінюється наявність пошукових слів у заголовку та підзаголовку документа.

Майбутня АСПІ повинна враховувати різні служби інтернету, які привертають увагу аналітиків спеціальних служб, такі як пошукові системи, тематичні каталоги ресурсів, сайти новин, RSS-повідомлення та інформаційні агентства, які транслюють новини онлайн. Для досягнення цієї мети першим кроком у створенні АСПІ є створення бази даних пошукових ресурсів, які доступні в інтернеті, з урахуванням їхніх особливостей щодо надання інформації за визначеною тематикою. Для цього пропонується використання конкретних рубрик у базі даних (див. табл. 1) [7].

Таблиця 1 – Рубрики бази даних пошукових ресурсів

№ з/п	Назва пошукового ресурсу	Адреса ресурсу	Рейтинг за результатами власника	Кількість запитів	Кількість знайдених посилань за запитом	Ступінь відповідності запиту (ступінь релевантності)	Частота оновлення інформації за визначеною тематикою
1	2	3	4	5	6	7	8

Уточнення пошуку можливе при використанні тематичної класифікації ресурсів – векторів простору словника (термінів індексації) системи [8]. Завдання полягатимуть у тому, щоб найкращим чином обрати такі ознаки та визначити правила, на основі яких буде прийматися рішення стосовно ресурсу до певної рубрики. Тематичні рубрики повинні бути контекстозалежними. Використання наданої бази даних перед безпосереднім пошуком дасть змогу оператору сформулювати метапошуковий запит з урахуванням апріорних даних, що ґрунтуються на оцінюванні результатів пошуку з конкретного ресурсу.

Окремо розглядається завдання щодо вилучення інформації з онлайн відеотрансляції новин. У цьому разі для своєчасного відпрацювання завдання оператору необхідно постійно відслідковувати визначені завчасно передачі. Одним із можливих варіантів вирішення цього завдання в умовах обмеженої кількості особового складу є здійснення запису передач шляхом формування плану їх перегляду. При цьому можна скористатися існуючими інтернет-

## ***Форми, методи і засоби виявлення, оцінювання і прогнозування загроз інформаційній безпеці України***

порталами, що надають інформацію про розклад телепередач. Це дасть можливість автоматично обирати типи передач, які відповідають тематичному пошуку.

Подальша робота АСПІ пов'язана з обробкою знайденого матеріалу. Для вирішення цього завдання необхідним етапом при створенні майбутньої системи є організація автоматичного реферування знайденої інформації. Порядок обробки документа при цьому буде складатися з етапів, наведених на рис. 2:

