

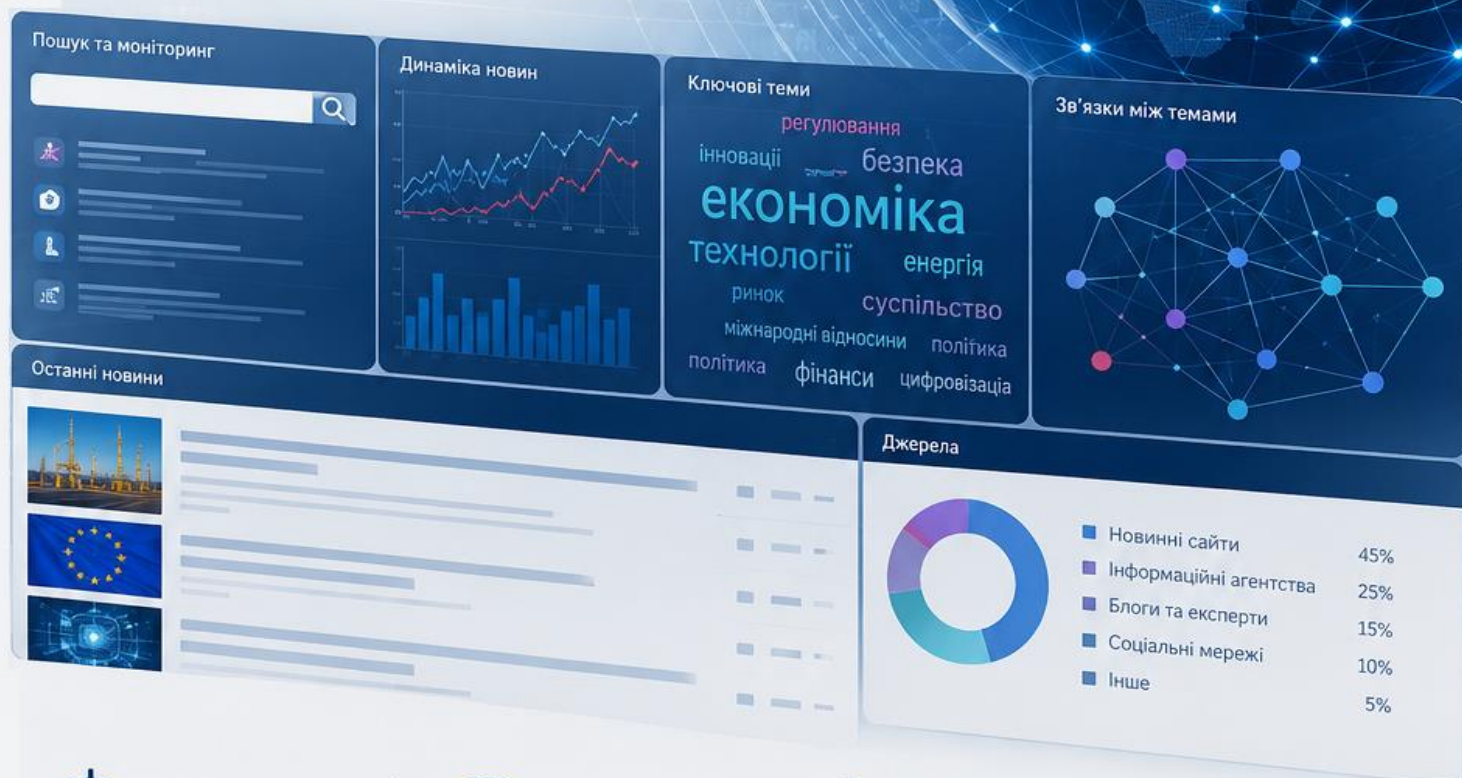


# InfoStream

Завжди у курсі

## ІНТЕЛЕКТУАЛЬНИЙ МОНІТОРИНГ НОВИН З ІНТЕРНЕТУ

НАУКОВО-МЕТОДИЧНИЙ  
ПОСІБНИК



ТОЧНІСТЬ  
ТА РЕЛЕВАНТНІСТЬ



ОПЕРАТИВНІСТЬ  
ТА ПОВНОТА



ДОСТОВІРНІСТЬ  
ТА ВЕРИФІКАЦІЯ



ПРОГНОЗУВАННЯ  
ТА АНАЛІТИКА

*InfoStream. Інтелектуальний моніторинг новин з Інтернету*: Науково-методичний посібник / за заг. ред. проф. Ланде Д.В. – Київ: Інформаційний центр «Електронні вісті», 2026 р. – 86 с. DOI: 10.13140/RG.2.2.11034.40643

У посібнику систематизовано теоретичні та практичні засади інформаційного, програмно-технологічного та аналітичного забезпечення системи моніторингу веб-ресурсів та соціальних мереж. Показано еволюцію системи від класичного повнотекстового пошуку до гібридної архітектури, де перевірені алгоритми інформаційної вибірки поєднуються з можливостями сучасних великих мовних моделей. Розглянуто класичну технологічну основу системи: формалізовану мову запитів InfoReS, інтуїтивний конструктор пошукових запитів із багаторівневою булевою логікою та контекстними операторами, механізми автоматичної детекції змістовних дублів та інформаційні портрети для динамічного уточнення релевантних вибірок. Унікальною методологічною перевагою є доступ до структурованої ретроспективної бази новин, яка безперервно формується протягом понад 30 років, що забезпечує часову глибину досліджень та відтворюваність історичних зрізів.

Значну увагу приділено аналітичним модулям нового покоління. Детально описано алгоритми автоматичного реферування та генерації структурованих резюме сюжетів, архітектуру, яка гарантує прив'язку кожного аналітичного висновку до верифікованих першоджерел, мінімізуючи ризик «галюцинацій» ШІ. Розкрито сучасні методи аналізу часових рядів: від класичних гістограм динаміки понять до виявлення інформаційних операцій. Okремо представлено підходи до прогнозування інформаційних процесів: динамічний прогноз обсягу потоку на основі статистичних моделей та семантичний прогноз розвитку тенденцій за допомогою великих мовних моделей.

Матеріал посібника орієнтований на інтеграцію отриманої інформації у контур систем підтримки прийняття рішень. Наведено методичні сценарії трансформації сирих новинних даних у структуровані звіти, дайджести та аналітичні довідки, придатні для використання в управлінській, комунікаційній та дослідницькій практиці.

Посібник адресовано інформаційним аналітикам, фахівцям у сфері стратегічних комунікацій, дослідникам медіасередовища, науковцям та студентам, які потребують системного підходу до пошуку, обробки та прогнозування інформаційних процесів у сучасному цифровому просторі.

© Інформаційний центр  
«Електронні Вісті», 2026

© Ланде Д.В., 2026



## ПУТІВНИК ПО ПОСІБНИКУ INFOSTREAM

Від новинного потоку до управлінського рішення



## ЗМІСТ

ВСТУП .....	6
РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ІНФОРМАЦІЙНОГО МОНІТОРИНГУ .....	8
1.1. Інформаційний простір Інтернет.....	10
1.2. Моніторинг як інструмент підтримки прийняття рішень .....	11
1.3. Еволюція підходів: від к пошуку до семантико-аналітичних процедур	13
1.4. Релевантність, достовірність, оперативність, повнота.....	14
РОЗДІЛ 2. ТЕХНОЛОГІЧНЕ ЯДРО ТА АРХІТЕКТУРА СИСТЕМИ INFOSTREAM .....	16
2.1. Програмно-технологічна структура.....	17
2.2. Пошукове ядро InfoReS .....	19
2.3. Механізми фільтрації .....	22
2.4. Ретроспективний фонд.....	23
РОЗДІЛ 3. КЛАСИЧНИЙ ІНСТРУМЕНТАРІЙ: МОВА ЗАПИТІВ ТА КОНСТРУКТОР ПОШУКУ .....	26
3.1. Синтаксис та пріоритети операторів InfoReS .....	27
3.2. Морфологічний пошук, точні входження, праві скорочення.....	29
3.3. Конструктор запитів .....	32
3.4. Інформаційний портрет.....	34
3.5. Класифікатор-навігатор .....	37
3.6. Типові сценарії пошуку .....	39
РОЗДІЛ 4. СУЧАСНА АНАЛІТИКА: LLM, RAG, ЧАСОВІ РЯДИ ТА ПРОГНОЗУВАННЯ .....	42
4.1. Автоматичне реферування та генерація структурованих резюме сюжетів .....	43
4.2. RAG-архітектура в InfoStream .....	46
4.3. Аналіз часових рядів.....	49
4.4. Прогнозування інформаційних процесів .....	51

РОЗДІЛ 5. ІНТЕГРАЦІЯ В СППР: ВІД СИРИХ ДАНИХ ДО УПРАВЛІНСЬКИХ РІШЕНЬ .....	55
5.1. Трансформація новинного потоку в аналітичні інсайти .....	56
5.2. Сценарії застосування.....	59
5.3. Алгоритми побудови аналітичних довідок.....	62
5.4. Побудова семантичних карт на основі великих мовних моделей.....	64
РОЗДІЛ 6. СЕРВІСНІ МОДУЛІ ТА ЕКСПЛУАТАЦІЯ СИСТЕМИ .....	68
6.1. Онлайн-доступ до оперативних та ретроспективних баз даних .....	69
6.2. Автоматизована доставка контенту .....	73
6.3. Кабінет користувача .....	75
6.4. Каталог джерел .....	76
ВИСНОВКИ .....	79
ДОДАТОК А. Довідник мови запитів InfoReS.....	81
ДОДАТОК Б. Глосарій термінів та скорочень .....	84

## ВСТУП

Метою даного науково-методичного посібника є систематизація теоретичних, технологічних та практичних засад інформаційного моніторингу новинних потоків з відкритих ресурсів Інтернету. Видання покликане ознайомити із сучасними підходами до інформаційного пошуку, навчити принципам конструювання складних запитів, ефективному використанню аналітичного інструментарію та інтеграції отриманих результатів у контур підтримки прийняття рішень.

Основними завданнями посібника є формування у користувачів навичок роботи з формалізованою мовою запитів та конструктором пошукових сценаріїв; розкриття механізмів автоматичного реферування, семантичної кластеризації та верифікації даних; освоєння методів аналізу часових рядів, хвильового моделювання та прогнозування інформаційних процесів; надання алгоритмів трансформації сирих новинних даних у структуровані аналітичні довідки. Цільовою аудиторією видання є інформаційні аналітики, фахівці з комунікацій, дослідники медіасередовища, науковці, а також студенти та аспіранти, які спеціалізуються у сфері цифрової аналітики, OSINT та систем підтримки прийняття рішень.

Структура посібника побудована за принципом поступового ускладнення: від теоретико-методологічного фундаменту до практичного застосування гібридних аналітичних процедур. На початку викладу детально розглянуто класичне технологічне ядро системи – пошуковий механізм InfoReS. Наведено формалізовану мову запитів з багаторівневою булевою логікою та контекстними операторами, механізми детекції змістовних дублів та надано інформацію про ретроспективний фонд новин, що формується протягом понад 30 років.

На цій основі викладено принципи інтеграції сучасних методів штучного інтелекту. ІІІ-модулі не замінюють класичний пошук, а доповнюють його: спочатку система відбирає релевантний масив документів за строгими формальними критеріями, після чого задіюються великі мовні моделі (LLM) для семантичного реферування, витягнення сутностей та оцінки тональності. Архітектура RAG (Retrieval-Augmented Generation) забезпечує прив'язку кожного згенерованого висновку до верифікованих першоджерел. Цим мінімізується ризик «галюцинацій» та забезпечується придатність інформації для офіційних звітів. Окремий акцент зроблено на аналізі часових рядів, від класичних гістограм динаміки понять до виявлення хвильових патернів, аномалій інформаційних потоків, а також до двокомпонентного прогнозування – динамічного (обсяг потоку) та семантичного (розвиток тенденцій за допомогою LLM). Кожен технологічний блок ілюструється скріншотами інтерфейсу та покроковими сценаріями, що забезпечує безперервний зв'язок між теорією, алгоритмами та реальним робочим середовищем.

**Активна база даних:** Система інтеграції ресурсів Інтернету

Головна | Допомога | Кабінет | Джерела | Статистика | Новини проекту

Період: 7 днів | Вилучити дублі | Морфологія | Посторинково

ПОШУК | Динаміка | Дайджест | Мова запитів Конструктор

Очистити | ШІ | Сюжети

**Рубрики**

- Автотранспорт
- АПК
- Банки
- Бізнес
- Зовнішня економіка
- Інвестиції
- Інтернет
- Калейдоскоп
- Катастрофи
- ІТ
- Протидія криміналу
- Культура
- Мас-медіа
- Медицина
- Медіатургтя
- Наука\_техніка
- Нерухомість
- Нормативні акти
- Оборона
- Освіта
- Огляди преси
- Офіційна хроніка
- Політика
- Приватизація
- Реклама\_маркетинг

**Основні сюжети**

**Кошта: ЄС зробив важливий крок вперед - всі країни ЄС погодилися відкрити переговорний кластер з Укр... (58)**  
 Фото: <https://x.com/antoniosostartm> Президент Європейської ради Антоніу Кошта назвав "важливим кроком вперед", який зробив Європейський союз, коли всі держави-члени Євросоюзу погодилися відкрити перший переговорний кластер з Україною та Молдовою. "Сьогодні Європейський Союз зробив важливий крок уперед."

**Російська мова більше не захищена в Україні: Зеленський підписав закон (41)**  
 Документ змінює підхід до мовної політики та переглядає перелік захищених мов. Анастасія Павленко Коментарі Президент України Володимир Зеленський підписав закон, який позбавляє російську мову захисту Європейської хартії регіональних або міноритарних мов в Україні. Про це повідомив очільник Верховної Ради Руслан Стефанчук. Стефанчук пояснив, що мова держави-агресорки не може користуватися захистом.

**Страны ЕС дали "зеленый свет" на переговоры о вступлении с Украиной и Молдовой (26)**  
 20:37 12.06.2026 Пт 2 мин Когда переговоры о вступлении Украины в ЕС начнутся фактически? Иван Носальский Фото: глава Євросовета Антоніу Кошта, президент України Володимир Зеленський і глава ЕК Урсула фон дер Ляйен (Getty Images) Не тратьте время на шум!

**Росія атакувала енергетику України: загинув працівник, є поранений (22)**  
 12.06.2026 Росія атакувала енергетику України: загинув працівник, є поранений Джерело: tsn.ua ДТЕК Після російського обстрілу однієї з ТЕС ДТЕК загинув працівник, ще одного поранено. Про це повідомили у пресслужбі компанії ДТЕК. Росія атакувала ТЕС ДТЕК - що відомо "Щирі співчуття рідним та близьким загиблого."

**Персональні канали**

- Мониторинг
- Архів
- Антиплагиат
- Правова наука
- Захист персональних даних
- Електронне декларування
- Кібератака
- LLM
- ШІ

Віддалити

**Персональні джерела**

- MediaPort
- Nadra.info
- АПК-Інформ
- АрміяІнформ
- ВичЕрпно
- ГолосUA
- Закарпатський Кореспондент

Уточнити запит | Вилучити

*Основний інтерфейс системи InfoStream: ліва частина містить ієрархічний каталог попередньо визначених тематичних рубрик, що забезпечують швидку навігацію інформаційними ресурсами та формування тематичних вибірок; центральна частина відображає дайджест основних новин, відібраних відповідно до поточного запиту або обраної рубрики; права частина призначена для роботи з персональними каналами користувача (збереженими пошуковими запитамі), а також персональними наборами джерел. У верхній частині інтерфейсу розташовано поле введення пошукового запиту та панель виклику основних режимів роботи системи: пошуку документів, аналізу динаміки появи повідомлень, формування дайджестів, аналізу сюжетів і режимів штучного інтелекту (ШІ)*

Посібник адаптовано до різних рівнів підготовки користувачів та контекстів застосування. У навчальному процесі він може використовуватися як базовий підручник для курсів з інформаційного пошуку, медіааналітики, цифрових досліджень та штучного інтелекту в аналітичних системах. Поетапна структура дозволяє студентам послідовно опановувати синтаксис запитів, інтерфейс класифікатора-навігатора, принципи побудови інформаційних портретів та основи роботи з LLM-звітами.

## РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ІНФОРМАЦІЙНОГО МОНІТОРИНГУ

### БЛОК-СХЕМА ВИКЛАДУ МАТЕРІАЛУ



Для інформаційних аналітиків це видання є довідково-методичним керівництвом з налаштування персональних моніторингових каналів, валідації результатів пошуку та інтеграції аналітичних дашбордів у корпоративні середовища. Особливу увагу приділено алгоритмам побудови сюжетних ланцюжків, автоматичних дайджестів та матриць взаємозв'язків, що суттєво скорочують час на первинну обробку великих масивів даних.

В управлінських процесах посібник орієнтовано на інтеграцію моніторингу в контур систем підтримки прийняття рішень (СППР). Описані методичні підходи дозволяють трансформувати оперативні новинні потоки у структуровані інсайти для кризового реагування, формування медіа-стратегій, конкурентної розвідки та оцінки регуляторних ризиків.

Матеріал посібника рекомендується використовувати ітеративно: від опрацювання базових сценаріїв пошуку до налаштування гібридних процесів з семантичним прогнозуванням, обов'язково поєднуючи автоматизовані висновки з експертною верифікацією та посиланнями на першоджерела. Такий підхід гарантує не лише технологічну ефективність, а й методологічну прозорість, відтворюваність результатів та відповідність сучасним стандартам аналітичної діяльності.

Цей розділ присвячено фундаментальним теоретико-методологічним засадам інформаційного моніторингу, які становлять концептуальну основу як для класичних пошукових технологій, так і для сучасних аналітичних рішень, що базуються на великих мовних моделях. У цифрову епоху інформаційний простір Інтернету характеризується експоненційним зростанням обсягів даних, високою динамікою оновлень та гетерогенністю джерел, що зумовлює необхідність системного, відтворюваного підходу до відбору, верифікації та інтерпретації новинних потоків. Моніторинг перестає бути суто технічною процедурою пошуку та трансформується в інструмент підтримки прийняття управлінських, комунікаційних та наукових рішень.

У розділі послідовно розглядаються головні аспекти, що визначають ефективність моніторингових процесів: структура та характеристики інформаційного середовища, завдання та критерії якості інформаційних вибірок, технологічна еволюція методів пошуку – від формалізованих булевих конструкцій до семантико-аналітичних процедур нового покоління. Окрема увага приділяється методології оцінки релевантності, достовірності, оперативності та повноти даних, які є незамінними для побудови стабільних інформаційних каналів та подальшої інтеграції результатів у контур систем підтримки прийняття рішень. Наведені теоретичні положення є підґрунтями для подальшого опису програмно-технологічного ядра системи InfoStream, архітектури гібридної аналітики та практичних сценаріїв її застосування в реальних робочих процесах.

## 1.1. Інформаційний простір Інтернет

Інтернет як інформаційний простір є розподіленим, гетерогенним та безперервно змінюваним середовищем, у якому новинні потоки формуються мільйонами незалежних джерел без єдиного стандарту публікації чи централізованої модерації. На відміну від статичних архівів або вузькопрофільних баз даних, цей простір характеризується мовною та тематичною різноманітністю, а також постійною зміною структури контенту. Для систем контент-моніторингу це означає необхідність інтегрувати різноманітні дані в єдине індексоване сховище з мінімальною затримкою між моментом публікації та появою матеріалу в пошуковому контурі, зберігаючи при цьому відтворюваність, верифіковуваність та аналітичну придатність вибірок.

**Класифікація та архітектура джерел.** Джерельна база новинного моніторингу охоплює декілька функціональних рівнів: офіційні інформаційні агентства та державні портали; комерційні ЗМІ (онлайн-видання, телерадіомовні компанії, друковані електронні архіви); тематичні блоги, експертні спільноти та аналітичні центри; соціальні мережі та мікроблогінг-платформи, де новинний імпульс часто формується раніше за традиційні медіа; RSS-стрічки, а також архівні сховища з історичними публікаціями. Кожне джерело має власну частоту оновлення, мову подання, формат контенту (текст, мультимедіа, структуровані дані) та рівень редакційного контролю. У практиці моніторингу критично важливою є не лише кількісна репрезентативність джерел, а їхня здатність генерувати первинні новинні події, а також стабільність доступу та прозорість редакційної політики, що впливає на достовірність подальшого аналізу.

**Характеристики інформаційних потоків.** Новинні потоки в Інтернеті підкоряються законам масштабного розподілу: невелика частка джерел формує більшість унікальних подій, тоді як переважна більшість публікацій є варіаціями, коментарями, ретрансляціями або аналітичними реакціями. Обсяг щоденного новинного трафіку вимірюється сотнями тисяч документів, що створює проблему «інформаційного шуму» – співвідношення сигналу до релевантних даних без додаткової фільтрації часто не перевищує 15–20%. Додатковими ускладненнями виступають змістовні дублі, семантична близькість документів, мовна неоднорідність, а також маніпулятивні публікації, що імітують органічний розвиток подій. Класичний повнотекстовий пошук без лінгво-статистичної обробки та багаторівневої фільтрації не здатний ефективно відокремити ядро інформаційного процесу від його периферії, що зумовлює необхідність впровадження автоматичного виявлення дублікатів, побудови інформаційних портретів та динамічного уточнення запитів.

**Динаміка оновлень.** Динаміка оновлень новинного простору має виражений часовий характер і може бути структурована за трьома рівнями:

(1) оперативний – публікації з інтервалом від хвилин до годин, характерні для стрічкових агентств, прес-служб та соціальних мереж; (2) періодичний – щоденні, тижневі або місячні випуски аналітичних видань, звітів, оглядів та нормативних документів; (3) ретроспективний – накопичення історичних зрізів, що дозволяє будувати довгі ряди та відстежувати еволюцію дискурсів. Часові ряди новинних потоків демонструють циклічність (добові, тижневі, сезонні патерни), спалахи активності під час кризових або подієвих ситуацій, а також поступове «затухання» інтересу до теми після проходження інформаційної хвилі. Аналіз цієї динаміки вимагає не лише фіксації факту публікації, а й оцінки її ваги в контексті загального потоку, що є підґрунтям для виявлення аномалій, прогнозування обсягу потоку та семантичного моделювання розвитку тенденцій.

На цей час моніторинг перестає бути процедурою простого пошуку за ключовими словами. Він трансформується у багаторівневий аналітичний процес: від сканування та нормалізації джерел до фільтрації змістовних дублів, побудови інформаційних портретів, автоматичного реферування та аналізу часових рядів. Саме симбіоз перевіреної класичної архітектури повнотекстового індексування, формалізованої мови запитів та сучасних методів штучного інтелекту (LLM для семантичного групування та витягнення сутностей, RAG-архітектура для прив'язки висновків до першоджерел, статистичні та нейронні моделі для прогнозу динаміки) дозволяє трансформувати хаотичний потік даних у структуровану основу для підтримки прийняття рішень. Наявність ретроспективного фонду, що безперервно формується протягом понад трьох десятиліть, забезпечує необхідну глибину для калібрування аналітичних моделей, виявлення довгострокових трендів та валідації семантичних прогнозів, що робить систему моніторингу не лише інструментом спостереження, а й компонентом прогнозно-аналітичного контуру сучасних СППР.

## **1.2. Моніторинг як інструмент підтримки прийняття рішень**

У сучасному цифровому середовищі інформаційний моніторинг трансформувався з допоміжної функції збору даних у стратегічну компоненту систем підтримки прийняття рішень. Його роль полягає у перетворенні неструктурованого потоку веб-публікацій та соціальних мереж на верифіковані та структуровані дані здатні впливати на управлінські, комунікаційні та дослідницькі рішення. Ефективність такого контуру залежить від чіткого визначення прикладних задач моніторингу та дотримання високих критеріїв якості інформаційних вибірок на всіх етапах процесу.

Функціональне призначення системи моніторингу визначається спектром управлінських потреб, до яких належать:

- Ситуаційна обізнаність та раннє попередження – своєчасне виявлення кризових сигналів, регуляторних ініціатив, ринкових змін або інформаційних атак.
- Виявлення тенденцій та прогнозування – аналіз довгострокової динаміки тематичних дискурсів, циклічності та семантичних зрушень в інформаційному просторі.
- Оцінка ефективності комунікаційних та управлінських впливів – вимірювання репутації, аналіз резонансу прес-релізів, кампаній або законодавчих змін.
- Верифікація інформаційного середовища – відокремлення органічного розвитку події від штучного втручання, детекція маніпулятивних патернів, координованих нарративів та інформаційних спалахів.
- Ретроспективний аналіз – використання архівного масиву даних, що безперервно накопичується протягом десятиліть, для навчання аналітичних моделей та виявлення повторюваних сценаріїв.

Для того щоб результати моніторингу могли інтегруватися в системи прийняття рішень без додаткової «ручної» фільтрації, вони мають відповідати низці вимог:

- Релевантність (тематична відповідність) як здатність системи відібрати матеріали, що точно відповідають семантиці запита, мінімізуючи тематичне відхилення дрейф.
- Повнота та репрезентативність, тобто охоплення всіх значущих джерел у межах заданої тематики без системних пропусків.
- Оперативність – мінімальна затримка між моментом публікації матеріалу в мережі та його появи в індексованому контурі системи.
- Достовірність – гарантована прив'язка кожного аналітичного висновку до конкретного першоджерела, дати публікації та оригінального уривка тексту.
- Мінімізація інформаційного шуму та дублів: автоматичне виявлення змістовних дублів (дублікатів) та семантично близьких матеріалів.
- Аналітична придатність – структурованість даних, наявність розширених метаданих, можливість експорту у стандартних форматах та прямої інтеграції із зовнішніми платформами.

Класичне ядро системи (повнотекстовий пошук, формалізовані запити, ретроспективний фонд) забезпечує точність, контрольованість та відтворюваність вибірки. Сучасні аналітичні модулі (LLM для семантичного групування та автоматичного реферування, RAG для прив'язки до джерел та даних, методи аналізу часових рядів для виявлення патернів інформаційних операцій, аномалій та прогнозування динаміки) трансформують відібрану вибірку у динамічну аналітичну модель. Саме таке поєднання дозволяє перейти від пасивного спостереження до активного прогнозування, де кожен етап – від конструювання запиту до генерації структурованого звіту –

підпорядкований єдиній методологічній логіці підтримки обґрунтованих рішень.

### 1.3. Еволюція підходів: від к пошуку до семантико-аналітичних процедур

Початок розвитку автоматизованого пошуку інформації в мережі Інтернет був зумовлений необхідністю швидкого знаходження документів за формальними запитами. Класичні інформаційно-пошукові системи (ПС) базувалися на індексації слів, булевій логіці та ранжуванні за частотою входження. В умовах зростання обсягів даних пошук за ключовими словами перетворився на інструмент первинної фільтрації, що потребує подальшого аналітичного збагачення.

Першим кроком у подоланні цих обмежень стало розроблення формалізованих мов запитів, що поєднують булеву логіку з контекстними операторами та морфологічною обробкою. У архітектурі систем моніторингу нового покоління, зокрема InfoStream, ядром виступає повнотекстова ПС InfoReS, яка реалізує пріоритезовані оператори контекстного слідування (~) та контекстної близькості (@), дозволяючи фіксувати не лише наявність термінів, а й їхнє семантичне та синтаксичне взаєморозташування в тексті. Додавання морфологічного аналізу, точних входжень (|), правих скорочень та автоматичної детекції змістовних дублів трансформувало пошук з механічного збігу в контрольовану лінгво-статистичну процедуру. Цей етап заклав фундамент відтворюваності та прозорості вибірок, що є критично важливим для подальшого аналізу.

Наступна хвиля еволюції була пов'язана з інтеграцією методів Text Mining, що дозволили переходити від пошуку окремих документів до аналізу масивів. Важливим інноваційним рішенням стало побудова «інформаційних портретів» на основі частотно-статистичних моделей, які візуалізують вагові характеристики вибірки за рубриками, мовами, географією, персоналіями, тональністю та цифровою насиченістю. Це забезпечило динамічне уточнення запитів без зміни їхнього синтаксису. Паралельно з'явилися алгоритми автоматичного реферування, ранжування сюжетних ланцюжків, побудови гістограм динаміки понять та матриць взаємозв'язків рубрик. Система навчилася відповідати не лише на запит «що знайти?», а й на питання «що нового?», «про що пишуть найбільше?» та «які теми корелюють?», перетворюючи сирі результати пошуку на структуровані аналітичні об'єкти.

Зі зростанням обсягів історичних даних (ретроспективні фонди, що безперервно формуються протягом вже трьох десятиліть) акцент змістився у бік аналізу часових рядів та виявлення патернів. Класичні гістограми частоти появи понять доповнено алгоритмами детекції аномалій, волатильності медіауваги, хвильової динаміки інформаційних потоків на основі порівняння локального лексичного складу з довгостроковим фоном. Статистичні та гібридні моделі дозволили перейти від ретроспективної констатації до динамічного прогнозу обсягу потоку, циклічності тем та ймовірності

інформаційних спалахів. Це створило технологічний міст між описовою аналітикою та прогностичними моделями, де часовий розріз стає інструментом передбачення.

Сучасний етап еволюції визначається симбіозом перевіреної класичної архітектури з великими мовними моделями та архітектурою RAG (Retrieval-Augmented Generation). Замість ізольованої генерації текстів, сучасні процедури спочатку формують релевантну вибірку за строгими формальними критеріями, після чого LLM застосовується для глибокого семантичного кластерування, витягнення сутностей, оцінки тональності та стислого реферування. RAG-підхід гарантує, що кожен згенерований висновок прив'язаний до верифікованих першоджерел, дат та уривків оригінального тексту, мінімізуючи ризик «галюцинацій» та забезпечуючи аудиторваність результатів. Семантичний прогноз розвитку тенденцій, побудований на основі LLM, доповнює статистичний прогноз динаміки, формуючи гібридну аналітичну модель, здатну оперувати як кількісними, так і якісними параметрами інформаційного простору.

Таким чином, еволюція підходів від пошуку до семантико-аналітичних процедур відображає фундаментальну зміну парадигми: від пасивного відбору документів до активного конструювання інформаційних моделей реальності. Класичне ядро (мови запитів, детекція дублів, ретроспективні фонди) забезпечує точність, контрольованість та відтворюваність вибірки, тоді як сучасні модулі (LLM, RAG, аналіз часових рядів, сюжетне моделювання) трансформують її у динамічну основу для прийняття рішень.