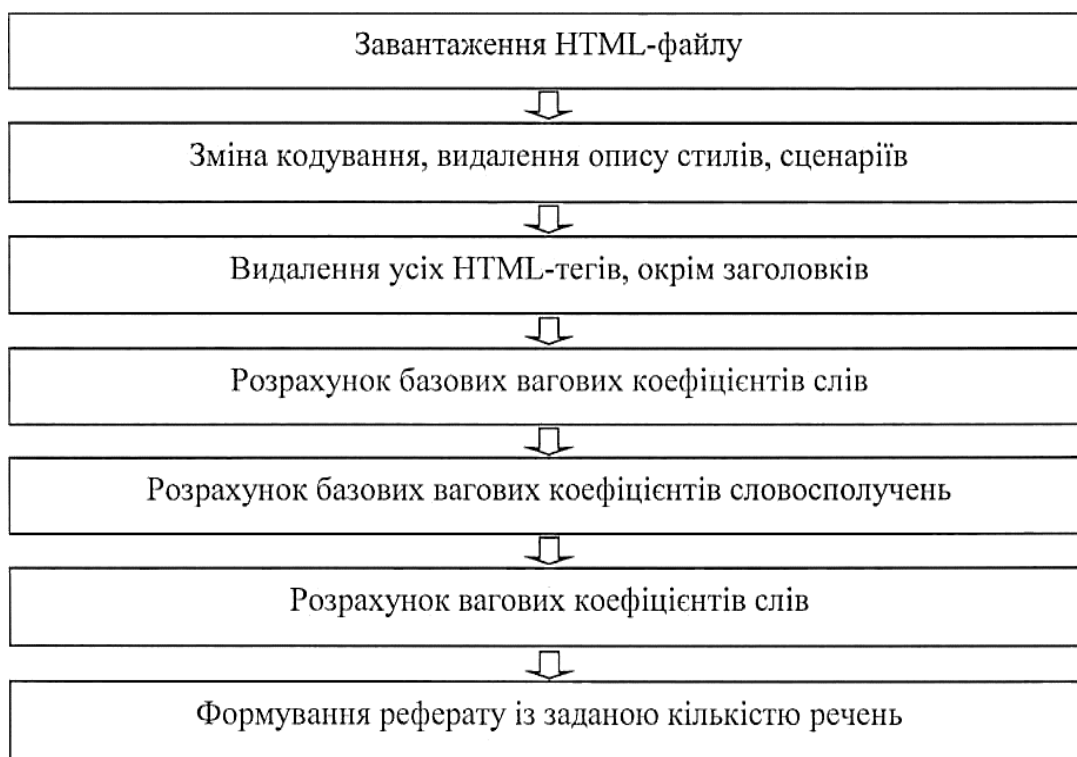


Рисунок 2 – Етапи автоматичного реферування документів

Для візуалізації знайденої інформації з метою її подальшого аналізу доцільно використати технологію побудови семантичних мереж [8]. Порівняння семантичних мереж різних текстів дає можливість установити ступінь їхньої змістової близькості, що може використовуватися для автоматичної класифікації документів за заданими рубриками, їх пошуку за подібністю заданого тексту, а також розбиття інформаційного масиву на класи документів близького змісту.

Отже, модель АСПІ за визначеною тематикою в мережі Інтернет буде такою, як показано на рис. 3.

## *Forms, methods and means of detecting, assessing and forecasting information security threats to Ukraine*

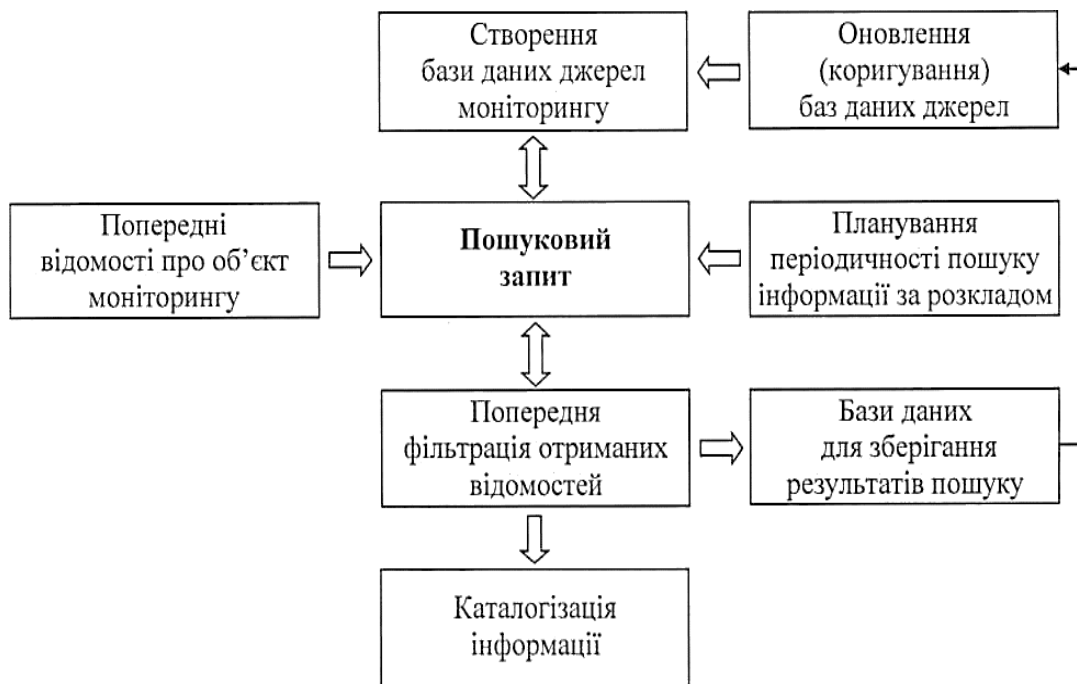


Рисунок 3 – Модель автоматизованої системи пошуку тематичної інформації в мережі Інтернет

**Висновки.** Використання запропонованої моделі автоматизованої системи пошуку тематичної інформації в мережі Інтернет під час інформаційно-аналітичної діяльності у спеціальних підрозділах дасть змогу провести повний і точний пошук інформації за визначеною тематикою шляхом:

- забезпечення оператора-аналітика засобами швидкого й ефективного пошуку різномірної інформації за об'єктами моніторингу;
- надання засобів швидкого виявлення неявних зв'язків між об'єктами моніторингу й пов'язаними з ними фактами та подіями;
- проведення фіксації та візуалізації результатів аналітичних досліджень шляхом генерації дайджестів статей, фактів, формалізованих досьє, семантичних мереж та інших аналітичних звітів.