#### **1.4. Релевантність, достовірність, оперативність, повнота**

Ефективність будь-якої системи контент-моніторингу визначається не обсягом зібраних даних, а їхньою здатністю забезпечувати прийняття обґрунтованих рішень. Ця здатність формується на перетині чотирьох базових критеріїв якості інформаційних каналів: релевантності, достовірності, оперативності та повноти. У науково-методичному контурі InfoStream ці параметри розглядаються не як взаємопов'язані змінні, що підлягають кількісному вимірюванню та інтеграції в автоматизовані аналітичні процедури.

Релевантність визначає ступінь відповідності відібраної інформації семантичному ядру запиту та управлінській задачі. У класичному контурі вона забезпечується формалізованою мовою запитів InfoReS, контекстними операторами слідування (~) та близькості (@), морфологічною обробкою та багаторівневою фільтрацією за рубриками, мовами, країнами та джерелами. На сучасному етапі релевантність посилюється за рахунок інтеграції LLM, що виконують семантичну кластеризацію, виявлення прихованих тематичних зв'язків та автоматичне ранжування документів за контекстуальною близькістю до цілі моніторингу.

Достовірність відображає ступінь верифікованості даних та відсутність маніпулятивного впливу на інформаційний потік. У практиці моніторингу достовірність забезпечується трьома рівнями: (1) ідентифікація та ранжування джерел за редакційною політикою, історичною стабільністю та рівнем первинності інформації; (2) автоматична детекція змістовних дублів та семантично близьких публікацій, що дозволяє відокремити оригінальний сигнал від інформаційного шуму; (3) обов'язкова прив'язка кожного аналітичного висновку до першоджерела.

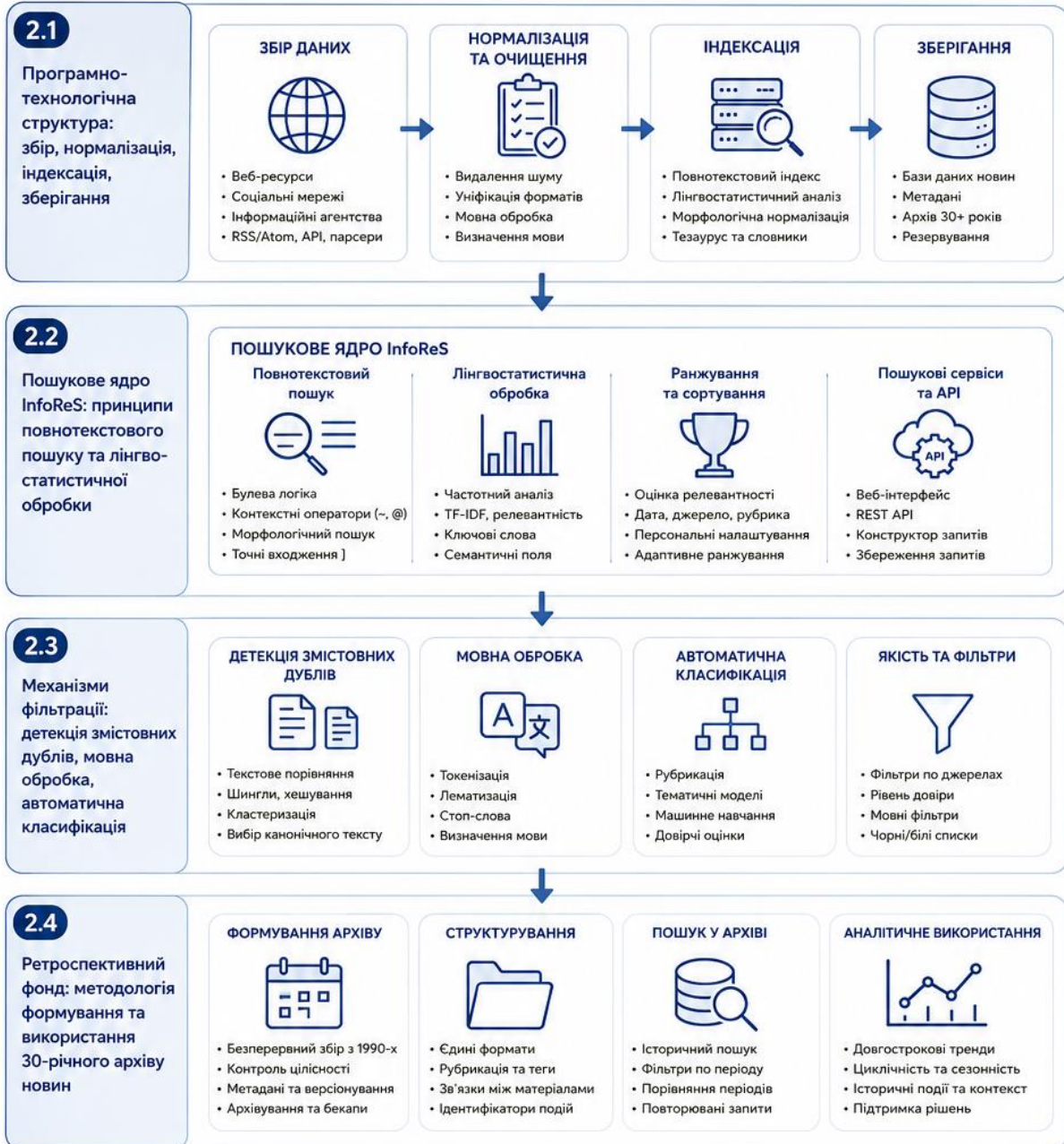
Оперативність характеризує мінімальну затримку між моментом появи публікації в мережі та її доступністю в аналітичному контурі системи. У кризових сценаріях, моніторингу регуляторних змін або ринкових коливань цей параметр є критичним. InfoStream забезпечує оперативність за рахунок безперервного сканування джерел з періодичністю оновлення індексу (до 15 хвилин для стрічкових агентств та офіційних порталів), автоматичної нормалізації контенту та підтримки режимів доставки в реальному часі (email-розсилки, RSS-потоки, веб-віджети, API).

Повнота визначає ступінь охоплення всіх значущих джерел, мовних версій, регіональних сегментів та типів документів у межах заданої тематики. Низька повнота призводить до системних пропусків, упередженості вибірок та хибних управлінських висновків. У InfoStream повнота забезпечується структурованим каталогом джерел, підтримкою багатомовного індексування, інтеграцією офіційних порталів, ЗМІ, соціальних мереж, експертних блогів та RSS-стрічок, а також наявністю ретроспективного фонду, що безперервно формується понад 30 років.



## РОЗДІЛ 2. ТЕХНОЛОГІЧНЕ ЯДРО ТА АРХІТЕКТУРА СИСТЕМИ INFOSTREAM

### БЛОК-СХЕМА ВИКЛАДУ МАТЕРІАЛУ



### РЕЗУЛЬТАТ РОЗДІЛУ 2



Розділ 2 описує технологічне ядро InfoStream – від збору новин до формування перевіреної, структурованої та доступної для аналітики бази даних.

Архітектура InfoStream побудована за модульним принципом і реалізує багаторівневу технологічну процедуру. На першому етапі система здійснює безперервне сканування гетерогенних джерел (новинні портали, офіційні сайти, соціальні мережі, RSS-стрічки), після чого виконується автоматична нормалізація контенту, лінгво-статистична обробка, індексація та збереження в єдиному сховищі. Програмно-технологічним фундаментом цього процесу виступає повнотекстова інформаційно-пошукова система InfoReS, що забезпечує високоточний пошук із підтримкою формалізованої мови запитів, контекстних операторів, морфологічного аналізу та багаторівневої фільтрації.

На цій класичній основі побудовано гібридний аналітичний контур. Відібрані за строгими формальними критеріями масиви документів подаються на вхід модулям великих мовних моделей для семантичної кластеризації, витягнення сутностей, оцінки тональності та автоматичного реферування. Архітектура Retrieval-Augmented Generation інтегрована безпосередньо в процедури видачі, що гарантує прив'язку кожного згенерованого висновку чи резюме до першоджерела, дати публікації та оригінального уривка, мінімізуючи ризик неперевірених тверджень. Паралельно функціонує підсистема аналізу часових рядів, яка трансформує хронологічні метадані, частотність понять та динаміку джерельного потоку у інтерактивні гістограми, патерни інформаційних операцій та прогнозні моделі.

Унікальним методологічним активом системи є ретроспективний фонд новин, який безперервно формується понад 30 років. Цей архів забезпечує необхідну часову глибину досліджень, калібрування аналітичних алгоритмів, валідації семантичних прогнозів та відтворення історичних інформаційних зрізів без втрати контексту чи втрати видалених з мережі матеріалів. Саме поєднання стабільного ядра, автоматизованої фільтрації шуму та дублів і сучасних ШІ-модулів трансформує хаотичний цифровий потік у керовану аналітичну модель, придатну для інтеграції в контур систем підтримки прийняття рішень.

## **2.1. Програмно-технологічна структура**

Програмно-технологічна архітектура системи InfoStream реалізує замкнену багаторівневу процедуру обробки інформації, який забезпечує трансформацію потоку веб-публікацій у верифіковані, індексовані та придатні для аналітичної обробки масиви даних. Архітектура побудована за модульним принципом і функціонально поділяється на три взаємопов'язані комплекси: збір та селективного поширення інформації, інтерактивного доступу до повнотекстових баз даних та контент-моніторингу на основі Text Mining. Кожен етап технологічного циклу – від сканування джерел до довгострокового зберігання – підпорядкований єдиній методологічній логіці

забезпечення оперативності, повноти, релевантності та відтворюваності результатів.

**Збір інформації.** Фундаментом системи є автоматизований режим сканування мережевих ресурсів у реальному часі. Система безперервно опитує тисячі інформаційних веб-сайтів, новинних агентств, офіційних порталів, соціальних мереж та RSS-стрічок. Головною особливістю при цьому є висока частота оновлення індексу (до 15 хвилин для пріоритетних джерел), що забезпечує мінімальну затримку між моментом публікації та появою матеріалу в пошуковому контурі. На відміну від глобальних пошукових систем, сканування в InfoStream не піддається впливу таргетингу, маркетингу чи самоцензури, що гарантує об'єктивність та прозорість первинної вибірки. Механізми збору адаптовані до гетерогенності джерел, система коректно обробляє різні мови, формати подання контенту, рівні цифрової насиченості та індивідуальну динаміку оновлення кожного окремого ресурсу.

**Нормалізація та структурування.** Отримані дані надходять на етап автоматичної нормалізації, де приводяться до єдиного текстового формату. Процес включає очищення від HTML-розмітки, скриптів, рекламних блоків та навігаційних елементів, виділення основного змісту публікації (заголовки, текст, метадані, дата, джерело), а також уніфікацію кодувань. Паралельно здійснюється попередня класифікація інформації за семантичними ознаками та тематичними рубриками. На цьому етапі застосовуються лінгво-статистичні алгоритми для визначення мови документа, оцінки його розміру та цифрової насиченості, що є критично важливим для подальшої фільтрації.

**Індексація та пошукове ядро.** Програмно-технологічним фундаментом обробки індексованих даних виступає повнотекстова інформаційно-пошукова система InfoReS. Індесування реалізує підтримку формалізованої мови запитів із багаторівневою булевою логікою, контекстними операторами, морфологічним аналізом природних мов та можливістю пошуку за точними входженнями чи правими скороченнями. Системою автоматично детектуються змістовні дублі та семантично близькі документи на основі лінгво-статистичних алгоритмів, що забезпечує високу щільність корисного сигналу у видачі. Індекс структуровано так, щоб підтримувати не лише класичний лексичний пошук, а й векторні представлення документів, необхідні для архітектури RAG. Це дозволяє генерувати аналітичні резюме з прив'язкою до першоджерел, мінімізуючи ризик неперевіраних тверджень.

**Зберігання та організація даних.** Нормалізовані та індексовані матеріали зберігаються у захищеному інформаційному сховищі, архітектурно розподіленому на кілька логічних контурів: оперативна база (документи за останні 7 днів, оновлюється в реальному часі), ретроспективна база (архів, що безперервно формується понад 30 років), а також база даних повідомлень із соціальних мереж. Унікальною методологічною перевагою є принцип незалежності архіву від стану веб-джерела: навіть якщо публікація видалена

з оригінального сайту, вона зберігається в інформаційному сховищі системи, забезпечуючи цілісність та відтворюваність історичних зрізів. Сховище оптимізовано для високошвидкісного доступу, підтримує санкціонований багаторівневий контроль доступу та інтегрується з підсистемами аналізу часових рядів, де хронологічні метадані трансформуються у динамічні гістограми, патерни інформаційних операцій та прогнозні моделі.

Така програмно-технологічна структура трансформує хаотичний цифровий потік у керовану аналітичну модель. Класичне ядро (InfoReS, нормалізація, детекція дублікатів, ретроспективний архів тощо) забезпечує точність, контрольованість та відтворюваність вибірки, а інтеграція з сучасними методами векторного пошуку, LLM-семантики та формує підґрунтя для гібридної аналітики, придатної для безпосередньої інтеграції в контур систем підтримки прийняття рішень.

## **2.2. Пошукове ядро InfoReS**

Програмно-технологічним фундаментом системи InfoStream виступає повнотекстова інформаційно-пошукова система InfoReS. Саме ця компонента забезпечує первинний цикл обробки даних: від безперервного індексування нових публікацій до формування релевантних вибірок за складними пошуковими предписаннями. На відміну від загальних пошукових систем, InfoReS архітектурно орієнтований на специфіку новинних потоків, де критичними є оперативність оновлення індексу, точна семантична відповідність запиту та здатність фільтрувати інформаційний шум у реальному часі.

Ядро системи InfoReS працює з інтенсивним потоком документів, що надходять від тисяч веб-ресурсів, офіційних порталів, RSS-стрічок та соціальних мереж. Кожна публікація проходить етап нормалізації, очищення від службового коду, мовної сегментації та індексації в єдиному повнотекстовому сховищі. Важливою технологічною особливістю є циклічність оновлення індексу: бази даних системи поповнюються з періодичністю до 15 хвилин, що забезпечує мінімальну затримку між моментом публікації матеріалу в мережі та його доступністю в пошуковому контурі.