### Список використаних джерел

1. Nych L. Y., Kaminskyj R. M., & Shakhovska N. B. Effectiveness Evaluation of Search in Information Systems with Consolidated Information. *Radio Electronics, Computer Science, Control*. 2016. URL: <https://doi.org/10.15588/1607-3274-2016-2-13> (дата звернення: 10.12.2022).
2. Поташова А. В. Проблеми пошуку інформації в глобальній мережі Інтернет. *Науковий огляд*. 2018. № 5 (48). С. 130–139.
3. Рогушина Ю. В. Методика розробки термінології інформаційних ресурсів як базису формування онтологій та тезаурусів для семантичного. *Інженерія*

## Форми, методи і засоби виявлення, оцінювання і прогнозування загроз інформаційній безпеці України

програмного забезпечення. 2014. № 1. С. 41–51. URL: [http://nbuv.gov.ua/UJRN/Iprz\\_2014\\_1\\_7](http://nbuv.gov.ua/UJRN/Iprz_2014_1_7) (дата звернення: 15.12.2022).

4. Крайовський В. Я., Литвин В. В., Шаховська Н. Б. Основні підходи до розроблення програмного комплексу автоматичного реферування текстових документів. *Збірник наукових праць Інституту проблем моделювання в енергетиці імені Г. Є. Пухова НАН України*. Київ : ППМЕ ім. Г. Є. Пухова НАН України, 2009. Вип. 51. С. 178–186.

5. Гришук Р. В., Даник Ю. Г. Основи кібернетичної безпеки : монографія / за заг. ред. проф. Ю. Г. Даника. Житомир : ЖНАЕУ, 2016. С. 549.

6. Варенко В. М., Братусь І. В., Дорошенко В. С., Смольников Ю. Б., Юрченко В. О. Системний аналіз інформаційних процесів : навчальний посібник. Київ : Університет «Україна», 2013. С. 203.

7. Міхеєв Ю. І., Чернявський Г. П., Манько О. В., Токар А. М. Контент-моніторинг у системі виявлення інформаційних небезпек. *Перспективи розвитку озброєння та військової техніки сухопутних військ* : збірка тез доповідей Міжнародної науково-технічної конференції, (Львів, 14–16 трав. 2014 р.). Львів : АСВ, 2014. С. 39.

8. Сухий О. Л., Міленін В. М., Тарадайнік В. М. Алгоритми пошуку в інформаційних системах : методичні рекомендації. Київ, 2015. С. 70.

Рецензенти:

доктор технічних наук, доцент

В. Шестаков,

кандидат технічних наук, професор

В. Настрадін

**Анотація.** У статті запропоновано підхід до створення моделі автоматизованої системи пошуку тематичної інформації в інтернеті, що базується на унікальному характері оперативно-аналітичної діяльності спеціальних підрозділів. Представлена модель автоматизованої системи враховує переваги та недоліки існуючих підходів щодо розроблення подібних систем. Передбачається, що впровадження розробленого спеціального забезпечення на базі запропонованої моделі автоматизованої системи дасть змогу провести повний та релевантний пошук інформації за визначеною тематикою під час інформаційно-аналітичної діяльності у спеціальних підрозділах.

**Ключові слова:** модель, мережа Інтернет, автоматизована система пошуку інформації, інформаційний запит.

**Abstract.** In this research, an approach to creating a model automated system for searching thematic information on the Internet based on the unique nature of the operational and analytical activities carried out by special units is proposed. The presented model takes into account the advantages and disadvantages of existing approaches to the development of such systems. The implementation of the developed special software on the basis proposed model of the automated System is expected to allow for a complete and relevant search information on a specific topic during information and analytical activities in special units.

**Key words:** model, Internet, automated information retrieval System, information request.