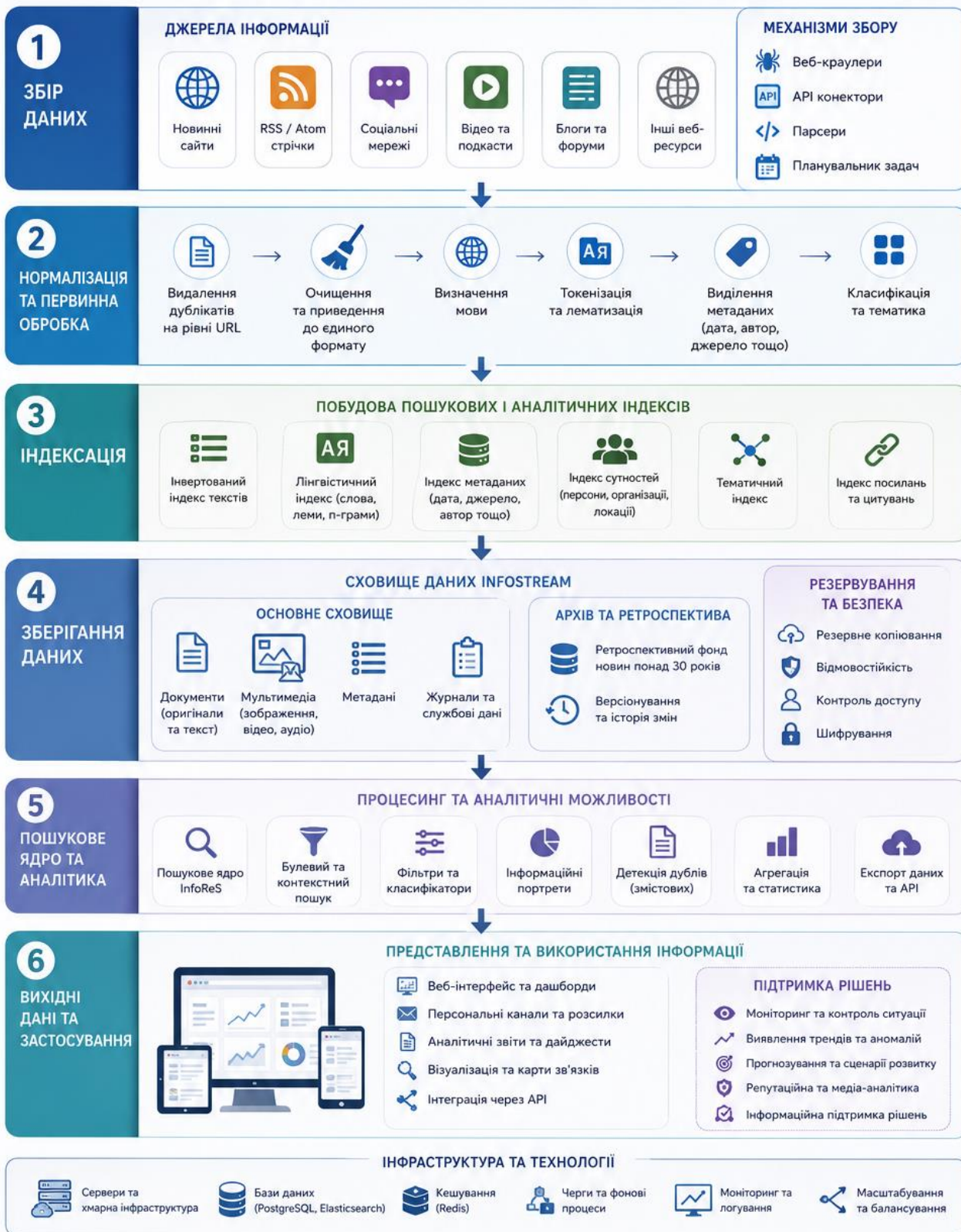


# InfoStream

Завжди у курсі

## ПРОГРАМНО-ТЕХНОЛОГІЧНА СТРУКТУРА INFOSTREAM

Від збору новин до аналітичних інсайтів та підтримки рішень



Взаємодія користувача з ядром системи InfoReS здійснюється через спеціалізовану мову запитів, що поєднує природномовні терміни з формалізованими логічними та контекстними операторами. Базову конструкцію складають булеві зв'язки: (&/+) – AND, (|/) – OR, (!/^) – NOT, пріоритети яких регулюються круглими дужками для формування багаторівневих предписань. Унікальні можливості InfoReS реалізуються через оператори контекстного слідування (~/-) та контекстної близькості (@), що дозволяють задавати допустиму відстань між термами в тексті документа з точністю до одного–двох слів. Така гнучкість є критично важливою для відокремлення цільових подій від тематичного дрейфу та формування вузькоспеціалізованих інформаційних каналів. Морфологічний режим забезпечує автоматичне відсікання закінчень та охоплення всіх словоформ, тоді як символ точного входження (□) дозволяє фіксувати конкретні назви, бренди чи аббревіатури без розширення вибірки за суміжними лексемами.

Лінгво-статистичний апарат обробки. Пошукове ядро не обмежується синтаксичним збігом слів. Його архітектура інтегрує алгоритми лінгво-статистичного аналізу, що формують основу для подальшої аналітики та візуалізації. До основних механізмів належать:

- Автоматична детекція дублікатів (змістовних дублів) – система ідентифікує не лише повні копії, а й документи зі значною семантичною близькістю, маркуючи їх та групуючи для зручності аналізу. Виявлення дублікатів базується на частотно-статистичному порівнянні лексичних ваг та структурної близькості фрагментів.
- Пошук подібних документів – на основі графу семантичних зв'язків InfoReS забезпечує перехід від знайденого матеріалу до логічно близьких публікацій.
- Формування інформаційних портретів – кожна вибірка документів автоматично профілюється за основними метриками – рубриками, мовами, географією, персоналіями, тональністю, рівнем цифрової насиченості та розміром тексту. Всі характеристики ранжуються за вагою в інформаційному каналі, надаючи користувачу інструмент для динамічного уточнення запиту через візуальну інтерактивну панель.
- Частотно-статистичне реферування – на базі вагових коефіцієнтів окремих термінів, речень та абзаців визначаються найбільш інформативні фрагменти, що лягає в основу автоматичного формування дайджестів та знаходження сюжетних ланцюжків.

Інтеграція в сучасний аналітичний контур. У класичній архітектурі InfoReS забезпечує точність, контрольованість та високу швидкість відпрацювання запитів до великих масивів даних. Відібрані за формальними та лінгво-статистичними критеріями документи передаються на вхід генеративним моделям та RAG-процедурам. Завдяки детермінованій логіці InfoReS, штучний інтелект отримує вже верифікований, релевантний та структурований контекст, що суттєво знижує ризик генерації неперевіраних

тверджень та підвищує якість автоматичних резюме, хвильових прогнозів та аналітичних довідок. Це гарантує методологічну прозорість, відтворюваність результатів та відповідність вимогам систем підтримки прийняття рішень.

### 2.3. Механізми фільтрації

В умовах високої інтенсивності інформаційних потоків (понад 100 000 документів щодобово з тисяч веб-ресурсів) ефективність моніторингу визначається не обсягом зібраних даних, а їхньою придатністю до змістовної аналітичної обробки. Саме на етапі первинної обробки та фільтрації формується якість інформаційної вибірки, яка згодом подається на вхід аналітичним модулям. У архітектурі InfoStream ця задача вирішується через три взаємопов'язані механізми: детекцію змістовних дублів, лінгвістичну обробку та автоматичну класифікацію.

Детекція змістовних дублів. Новинний простір характеризується високою частотою ретрансляцій, коли одна подія публікується десятками джерел з незначними редакційними змінами, синонімічними замінами або зміною структури абзаців. Класичний пошук за ключовими словами без спеціалізованої фільтрації призводить до зростання «інформаційного шуму» та ускладнює оцінку реального інформаційного резонансу. У системі InfoReS, що є технологічним ядром InfoStream, реалізовано лінгво-статистичний алгоритм виявлення як точних копій, так і документів зі змістовною близькістю. Алгоритм працює на етапі індексації та порівнює частотні ваги термінів, структуру синтаксичних конструкцій та лексичний розподіл у межах документів. Результатом є автоматичне маркування дублів у видачі, агрегація релевантних матеріалів навколо первинного джерела та корекція метрик частотності. У сучасному аналітичному контурі цей механізм є критично важливим для коректної роботи з LLM-моделями: мінімізація дублів запобігає посиленню «галюцинацій» ШІ, забезпечує чистоту вибірки для побудови часових рядів та дозволяє LLM-моделям працювати з унікальними семантичними ядрами подій, а не з їхніми повтореннями.

Для забезпечення точності пошуку та стабільності аналітичних вибірок InfoStream застосовує багаторівневу лінгвістичну обробку. На базовому рівні реалізовано морфологічний аналізатор, що автоматично відокремлює закінчення та приводить слова до базових форм (лем), що уможливлює охоплення всіх словоформ за одним запитом без втрати релевантності. Для тонкого налаштування передбачено режим точного входження (символ `]` ) та оператори контексту, які враховують порядок та допустиму відстань між термами в тексті. Система також автоматично оцінює «цифрову насиченість» документів та їхній об'єм, що дозволяє фільтрувати аналітичні звіти, прайслисти чи короткі новинні зведення.

Після фільтрації дублікатів та лінгвістичної нормалізації масив документів підлягає структурованню. У InfoStream ця задача реалізується

через так званий «Класифікатор-навігатор» та систему тематичних рубрик. На відміну від статичних тегів, класифікація в системі має динамічний характер – вона будується «на льоту» на основі результатів пошукового запиту, формуючи інтерактивні «пошукові папки». Алгоритм аналізує частотно-статистичний розподіл термінів, географічних назв, персоналій, мов та джерел у вибірці, після чого кластеризує документи за семантичною близькістю. Користувач отримує візуалізоване дерево уточнень, де кожна гілка відображає вагу підтеми в загальному потоці та дозволяє миттєво звзвити або розширити інформаційний канал.

Названі механізми трансформують сирий новинний потік у контрольовану, відтворювану та аналітично придатну вибірку. Вони забезпечують технологічний фундамент, на якому базуються сучасні модулі штучного інтелекту, архітектура RAG та аналіз часових рядів, гарантуючи, що кожен етап процесу моніторингу, від сканування джерела до генерації прогнозу, спирається на верифіковані, очищені від шуму та структуровані дані.

#### **2.4. Ретроспективний фонд**

Ретроспективний інформаційний фонд є одним із основних технологічних та методологічних активів системи InfoStream. Його формування розпочалося з моменту введення першої версії системи в експлуатацію та триває безперервно, забезпечуючи накопичення, збереження та структурований доступ до новинних матеріалів протягом понад трьох десятиліть. На відміну від фрагментарних веб-архівів або кешованих копій пошукових систем, ретроспективний фонд InfoStream будується на основі стандартизованого технологічного циклу, що гарантує цілісність даних, відтворюваність історичних зрізів та їхню придатність для сучасних аналітичних і дослідницьких задач.

Наповнення ретроспективного фонду здійснюється автоматизовано через послідовність операцій, а саме, сканування відкритих джерел Інтернету, нормалізації контенту до єдиного текстового формату, лінгво-статистичної класифікації, індексації та збереження у централізованому сховищі. Критично важливим є принцип незалежності архіву від життєвого циклу джерела: навіть якщо публікація видалена з оригінального веб-сайту або ресурс припиняє функціонування, її повний текст, метадані та контекстні зв'язки залишаються незмінними в інформаційному сховищі системи. Це усуває проблему «цифрової ерозії» та забезпечує стабільну основу для аналітичних досліджень.

Архітектура сховища реалізована за модульним принципом, кожний запис зберігається з розширеним набором метаданих: дата та час публікації, джерело, мова, тематична рубрика, географічна прив'язка, витягнені сутності, оцінка тональності, рівень цифрової насиченості та структурні характеристики тексту. Базовий обсяг фонду на початок експлуатації

перевищував 50 млн записів; станом на 2026 рік архів збільшився на порядок, охоплюючи новинні потоки України та закордонних веб-сайтів і міжнародних інформаційних агентств.

Ретроспективний фонд не є пасивним сховищем; він інтегрований з аналітичним інструментарієм як джерело контексту, верифікації та навчання. Його застосування реалізується у кількох методологічних напрямках:

1. Архів дозволяє відстежувати еволюцію дискурсів, циклічність медіауваги, зміни в регуляторному полі, динаміку медіа-репутації та структурні зрушення в інформаційних потоках. Користувач може формувати історичні зрізи за довільними часовими інтервалами, порівнювати семантичні патерни між періодами та виявляти повторювані сценарії розвитку подій.
2. Великі мовні моделі, інтегровані в InfoStream для семантичної кластеризації, реферування та витягнення сутностей, потребують репрезентативного історичного корпусу. Ретроспективний фонд забезпечує дані для тонкого налаштування промптів.
3. Архітектура Retrieval-Augmented Generation у системі моніторингу спирається на індексовані джерела та документи. Завдяки збереженню повних текстів та метаданих навіть для видалених з інтернету публікацій, кожний згенерований аналітичний висновок може бути підтверджений прямим посиланням на архівний документ, дату та уривок оригіналу. Це усуває ризик «галюцинацій» ШІ та робить результати придатними для офіційних довідок, експертизи та академічних публікацій.

Тридцятирічний ретроспективний фонд створює унікальну базу для статистичного та машинного моделювання новинних потоків:

- Аналіз часових рядів. Довготривалі послідовності частотності понять, активності джерел, змін тональності та обсягу потоку дозволяють застосовувати сучасні методи для виявлення структурних зламів, аномалій, циклічності та латентних трендів.
- Статистичні моделі, навчені на історичних даних, прогнозують обсяг тематичних інформаційних потоків, ймовірність інформаційних операцій, спалахів та рівень медіа-резонансу.
- Великі мовні моделі, отримуючи історичний контекст розвитку теми, генерують сценарії змісту тенденцій: ймовірні напрямки дискурсу, очікувані реакції стейкхолдерів, потенційні точки кристалізації нових подій. Поєднання динамічного та семантичного прогнозу трансформує моніторинг з реактивного спостереження у превентивний аналітичний інструмент.

Ретроспективний фонд експлуатується з дотриманням стандартів дослідницької інфраструктури, а саме, цілісності даних, структурованого доступу, аудиту та відтворюваності:

- архівні записи є незмінними після індексації; версіонування та контрольні суми гарантують відсутність несанкціонованих змін;
- користувачі можуть здійснювати часове фільтрування, інтеграцію з ВІ-платформами та автоматизацію аналітичних процедур;
- кожен запит до архіву фіксується з метаданими пошуку, що дозволяє незалежно відтворити вибірку, перевірити релевантність та валідувати аналітичні висновки в рамках СППР.

Таким чином, 30-річний ретроспективний фонд InfoStream виступає не лише технологічним сховищем, а й методологічним фундаментом сучасної аналітики. Він забезпечує глибину контексту, відтворюваність результатів, надійну верифікацію ШІ-генерацій та статистичну базу для прогнозування, трансформуючи систему моніторингу з інструменту збору даних у повноцінний компонент інтелектуальної підтримки прийняття рішень.



## РОЗДІЛ 3. КЛАСИЧНИЙ ІНСТРУМЕНТАРІЙ: МОВА ЗАПИТІВ ТА КОНСТРУКТОР ПОШУКУ

### ЛОГІКА ВИКЛАДУ МАТЕРІАЛУ



### ПІДСУМОК РОЗДІЛУ 3

Класичний інструментарій InfoStream забезпечує точний пошук, гнучке налаштування та глибоке розуміння інформаційного масиву для подальшої аналітичної обробки та прийняття рішень.



Сформулювати запит



Налаштувати та уточнити



Проаналізувати портрет вибірки



Отримати релевантні результати



Передати в аналітику та прийняття рішень

Якщо попередні розділи окреслили архітектуру системи, принципи формування ретроспективного фонду та критерії якості інформаційних каналів, то тут розкривається механізм їхнього формування – класичний пошуковий контур, що залишається незамінним фундаментом процесу контент-моніторингу. Ефективність сучасної аналітики, включно з модулями штучного інтелекту та архітектурою RAG, прямо залежить від повноти і точності первинного відбору даних.

Основою цього контуру виступає повнотекстова інформаційно-пошукова система InfoReS та її формалізована мова запитів. Окрему увагу у цьому розділі приділено візуально-аналітичним інструментам динамічного уточнення, інформаційному портрету, який трансформує частотно-статистичні характеристики масиву документів у інтерактивні фільтри за рубриками, мовами, географією, тональністю, персоналіями та рівнем цифрової насиченості, та адаптивному класифікатору-навігатору, що реалізує принцип інтерактивних «пошукових папок» для семантичного групування результатів без повторного формулювання запиту.

Важливо підкреслити, що класичний інструментарій InfoStream не суперечить сучасним ІІІ-модулям, а формує для них контрольоване середовище. Саме строга формалізація запиту та багаторівнева фільтрація гарантують, що на вхід LLM-алгоритмам, RAG-генераторам та моделям аналізу часових рядів надходить релевантна, верифікована та структурована вибірка.

### **3.1. Синтаксис та пріоритети операторів InfoReS**

Інформаційно-пошукова система InfoReS, що становить програмно-технологічне ядро InfoStream, реалізує формалізовану мову запитів, адаптовану до специфіки новинних потоків та аналітичних задач контент-моніторингу. На відміну від стандартних пошукових інтерфейсів, де запити інтерпретуються як неструктуровані ключові слова, мова InfoReS забезпечує детерміновану, відтворювану та логічно контрольовану вибірку документів. Її синтаксис побудовано на поєднанні термів (слів природної мови або їхніх правих скорочень), логічних операторів, контекстних конструкцій та засобів групування, що дозволяє формалізувати складні інформаційні сценарії без втрати семантичної точності.

#### ***Базові принципи синтаксису та обробки термів***

За замовчуванням система не чутлива до регістру символів, тому запити *інформація* та *ІНФОРМАЦІЯ* інтерпретуються ідентично. Кожне введене слово автоматично сприймається як праве скорочення (усічена основа), що забезпечує пошук похідних словоформ (наприклад, за запитом *банк* знайдуться *банку*, *банків*, *банківський*). Якщо необхідне точне входження слова без морфологічного розширення, використовується символ ] після терма (наприклад, *фонд]* відфільтрує *фонди*, *фонду*). Режим морфології

можна активувати або вимкнути в інтерфейсі: у вимкненому стані система працює виключно з правими скороченнями, у активованому – попередньо відкидає змінювані закінчення згідно з лінгвістичними правилами української та російської мов. Зірочка (\*) як символ довільного закінчення у мові InfoReS не застосовується, але не сприймається як помилка.

### ***Ієрархія та пріоритети операторів***

Коректна інтерпретація багаторівневих запитів забезпечується чітко визначеною послідовністю виконання операторів. Пріоритети розташовані за спаданням:

1. ~ – оператор контекстного слідування; @ – оператор контекстної близькості;
2. !, ^ – логічне заперечення (І-НІ);
3. &, +, пробіл – логічна кон'юнкція (І);
4. |, кома – логічна диз'юнкція (АБО).

Пріоритети можуть бути змінені за допомогою круглих дужок, що дозволяє конструювати вкладені логічні конструкції довільної глибини та усувати двозначність інтерпретації складних умов.

### ***Булева логіка та комбінування умов***

Оператори ТА (&, +, пробіл) і АБО (кома, |) реалізують класичні булеві функції кон'юнкції та диз'юнкції. Наприклад, запит

#### ***криза & енергетика | альтернативи***

дозволить відібрати документи, де обов'язково присутній термін криза, а також хоча б один із термів енергетика або альтернативи. Заперечення реалізується через ! або ^ у формі «І-НІ»: документ повинен відповідати лівій частині умови, але не містити термів правої частини, наприклад:

#### ***(кава | какао) ! (мелен | зерн) & розчин***

означає: знайти документи, де є **розчин**, а також **кава** або **какао** (або обидва), за умови відсутності мелен та зерн. Дужки тут фіксують рівні логіки, гарантуючи, що заперечення застосовується лише до вказаної групи термів перед виконанням кон'юнкції.

### ***Контекстне слідування (~) та близькість (@)***

Для відбору документів із просторовою прив'язкою термів використовуються два спеціалізовані оператори, що мають вищий пріоритет за булеві конструкції:

- Оператор ~ (контекстне слідування) відбирає пари або трійки термів, які зустрічаються в тексті у вказаній послідовності. Відстань між термами можна обмежити синтаксисом ~/n/, де n – максимальна кількість слів між ключовими термами. Наприклад, **чашка~/2/кави** знайде **чашка кави**, **чашка міцної кави**, але не **кава у чашці**.
- Оператор @ (контекстна близькість) працює аналогічно, але не фіксує порядок слідування термів. Запит **чашка@/2/кави** відбере **чашка**

*кави, кава у чашці, міцної кави чашка* тощо, за умови що між термами не більше двох слів.

Обмеження – операндом не може бути вираз у дужках. За замовчуванням відстань дорівнює нулю (слова йдуть підряд без проміжних лексем). Ці конструкції критично важливі для пошуку сталих словосполучень, назв установ, імен з прізвищами або технічних термінів, де розрив або зміна порядку змінює семантику або призводить до хибних спрацьовувань.

### ***Багаторівнева побудова запитів та методичні рекомендації***

Складні інформаційні сценарії потребують комбінування всіх рівнів синтаксису. Рекомендований алгоритм конструювання:

1. Визначення ядра запиту через булеві оператори (&, |);
2. Фіксація контекстних зв'язків через ~/@ з урахуванням допустимої відстані;
3. Виключення інформаційного шуму через !;
4. Групування дужками для контролю пріоритетів та ізоляції логічних блоків.

Типовою методичною помилкою є відсутність дужок при комбінуванні & та |, що призводить до некоректної інтерпретації через пріоритет операторів. Рекомендується завжди тестувати запит на обмеженій часовій вибірці, аналізувати «Інформаційний портрет» результатів та коригувати конструкцію шляхом додавання контекстних обмежень або уточнення термів. Мова InfoReS інтегрована в усі режими системи: онлайн-пошук, автоматичні email-розсилки, RSS-потоків та формування персональних каналів, що гарантує відтворюваність вибірок на всіх етапах технологічного процесу та забезпечує стабільну основу для подальшої LLM-аналітики.

## **3.2. Морфологічний пошук, точні входження, праві скорочення**

Ефективність пошукового запиту в системі InfoStream значною мірою визначається коректним налаштуванням лексико-морфологічних параметрів. Пошукове ядро InfoReS реалізує гнучкий механізм обробки термів, що дозволяє користувачеві свідомо балансувати між повнотою вибірки (recall) та її точністю (precision). Нижче детально розглянуто чотири базові інструменти керування формою слів, які формують фундамент конструювання запитів у класичному контурі системи.

### ***Морфологічний пошук***

За замовчуванням у системі активований режим морфологічної обробки. Під час індексації та виконання запиту лінгвістичний модуль автоматично відсікає змінювані закінчення, відокремлюючи незмінну основу терміна. Це дозволяє одночасно знаходити всі граматичні форми заданого слова. Наприклад, запит *банк* у морфологічному режимі поверне документи, що

містять «банку», «банки», «банківський», «банків», «банківській» тощо. Цей режим є критично важливим для охоплення повного тематичного поля, особливо в мовах з розвиненою флексією (українська, російська), де зміна закінчення не змінює базової семантики терміна. Користувач має можливість примусово вимкнути морфологію через перемикач у пошуковій формі, що переводить систему в режим суворого лексичного співпадіння. У науково-аналітичній практиці морфологію доцільно використовувати для широких галузевих запитів, моніторингу тенденцій та побудови інформаційних портретів.

### ***Точні входження (оператор ])***

У випадках, коли морфологічне розширення призводить до надмірного інформаційного шуму, омонімії або спотворення сенсу, InfoStream підтримує символ точного входження ]. Цей оператор інструктує пошуковий рушій шукати виключно задану графему без будь-якого автоматичного розширення закінчень. Наприклад, за запитом фон] система відфільтрує словоформи «фонди», «фондам», «фоновий» і поверне лише документи з точним написом «фон». Практичне застосування цього оператора рекомендоване у таких сценаріях:

- моніторинг абревіатур, акронімів та технічних кодів;
- пошук власних назв, прізвищ або географічних об'єктів зі спільним коренем, але різною семантикою;
- відстеження фіксованих юридичних, фінансових або нормативних конструкцій, де зміна закінчення змінює правовий або економічний зміст;
- фільтрація омонімічних основ, що належать до різних тематичних кластерів. Оператор ] може застосовуватися вибірково до окремих термів у складі багатокритеріального запиту.

### ***Праві скорочення***

Архітектура InfoReS підтримує концепцію правого усічення як базовий режим пошуку. Важливою особливістю системи є відсутність необхідності у використанні символу-джокера (наприклад, \*): у мові InfoStream зірочка після скороченого слова не застосовується, оскільки будь-яке введене слово автоматично сприймається як права основа. У поєднанні з вимкненою морфологією праве скорочення працює на рівні лексичного префікса, що суттєво зменшує помилки першого роду при обробці складних багатокомпонентних конструкцій. У комбінації з операторами контекстної близькості (@, ~) праве скорочення дозволяє фіксувати початок терміна без прив'язки до його граматичної ролі в реченні, що особливо корисно при моніторингу назв проектів, продуктів або ініціатив, які можуть змінювати закінчення залежно від синтаксичного контексту.

### ***Не чутливість до регістру літер***

Пошуковий контур InfoStream є повністю регістронезалежним (case-insensitive). Запити ІнфоСтрім, інфострім та ІНФОСТРІМ обробляються ідентично на рівні індексу та видачі. Ця властивість:

- усуває необхідність урахування регістру при копіюванні термінів з новинних стрічок, офіційних документів або соціальних мереж;
- гарантує, що форматування тексту, використання великих літер у заголовках або акцентування в оригінальних публікаціях не впливає на результати пошуку;
- спрощує автоматизацію процесів, оскільки API-запити та експортовані конфігурації не потребують додаткової нормалізації регістру.

### ***Круглі дужки як інструмент контролю пріоритетів***

Круглі дужки «(», «)» у конструкторі виконують функцію явного визначення рівнів обчислення, перевизначаючи стандартну ієрархію операторів. Кожна пара дужок формує логічний блок, який обробляється як єдиний операнд перед застосуванням зовнішніх зв'язків. Методично доцільно використовувати дужки у трьох випадках:

- Групування альтернатив: (банк | фінанс | кредит) забезпечує вибір будь-якого з синонімічних термінів як єдиного логічного модуля.
- Ізоляція заперечень: ! (спам | реклама | прес-реліз) гарантує, що виключення застосовується до всього блоку, а не до останнього слова.
- Багаторівнева вкладеність: ((Україн | Українськ) & (економік | ринк | фінанс)) ! (криптовалют | ICO) дозволяє будувати складні тематичні фільтри з чіткою семантичною межею.

При конструюванні запитів рекомендується дотримуватися правила «одна пара дужок – один логічний рівень», уникати надмірної вкладеності (понад 3 рівні) та перевіряти збалансованість дужок перед запуском пошуку. Конструктор автоматично підсвічує некоректні конструкції та пропонує корекцію синтаксису.

### ***Методичні рекомендації щодо застосування***

Для досягнення оптимального співвідношення релевантності та повноти в аналітичній практиці рекомендовано дотримуватись наступних правил конструювання запитів:

1. За замовчуванням використовувати морфологічний режим для широких тематичних моніторингових та побудови сюжетних ланцюжків.
2. При виявленні тематичного дрейфу або омонімії в «Інформаційному портреті» додавати оператор ] до критичних термінів, що генерують шум.

3. У комбінаціях з операторами близькості (@, ~) уникати морфологічного розширення, якщо точна послідовність або фіксована форма слів є ключовою для контексту події.
4. Не використовувати символ \* для скорочення: система автоматично трактує будь-який термін як праву основу, що спрощує синтаксис і зменшує ризик синтаксичних помилок.
5. Регулярно перевіряти структуру вибірки через інтерактивні фільтри портрета та коригувати запит шляхом поєднання морфологічного охоплення з точковим обмеженням, перш ніж передавати масив на вхід LLM-модулів.

Такий підхід перетворює лексико-морфологічні налаштування з суто технічного параметра інтерфейсу на інструмент аналітичного контролю якості даних, що безпосередньо впливає на точність подальшої семантичної кластеризації, верифікації джерел та побудови управлінських інсайтів.

### ***Системні фільтри: джерело, мова, дата та їх синтаксичні еквіваленти***

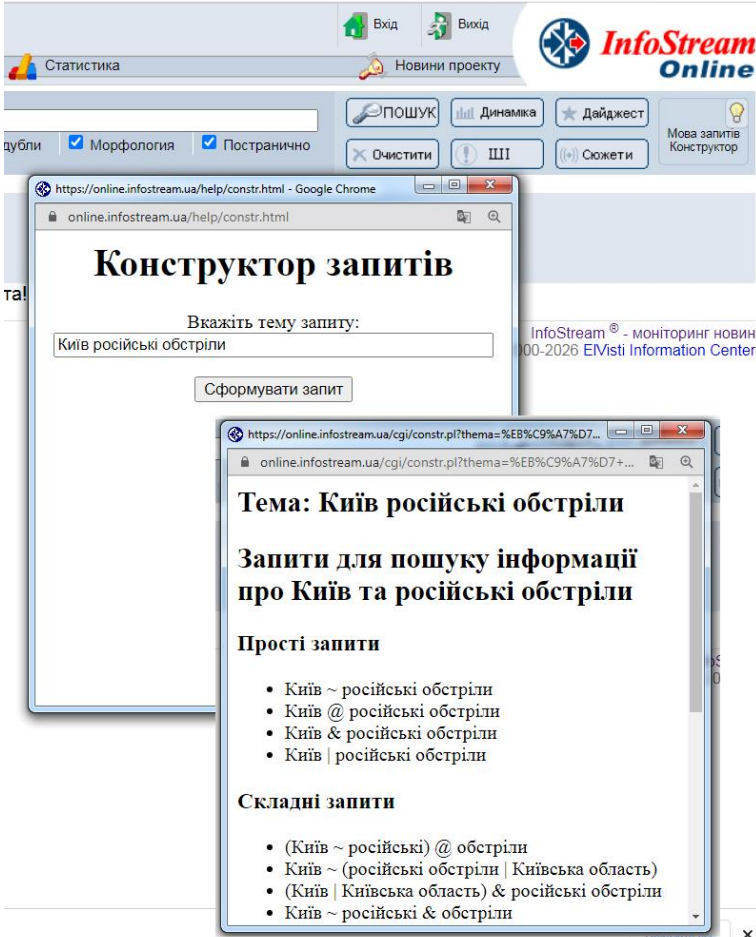
Окрім текстового поля введення, конструктор інтегрує візуальні фільтри метаданих, які трансформуються у внутрішні пошукові терми або застосовуються на етапі пост-фільтрації ядром InfoReS.

- Фільтр за джерелом: Реалізовано через каталог підключених веб-ресурсів з можливістю вибору окремих сайтів або груп. Синтаксично джерела ідентифікуються кодами джерел, наведених в розділі «Джерела» (наприклад, srd03031 для вебсайту «Zaxid.net», або srd02535 для джерела «24 канал»). Фільтр дозволяє обмежити вибірку джерелами, відібраними користувачем.
- Фільтр за країною джерела: Підтримується автоматичне розпізнавання мови документу та кодування країни джерела за стандартом ISO 3166-1 alpha-2. У запиті країна задається термом country.ua, country.de тощо.
- Фільтр за датою: Календарний інтерфейс дозволяє задавати довільні тимчасові інтервали для пошуку в оперативній та ретроспективній базах. Синтаксично дати подаються у форматі ГГГГ.ММ.ДД (наприклад, 2025.06.01). Підтримується праве скорочення: 2025.06.0 відбере документи з 1 по 9 червня 2025 року, а 2025.06 – всі публікації за червень. Конструктор автоматично конвертує візуальний вибір інтервалу у відповідні пошукові терми, що забезпечує узгодженість між UI та ядром.

### **3.3. Конструктор запитів**

Конструктор запитів у системі InfoStream є інтерактивним середовищем, що поєднує формалізований синтаксис пошукової системи InfoReS із візуальними інструментами багаторівневої фільтрації. Його методологічна

роль полягає у трансформації користувальницької інформаційної потреби у відтворювану, валідовану та оптимізовану пошукову конструкцію, яка стає основою для подальшої аналітичної обробки, автоматичного реферування та інтеграції в контур підтримки прийняття рішень. Конструктор реалізовано у вигляді структурованої форми введення з підтримкою логічних операторів, контекстних конструкцій, перемикачів морфології та дедуплікації, а також інтегрованих фільтрів за метаданими документів.



**Конструктор запитів**

Вкажіть тему запиту:  
Київ російські обстріли

Сформувати запит

**Тема: Київ російські обстріли**

**Запити для пошуку інформації про Київ та російські обстріли**

**Прості запити**

- Київ ~ російські обстріли
- Київ @ російські обстріли
- Київ & російські обстріли
- Київ | російські обстріли

**Складні запити**

- (Київ ~ російські) @ обстріли
- Київ ~ (російські обстріли | Київська область)
- (Київ | Київська область) & російські обстріли
- Київ ~ російські & обстріли

**Конструктор запитів:** Викликається із основного інтерфейсу за допомогою LLM. У відповідності з синтаксисом мови запитів системи InfoRes генеруються запити різної складності, які можна в подальшому застосовувати

Конструктор не обмежується одноразовим введенням тексту. Він є частиною циклу динамічного уточнення:

- Первинний пошук: Формування базової конструкції з урахуванням булевої логіки, контекстних операторів та фільтрів.
- Аналіз інформаційного портрета: Після отримання вибірки система генерує частотно-статистичну модель, де ранжуються рубрики, мови, джерела, географія, персоналії, тональність та характерні терміни.

Користувач може інтерактивно додавати або виключати елементи портрета, автоматично коригуючи запит без ручного редагування синтаксису.

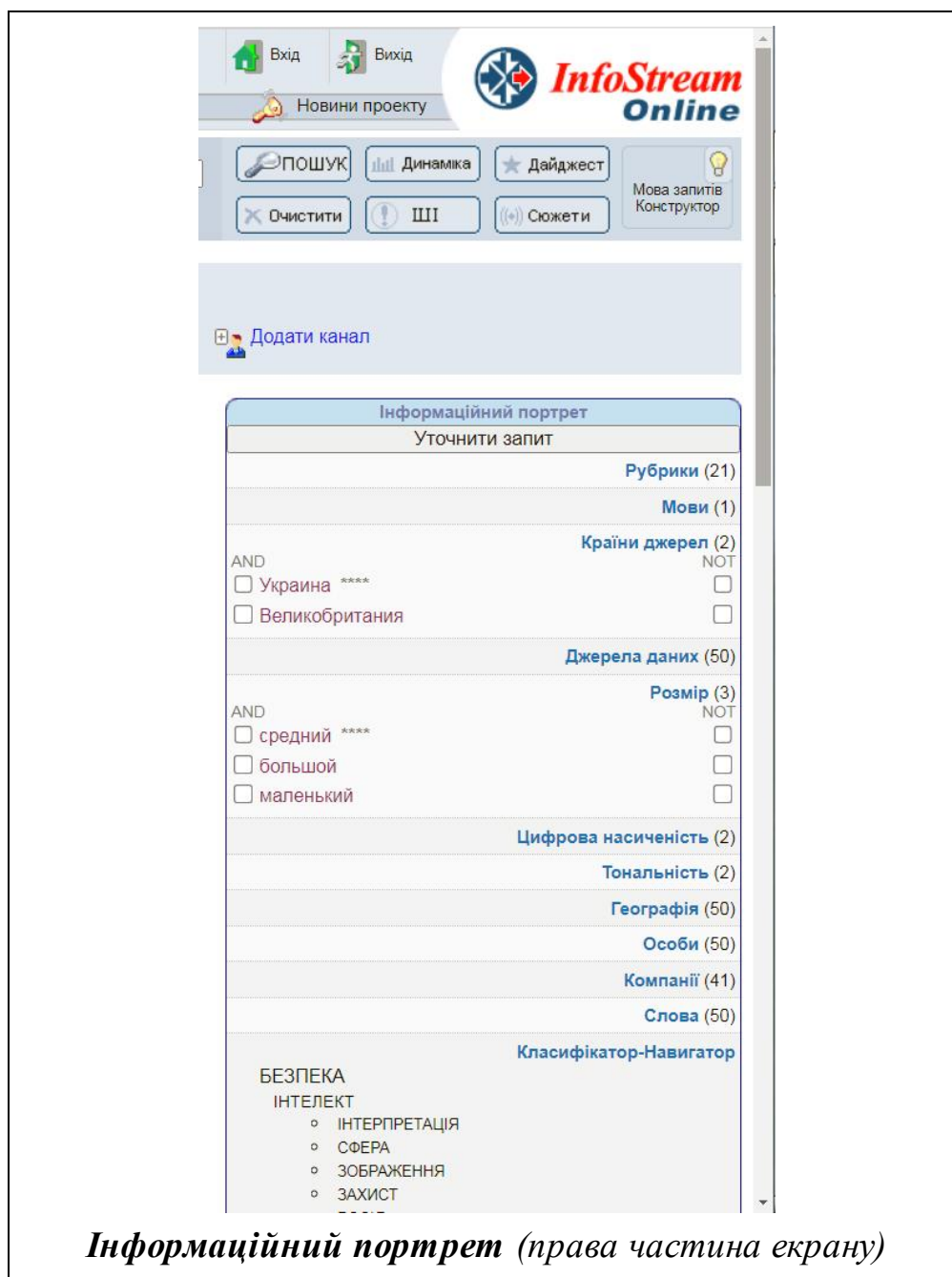
- Збереження та версіонування: Конфігурація запиту зберігається як «персональний інформаційний канал». Система фіксує усі параметри: текстову конструкцію, статус морфології, налаштування дедуплікації, діапазон дат, фільтри метаданих. Канал може бути переглянутий, відредагований, скопійований або видалений у «Кабинеті користувача».
- Автоматизація доставки: Збережений запит інтегрується у сервісні контури (email-розсилки, RSS-потоки, веб-віджети, API), що забезпечує безперервний моніторинг без повторного ручного конструювання.

Таким чином, конструктор запитів InfoStream виступає не лише інструментом введення синтаксису, але й методологічним каркасом формування контрольованої інформаційної вибірки. Поєднання багаторівневої булевої логіки, явного контролю пріоритетів через дужки, системних фільтрів метаданих та інтерактивного зворотного зв'язку через інформаційний портрет забезпечує перехід від довільного пошуку до відтворюваного, верифікованого та аналітично придатного каналу моніторингу.

### **3.4. Інформаційний портрет**

Інформаційний портрет є одним із основних аналітико-візуалізаційних інструментів системи InfoStream, який реалізує частотно-статистичну модель представлення результатів пошуку. На відміну від лінійного списку документів, портрет трансформує відібрану вибірку у динамічну структуру метаданих, де кожен атрибут (тематична рубрика, мова, джерело, географічна прив'язка, персоналії, тональність, рівень цифрової насиченості, обсяг тексту, характерні терміни) ранжується за його «вагою» – тобто частотою зустрічальності та статистичною значущістю в межах заданого інформаційного каналу. Таким чином, цей модуль забезпечує перехід від пасивного перегляду результатів до інтерактивного керування релевантністю вибірки.

Методологічна основа інформаційного портрета базується на емпірико-статистичних алгоритмах лінгвістичного аналізу. Система автоматично розраховує вагові коефіцієнти для кожного терміна та категоріального атрибута, нормалізує їх відносно загального обсягу вибірки та візуалізує у вигляді інтерактивного ранжованого списку. Найбільш репрезентативні характеристики позначаються маркерами високої ваги (символом «\*»), що дозволяє користувачу ідентифікувати ядро інформаційного потоку та відокремити його від «шуму» та змістовних дублів.



*Інформаційний портрет (права частина екрану)*

Структура інформаційного портрета містить такі ранжовані блоки:

- Тематичні рубрики та передвстановлені дескриптори;
- Мовна та географічна сегментація (країни джерел, згадані локації);
- Джерельний профіль (найменування ЗМІ, агенцій, офіційних порталів);
- Персоналії та іменовані сутності з текстів документів;
- Тональність документів (позитивна, нейтральна, негативна);
- Характерні терміни (слова з найвищою частотністю у вибірці);
- Технічні метрики: розмір повідомлень (малий/середній/великий), рівень цифрової насиченості (низький/середній/високий), що важливо для відбору аналітичних звітів, таблиць котировок або прес-релізів.

## ***Механізм динамічного уточнення вибірки***

Головна перевага інформаційного портрета полягає в його інтерактивності. Користувач може безпосередньо впливати на склад вибірки без ручного редагування синтаксису запиту:

- Одноетапне уточнення: клік на будь-яку характеристику автоматично доповнює пошуковий запит відповідним фільтром із урахуванням морфологічної обробки.
- Багатокритеріальне уточнення: використання логічних операторів «AND» (перетин множин) та «NOT» (виключення) безпосередньо в інтерфейсі портрета дозволяє конструювати складні фільтраційні сценарії.
- Конфігурування персонального каналу: після досягнення прийнятної релевантності параметри поточного портрета зберігаються як профіль персонального інформаційного каналу.

## ***Інтеграція з сучасними аналітичними процесами (LLM, RAG, часові ряди)***

У гібридній архітектурі InfoStream «Інформаційний портрет» виступає критичним контролем якості перед передачею даних у ШІ-модулі та статистичні моделі:

- Для RAG-архітектури – портрет забезпечує семантично чисту, верифіковану вибірку, мінімізуючи ризик подачі в контекстне вікно великої мовної моделі нерелевантних або маніпулятивних текстів. Метадані портрета використовуються для побудови фільтрів джерел та тематичних кластерів, що підвищує точність генерації звітів та гарантує прив'язку до першоджерел.
- Для LLM-аналітики частотно-статистичні ваги слів, персоналій та рубрик слугують підґрунтям для тонкого налаштування промптів, автоматичного екстрагування сутностей (NER) та оцінки тональності.
- Для аналізу часових рядів динаміка зміни вагових коефіцієнтів характеристик портрета у часі є базовим індикатором для побудови хвильових графіків, детекції інформаційних операцій, спалахів та прогнозування обсягу потоку. Портрет забезпечує темпоральну розмітку, необхідну для машинного навчання.

## ***Практичний сценарій та інтерфейс***

Інформаційний портрет інтегрований у екран результатів пошуку та перегляду персонального каналу. Він поділяється на логічні блоки, кожен з яких є інтерактивним. Типовий робочий цикл виглядає так:

- 1) аналітик формує базовий запит щодо регуляторних змін;
- 2) система видає вибірку;
- 3) портрет виявляє аномально високу вагу джерел з низькою редакційною перевіркою

- 4) аналітик виключає їх через «NOT» у портреті;
- 5) залишається верифіковане ядро, яке передається на генерацію RAG-звіту, автоматичне реферування та динамічний аналіз часових рядів.

У результаті користувач отримує не лише перелік документів, а й структуроване уявлення про склад, джерельну базу, тематичну структуру та інформаційну динаміку досліджуваного процесу, що суттєво підвищує якість аналітичних висновків і скорочує час підготовки управлінських рішень.

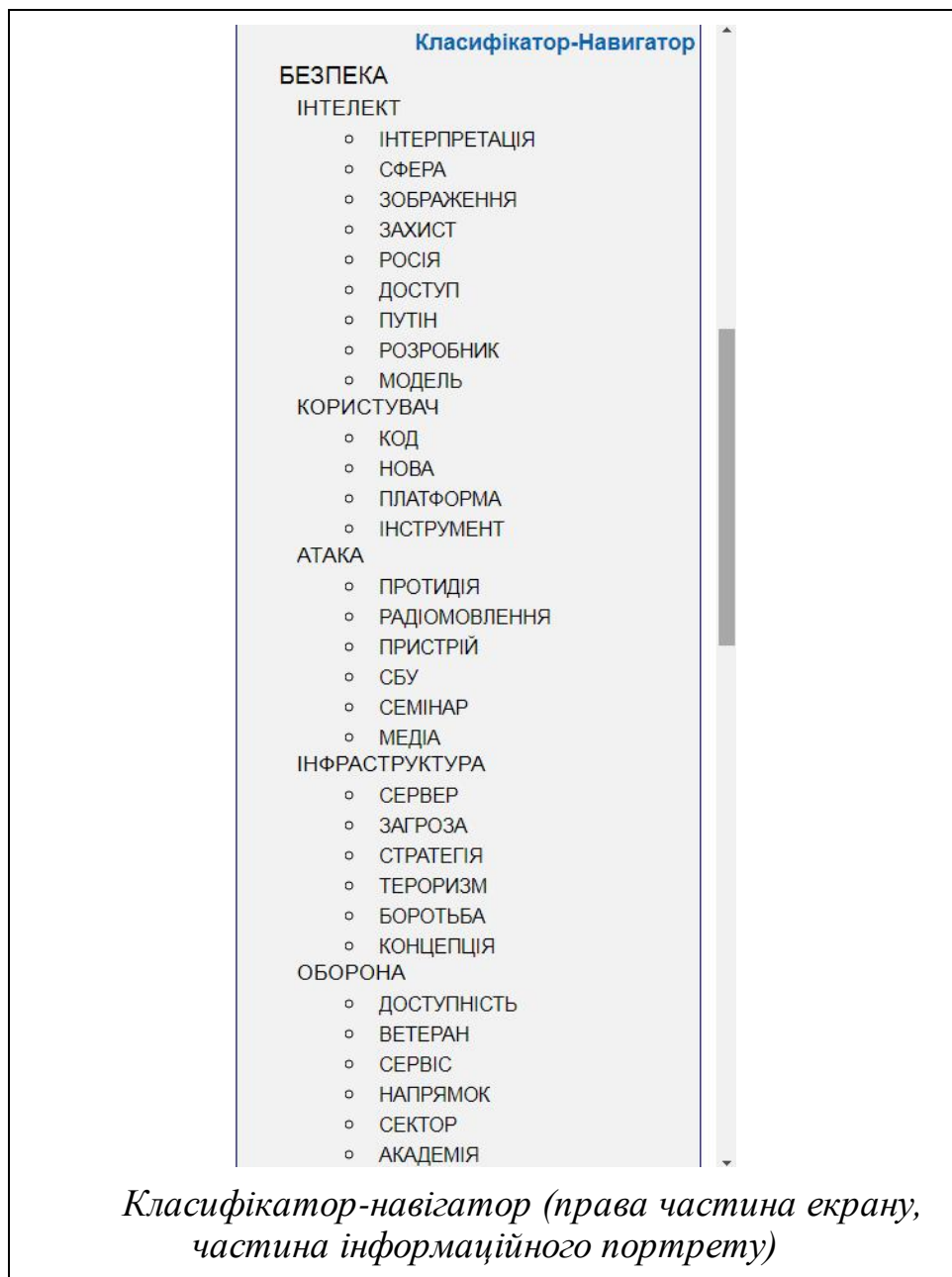
### ***Методологічне значення для СППР***

«Інформаційний портрет» поєднує класичну лінгво-статистичну строгість з інтерактивною гнучкістю. Він трансформує пошук із лінійного відбору у циклічний аналітичний процес: запит → вибірка → візуалізація портрета → динамічне уточнення → верифікований канал → подальша ІІІ-аналітика. Цей підхід гарантує, що кожен етап моніторингу підпорядкований критеріям релевантності, достовірності та придатності до аналітичної роботи.

### **3.5. Класифікатор-навігатор**

Класифікатор-навігатор InfoStream є інструментом динамічного уточнення результатів пошуку, який реалізує принцип інтерактивних «пошукових папок». На відміну від статичних рубрикаторів або фіксованих онтологій, що формуються заздалегідь і не враховують специфіку конкретного запиту, цей модуль генерує категорії безпосередньо на основі лексико-семантичного аналізу документів, відібраних системою на першому етапі. Такий підхід трансформує пошук з лінійної процедури відбору у багаторівневий аналітичний процес, де користувач послідовно звужує або розширює інформаційний канал, орієнтуючись на фактичну структуру знайденого контенту, а не на абстрактні схеми класифікації.

Основою цього механізму є автоматичне визначення близькості слів та термінів у масиві релевантних документів. Після обробки первинного запиту система будує дерево взаємопов'язаних кластерів, кожен із яких відображає окремий тематичний відтінок або контекстну площину пошуку. Наприклад, за запитом «бензин» класифікатор може автоматично сформувати гілки, пов'язані з «цінами на паливо», «екологічними нормами», «логістикою та поставками», «ринковими котировками» тощо. Кожна така гілка виступає як «пошукова папка» – динамічний фільтр, який миттєво звужує вибірку, додаючи до поточного запиту нові семантичні обмеження без необхідності ручного конструювання додаткових логічних операторів. Це дозволяє уникнути ситуацій, коли користувач не знає точних формулювань, не володіє повною термінологією предметної області або коли новинний дискурс містить неочікувані контекстні зсуви.



**Адаптивна кластеризація та алгоритмічна основа.** Механізм кластеризації базується на лінгво-статистичному аналізі частоти спільної появи термінів, оцінці контекстної близькості та виявленні латентних тематичних зв'язків у межах обраного масиву документів. Система не використовує фіксовані шаблони; замість цього вона адаптується до специфіки кожного запиту, мовного середовища та часового зрізу. Алгоритмічний контур враховує:

- аналіз спільної появи термінів – визначає, які слова найчастіше з'являються разом у релевантних документах, формуючи ядра кластерів;
- контекстну відстань – оцінює позиційну близькість слів у тексті;
- ієрархічну структуру – будує багаторівневе дерево, де кожен наступний рівень представляє більш деталізовану підтему або вузький

контекстний зріз. Глибина навігації регулюється користувачем, що забезпечує баланс між оглядовістю та деталізацією.

**Інтерактивний сценарій роботи.** Користувач формує первинний запит, після чого на екрані відображається результат пошуку та візуалізований класифікатор-навігатор. Аналітик може:

- оглядати автоматично згенеровані тематичні гілки та оцінювати їх розподіл за кількістю документів, мовою, країною походження або тональністю;
- активувати обрану «пошукову папку», що автоматично додає відповідні терміни або контекстні обмеження до поточного запиту;
- виключати нерелевантні кластери за допомогою оператора «НІ», не переписуючи синтаксичну конструкцію вручну;
- поглиблюватись у багаторівневу структуру, переходячи від загального контексту до вузьких підтем, або повертатись на попередній рівень для корекції напрямку аналізу.

Класифікатор-навігатор вирішує фундаментальну проблему інформаційного пошуку – розрив між формальним запитом і фактичною семантикою знайденого масиву.

У контексті архітектури InfoStream цей інструмент виступає містком між класичною булевою логікою та семантико-аналітичними модулями нового покоління. Принципи адаптивної кластеризації, реалізовані в класифікаторі-навігаторі, стали методологічною основою для подальшої інтеграції LLM-моделей тематичного групування.

### **3.6. Типові сценарії пошуку**

Типові сценарії пошуку в InfoStream відображають спектр практичних завдань, що виникають у процесі інформаційно-аналітичної, комунікаційної та управлінської роботи. Вони структуровані за принципом від загального до конкретного. Кожний сценарій базується на комбінації формалізованої мови запитів InfoReS та інструментів динамічного уточнення (інформаційний портрет, класифікатор-навігатор). У поточній системі ці класичні механізми інтегровані з LLM-реферуванням та аналізом часових рядів, що трансформує лінійний пошук у керований аналітичний процес, орієнтований на підтримку прийняття рішень.

#### ***Сценарій 1. Довгостроковий галузевий та ринковий моніторинг***

Мета: формування системного уявлення про стан галузі, регуляторні зміни, макроекономічні тенденції, структурні зрушення в інформаційному полі та циклічність медіауваги.

Технологічний підхід: використовується комбінація тематичних рубрик (наприклад, rubr02 для банківської сфери або rubr05 для енергетики), логічних операторів розширення семантики та фільтрів за мовою/країною.

Запит має широку предметну область, тому первинна вибірка потребує динамічного уточнення через «Інформаційний портрет».

Приклад запити: (банк | кредит | лізинг | депозит) & (регулюван | законодав | нбу | монетарн) ! (реклам | акці | промо | знижк)

Аналітичний контур: після формування вибірки активується ранжування джерел, географії, мов та тональності. Класифікатор-навігатор дозволяє інтерактивно кластеризувати результати за підтемами без повторного формулювання запити. На основі відібраного масиву генеруються автоматичні дайджести та сюжетні ланцюжки. У сучасній версії системи LLM виконує семантичне групування тематичних кластерів, а аналіз часових рядів виявляє сезонні патерни, довгострокові тренди та прогнозує обсяг інформаційного потоку. Ретроспективний фонд (понад 30 років) забезпечує відтворення історичних зрізів для порівняльного аналізу.

### ***Сценарій 2. Відстеження діяльності суб'єктів (конкуренти, партнери, бренди, персони)***

Мета: оцінка медіа-присутності, репутації, інформаційних активностей, реакції стейкхолдерів та виявлення потенційних ризиків або партнерських можливостей.

Технологічний підхід: вузькі, точні запити з використанням операторів точного входження (|), контекстної близькості та фільтрації за конкретними джерелами. Важливе відокремлення офіційних прес-релізів від експертних коментарів, аналітичних оглядів та новинного шуму.

Приклад запити: ("Національний банк України" | "НБУ" | нбу) & (рішен | ставка | інфляці | ліцензі | штраф) ! (конкурс | ваканс | тендер)

Аналітичний контур: система автоматично маркує змістовні дублі, ранжує тональність лексики та відокремлює первинні публікації від ретрансляцій. «Інформаційний портрет» дозволяє відфільтрувати низькоякісні або маргінальні джерела, зосередившись на медіа з високим індексом авторитетності. Система формує структуровані медіа-довідки з посиланням на першоджерело, дату публікації та уривок тексту, що усуває ризик «галюцинацій» ШІ та робить матеріал придатним для внутрішніх звітів, GR-комунікацій або експертних оцінок.

### ***Сценарій 3. Кризовий моніторинг та виявлення інформаційних ризиків***

Мета: оперативне виявлення загроз, інформаційних атак, техногенних/соціальних інцидентів або раптових регуляторних ініціатив для мінімізації ризиків та активації протоколів реагування.

Технологічний підхід: максимальна оперативність, використання спеціального режиму «Сюжети», комбінація «плюс-словників» (терміни, що сигналізують про кризу: аварія, конфлікт, санкція, відключення, розслідування, скандал) з географічними/галузевими фільтрами. Запити

налаштовуються на автоматичну доставку через email-розсилки, RSS-потоків або веб-віджети з періодичністю до режиму реального часу.

Приклад запиту: (авар | пожеж | відключен | затримк | штраф | конфіск) & (енергет | інфраструктур | логістик | фінанс) & country.ua

Аналітичний контур: режим «Сюжети» фільтрує публікації за критеріями: мінімальний час життя документа, лексична близькість до локального інформаційного потоку, суттєва відмінність від довгострокового фону, наявність сигналних термінів, високий ранг джерела та відсутність змістовного дублювання. Хвильовий аналіз часових рядів детектує аномальні спалахи активності, відокремлюючи органічний розвиток події від штучного нагнітання. LLM-модулі автоматично генерують попередні резюме інциденту для швидкого інформування керівництва, а RAG-архітектура гарантує, що кожен висновок прив'язаний до верифікованого першоджерела.

#### ***Сценарій 4. Відстеження конкретних подій, кампаній та законодавчих ініціатив***

Мета: моніторинг життєвого циклу окремої події (наприклад, виходу продукту, ухвалення закону, проведення міжнародного форуму, PR-кампанії) від етапу анонсу до пост-ефектів та суспільного резонансу. Технологічний підхід: тимчасові обмеження (інтерфейсний календар або датові терми у форматі РРРР.ММ.ДД), оператори контекстного слідування (~) для фіксації точних формулювань, фокусування на джерелах-інсайдерах та офіційних каналах комунікації.

Приклад запиту: (закон~цифров | законопроект~діджитал | законопроект~цифр) & (парламент | комітет | голосуван | читанн) & 2026.05.

Аналітичний контур: «Динамика понять» будує гістограму згадуваності ключових термінів у часовому розрізі, дозволяючи візуалізувати фази життєвого циклу події: анонс → суспільне обговорення → ухвалення → імплементація → оцінка наслідків. Сюжетні ланцюжки автоматично групують публікації за смисловими кластерами, а LLM-звіт формує хронологічну аналітичну довідку з посиланнями на основні рішення та реакції ЗМІ, а семантичний прогноз моделює ймовірні напрямки подальшого розвитку теми, що дозволяє планувати комунікаційні або управлінські кроки на випередження.

Користувач має можливість комбінувати сценарії відповідно до змін інформаційного контексту. Усі описані сценарії демонструють перехід від лінійного пошуку до багаторівневої аналітики. Класичні інструменти забезпечують контрольовану, відтворювану та семантично валідну вибірку. Сучасні аналітичні модулі трансформують відібрані дані у структуровані інсайти, придатні для інтеграції в контур систем підтримки прийняття рішень.

# РОЗДІЛ 4. СУЧАСНА АНАЛІТИКА: LLM, RAG, ЧАСОВІ РЯДИ ТА ПРОГНОЗУВАННЯ

## БЛОК-СХЕМА ВИКЛАДУ МАТЕРІАЛУ



Якщо класичний інструментарій мови запитів, конструктора пошуку та інформаційних портретів забезпечує контрольовану, відтворювану та семантично валідну вибірку, то сучасні аналітичні модулі трансформують цей масив даних у структуровані інсайти, придатні для безпосереднього використання в контурі систем підтримки прийняття рішень. Архітектура InfoStream доповнена гібридним аналітичним контуром, що поєднує перевірені алгоритми текстової аналітики з можливостями великих мовних моделей, архітектурою RAG та сучасними методами аналізу часових рядів.

У цьому розділі послідовно розкриваються три взаємопов'язані технологічні напрями, що формують основу сучасної новинної аналітики. По-перше, інтеграція LLM розглядається не як автономна генерація текстів, а як структурований процес обробки. Важливо підкреслити, що мовні моделі працюють виключно з релевантною вибіркою, відібраною засобами класичного ядра InfoReS, що гарантує стабільність результатів та мінімізує вплив інформаційного шуму. По-друге, архітектура RAG (Retrieval-Augmented Generation) представлена як методологічний стандарт верифікованості Ш-аналітики: кожен згенерований висновок, резюме або дайджест прив'язується до першоджерела. По-третє, аналіз часових рядів трансформує класичні гістограми динаміки понять у інтерактивні інструменти виявлення хвильових патернів, аномальних інформаційних спалахів, циклічності медіауваги та латентних трендів.

Окрему увагу приділено двокомпонентному прогнозуванню інформаційних процесів. Динамічний прогноз базується на статистичних та машинних моделях, що оцінюють ймовірність зміни обсягу потоку, рівня медіа-резонансу або виникнення інформаційних імпульсів. Семантичний прогноз, навпаки, використовує контекстуальні можливості LLM для моделювання ймовірних напрямків розвитку дискурсу, очікуваних реакцій стейкхолдерів та точок кристалізації нових подій. Поєднання цих підходів переводить моніторинг із режиму реактивного спостереження у превентивний аналітичний контур.

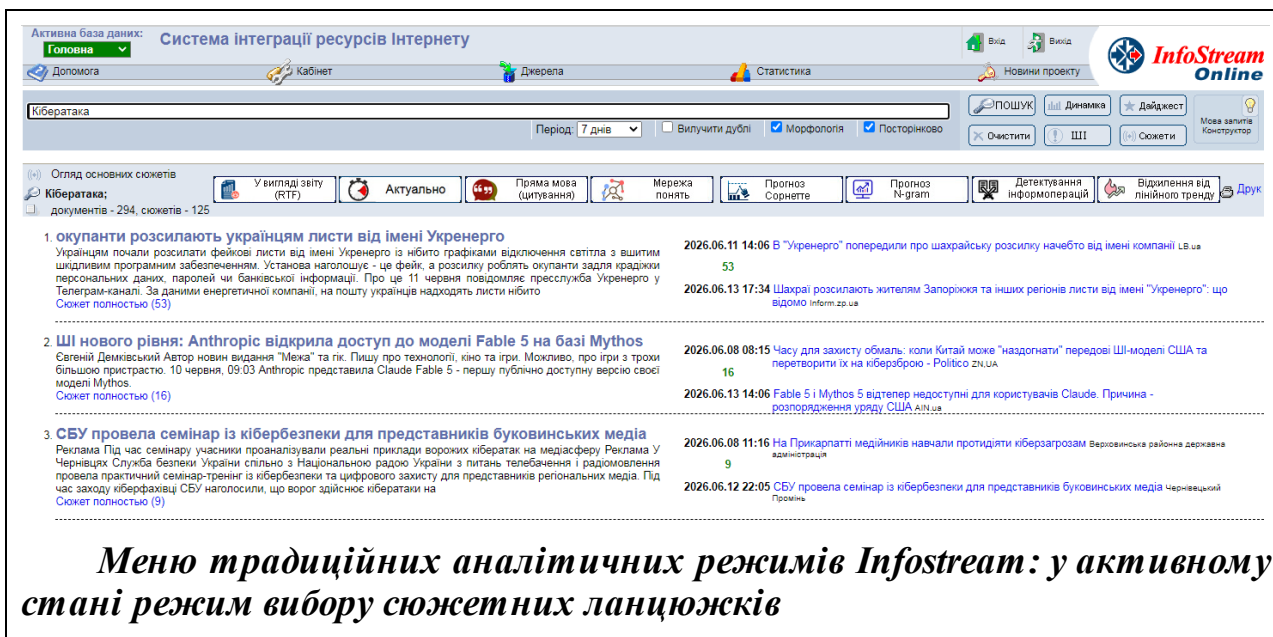
#### **4.1. Автоматичне реферування та генерація структурованих резюме сюжетів**

Автоматичне реферування масивів новинних документів є одним із головних етапів трансформації сирового інформаційного потоку у придатні для прийняття рішень аналітичні інсайти.

У системі InfoStream ця функція реалізується через гібридний підхід, що поєднує перевірені лінгво-статистичні алгоритми Text Mining із сучасними можливостями великих мовних моделей та архітектурою RAG. Такий симбіоз забезпечує високу швидкість обробки, семантичну точність та обов'язкову верифікацію кожного згенерованого твердження, що є критичним для інтеграції результатів у контур систем підтримки прийняття рішень.

## Частотно-статистичне реферування та формування дайджестів

Історично ядро автоматичного реферування в InfoStream базувалося на емпірико-статистичних методах, закладених в архітектуру інформаційно-пошукової системи InfoReS. Алгоритм визначає вагові коефіцієнти окремих термінів, речень, абзаців та цілих документів у межах релевантної вибірки. На основі цих ваг система відбирає найбільш інформативно насичені фрагменти, формує структурований дайджест, автоматично виключає змістовні дублі та прив'язує кожен фрагмент до дати публікації та гіперпосилання на першоджерело. Цей підхід гарантує відтворюваність, прозорість логіки відбору та мінімізацію суб'єктивного впливу, залишаючись технологічним стандартом для первинного структурного стиснення новинного потоку.



The screenshot displays the InfoStream Online web interface. At the top, there is a navigation bar with 'Активна база даних: Готова' and various menu items like 'Допомога', 'Кабінет', 'Джерела', 'Статистика', and 'Новини проекту'. A search bar contains the text 'Кибератака'. Below the search bar, there are filters for 'Період: 7 днів', 'Вилучити дублі', 'Морфологія', and 'Посторинково'. The main content area shows a list of search results with titles, dates, and snippets. The results are:

- 1 окупанти розсилають українцям листи від імені Укрнерго**  
Українцям почали розсилати фейкові листи від імені Укрнерго із нібито графіками відключення світла з шкідливим шкідливим програмним забезпеченням. Установа наголошує - це фейк, а розсилку роблять окупанти задля крадіжки персональних даних, паролів чи банківської інформації. Про це 11 червня повідомляє пресслужба Укрнерго у Телеграм-каналі. За даними енергетичної компанії, на пошту українців надходять листи нібито  
Сюжет повністю (53)  
2026.06.11 14:06 Б "Укрнерго" попередили про шахрайську розсилку начебто від імені компанії LB.ua  
53  
2026.06.13 17:34 Шахраї розсилають жителям Запоріжжя та інших регіонів листи від імені "Укрнерго", що відомо inform.zp.ua
- 2 ШІ нового рівня: Anthropic відкрила доступ до моделі Fable 5 на базі Mythos**  
Севеній Деміаський Автор новин видання "Межа" та гік. Пишу про технології, кіно та ігри. Можливо, про ігри з трохи більшою пристрасно. 10 червня, 09:03 Anthropic представила Claude Fable 5 - першу публічно доступну версію своєї моделі Mythos.  
Сюжет повністю (16)  
2026.06.08 08:15 Часу для захисту обмаль: коли Китай може "наздогнати" передові ШІ-моделі США та перетворити їх на кіберзброю - Politico.zn.ua  
16  
2026.06.13 14:06 Fable 5 і Mythos 5 відтепер недоступні для користувачів Claude. Причина - розпорядження уряду США a1n.ua
- 3 СБУ провела семінар із кібербезпеки для представників буковинських медіа**  
Резюме Під час семінару учасники проаналізували реальні приклади ворожих кібератак на медіасферу Републіки у Чернівцях Служба безпеки України спільно з Національною радою України з питань телебачення і радіомовлення провела практичний семінар-тренінг із кібербезпеки та цифрового захисту для представників регіональних медіа. Під час заходу кіберфахівці СБУ наголосили, що ворог здійснює кібератаки на  
Сюжет повністю (9)  
2026.06.08 11:16 На Прикарпатті медійників навчали протидіяти кіберзагрозам Верховинська районна державна адміністрація  
9  
2026.06.12 22:05 СБУ провела семінар із кібербезпеки для представників буковинських медіа Чернівцької Промінь

## Меню традиційних аналітичних режимів Infostream: у активному стані режим вибору сюжетних ланцюжків

### Інтеграція з LLM

Сучасний аналітичний контур InfoStream розширює класичне реферування шляхом інтеграції великих мовних моделей. LLM не замінюють пошукове ядро, а працюють виключно з попередньо відібраною, фільтрованою та семантично валідною вибіркою. При цьому реалізується:

- об'єднання розрізнених повідомлень чіткою хронологією, причинно-наслідковими зв'язками та виділенням основних подій;
- витягнення сутностей та контексту – автоматичне маркування персоналій, організацій, локацій, нормативних актів із збереженням семантичних зв'язків між ними;
- оцінка емоційного забарвлення та інформаційної ваги фрагментів, що дозволяє відокремити фактологічне ядро від коментарів, спекуляцій або маніпулятивних конструкцій;
- структурована генерація резюме – формування уніфікованих звітів за попередньо визначеними шаблонами.

Активна база даних: Система інтеграції ресурсів Інтернету

Головна | Допомога | Кабінет | Джерела | Статистика | Новини проекту

Кибератака

Період: 7 днів | Вилучити дублі | Морфологія | Посторинково

Пошук | Динаміка | Дайджест | Очистити | ШІ | Сюжети

Мова запитів: Конструктор

Масив текстів для аналізу (13 июня 2026)

Кибератака

**У режимі Аналітика штучного інтелекту:**  
 Формується репрезентативна повнотекстова вибірка документів за запитом  
 Робиться резюме за змістом цих документів  
 Формуються сценарії вирішення проблем, пов'язаних із запитом  
 Робиться прогноз (форсайт) розвитку проблем, пов'язаних із запитом  
 Формується інтерактивна семантична карта понять, пов'язаних із запитом

Наверх

**Аналітичні режими ШІ в поточній версії InfoStream**

## Архітектура RAG як методологічний стандарт верифікації

Критичною перевагою сучасного модуля реферування є інтеграція архітектури Retrieval-Augmented Generation (RAG). Замість вільної генерації текстів, система працює за принципом «спочатку пошук – потім синтез – завжди посилання». Кожен абзац або висновок у згенерованому резюме супроводжується цитуванням джерела, дати публікації та оригінального уривка тексту. Це усуває ризик «галюцинацій» ШІ, робить результати придатними для офіційних довідок, експертних висновків та аудиту.

Активна база даних: Система інтеграції ресурсів Інтернету

Головна | Допомога | Кабінет | Джерела | Статистика | Новини проекту

київ обстріл росія

Період: 7 днів | Вилучити дублі | Морфологія | Посторинково

Пошук | Динаміка | Дайджест | Очистити | ШІ | Сюжети

Мова запитів: Конструктор

Дайджест електронної преси (13 июня 2026)

київ обстріл росія

1. Генерал-лейтенант Крістофер Коутс: Смерть Путіна. Це - перша необхідна умова можливого тривалого миру
2. Битва за Україну. День тисяча п'ятсот шістдесят дев'ятий
3. Зеленський закликав реагувати на повітряні тривоги в найближчі дні
4. Не такого чекали.... Олена Тополя здивувала виплатою від держави після "прильоту" в квартири
5. Поки Кличко не звітує: на що КМДА законтракувала майже 26 млрд грн з початку року
6. путін погрожує наростити удари по Україні: кого торкнеться новий виток ескалації
7. Кількість жертв від атаки на Київ 2 червня зростає: померла студентка
8. Триває підготовка на російському полігоні "Капустин Яр": Військові попередили про загрозу запуску "Орешника" протягом доби
9. Вільний від "правильності". Пам'яті журналіста і військового тернополянина Віталія Дереха
10. Мистецький зв'язок між "Оригами-оленом" і Будапештським меморандумом? - головна художниця Українського павільйону на Венеційській бієнале

1. **Генерал-лейтенант Крістофер Коутс: Смерть Путіна. Це - перша необхідна умова можливого тривалого миру**  
 Сурма-UA 2026.06.13.10.18  
 Що, на вашу думку, західна аудиторія досі не розуміє про нинішній стан війни?  
 - Я впевнений в Україні, але не маю жодних ілюзій щодо масштабів викликів.  
 Водночас я не переконаний, що решта світу ставиться до цієї проблеми так само серйозно.  
 Росія не воює за власне виживання.  
 Це має наслідки не лише для армій, а й для всієї оборонно-промислової системи Заходу.

2. **Битва за Україну. День тисяча п'ятсот шістдесят дев'ятий**  
 Українформ 2026.06.11.21.01  
 Посли Великої Британії, Франції та Німеччини передали в МЗС РФ умови мирного врегулювання війни в Україні Ми збрали для вас інформацію про тисяча п'ятсот шістдесят дев'ятий день Великої війни: чим жили Україна та українці 11 червня 2026 року.  
 У Криму після атаки дронів пошкоджені два мости в Армянську, а також сталася пожежа в Севастополі в районі військової частини у Стрілецькій бухті - Телеграм-канал "Кримський ветер".  
 Міністр закордонних справ Андрій Сибіга обговорив з міністром у справах Європи та закордонних справ Албанії Ферітом Ходжею спільний шлях до вступу в ЄС; обговорив з очільницею МЗС Румунії Оаною Цою транскордонну безпеку в умовах війни з Росією.  
 США ввели санкції за постачання зброї для КВІР - у списку особи й компанії з Білорусі та Китаю - Томмі Пігготт, речник Держделу.  
 Польща підтримує інтеграцію України в Євросоюз та не має жодних застережень до цього процесу - Пйотр Лукасевич, тимчасовий повірений у справах Польщі в Україні.

**Тематичний дайджест, сформований за класичним алгоритмом**

-	Назва пункту	Зміст
1	Про що, події	Події війни в Україні, наслідки російських атак, міжнародна підтримка України.
2	Де це було	Україна, Росія, Київ, Крим, Севастополь, Албанія, Румунія, Польща.
3	Коли	11-12 червня 2026 року.
4	Хто	Володимир Зеленський, генерал-лейтенант Крістофер Коутс, Андрій Сибіга, Феріт Ходжею, Оана Цоя, Томмі Пігготт, Пйотр Лукасевич, Олена Тополя, Тарас Тополя, путін.
5	Перспективи розвитку	Можливий тривалий мир за умови зміни влади в Росії, інтеграція України в ЄС.
6	Переказ тексту (розширений)	Генерал-лейтенант Крістофер Коутс вважає, що смерть Путіна є необхідною умовою для тривалого миру. Він також підкреслив, що західна аудиторія досі не розуміє масштабів викликів, пов'язаних із війною в Україні. Міжнародна підтримка України продовжується: послы Великої Британії, Франції та Німеччини передали в МЗС РФ умови мирного врегулювання війни в Україні. Український президент Володимир Зеленський закликав реагувати на повітряні тривоги в найближчі дні. Росія продовжує здійснювати атаки на Україну, внаслідок чого є жертви та пошкодження інфраструктури.
7	Джерела	Сурма-UA, Укрінформ, Ракурс, Патріоти України, DoZorro, Стіна, Регион Киев Медиа, TeNews.org.ua, Стрічка новин Запоріжжя.

*Резюме, сформоване з використанням LLM і RAG-технології*

Поєднання класичного частотно-статистичного реферування, LLM-синтезу та RAG-верифікації перетворює автоматичне формування резюме з технічної утиліти в науково обґрунтований аналітичний інструмент. Система зберігає повну відтворюваність результатів, прозорість логіки відбору та прив'язку до першоджерел, що відповідає сучасним вимогам до інформаційної гігієни, аудиту та етичного використання ШІ в дослідницьких та управлінських процесах.

#### 4.2. RAG-архітектура в InfoStream

Архітектура Retrieval-Augmented Generation у системі InfoStream виступає методологічним мостом між класичним повнотекстовим пошуком та сучасними можливостями великих мовних моделей. На відміну від ізольованої генерації текстів, коли модель спирається виключно на свої внутрішні параметри, засоби RAG примусово «заземлюють» процес синтезу на реальних джерелах. У контексті новинного моніторингу це означає перехід від абстрактних відповідей до аналітичних довідок, де кожен висновок, дата, цифра чи оцінка тональності має пряме підтвердження у першоджерелі. Такий підхід трансформує ШІ з інструменту творчої генерації в інструмент верифікованої семантичної обробки, що є критичним для інтеграції в контур систем підтримки прийняття рішень.

## ***Архітектурний контур RAG у InfoStream***

Retrieval-Augmented Generation реалізовано у три послідовні етапи, кожен з яких інтегрований у існуючу технологічну структуру системи:

1. Retrieval (Відбір контексту): На вхід модуля подається результат роботи класичного пошукового ядра InfoReS. Користувач формує запит мовою системи, яка повертає ранжований масив документів із повними текстами та розширеними метаданими. Ретроспективний фонд забезпечує глибину контексту, а механізми автоматичного виявлення змістовних дублів гарантують відсутність інформаційного шуму.
2. Augmentation (Підготовка контексту): Відібрані документи нормалізуються та структуруються для подання LLM. Кожен фрагмент супроводжується обов'язковим набором метаданих. Контекст формується у вигляді структурованого промпту з чіткими інструкціями щодо пріоритетів, хронології та обмежень на генерацію.
3. Generation (Синтез звіту): Велика мовна модель обробляє підготовлений контекст і генерує структурований аналітичний текст: резюме сюжету, хронологію подій, оцінку медіа-резонансу, витягнення основних рішень чи нормативних актів, або відповідь на конкретне аналітичне запитання.

### ***Контекстуалізація пошуку***

У класичному моніторингу пошук завершується видачею списку документів. У RAG-контурі InfoStream пошук стає першою ланкою аналітичного процесу. Контекстуалізація досягається завдяки поєднанню формалізованої мови запитів та семантичного групування. На вхід LLM надходить не хаотичний потік, а тематично однорідний, хронологічно впорядкований та метаданими збагачений корпус. Це дозволяє моделі зберігати причинно-наслідкові зв'язки, розрізняти первинні повідомлення та коментарі, а також відстежувати еволюцію дискурсу в межах заданого тимчасового вікна.

### ***Верифікація джерел та обов'язкове цитування***

Головною вимогою архітектури є принцип «жодного твердження без посилання». Кожен абзац, факт або аналітичний маркер у згенерованому звіті супроводжується гіперпосиланням на конкретний документ у базі, дату публікації та уривок оригінального тексту. Це перетворює ШІ-звіт на аудитовану аналітичну довідку, придатну для офіційного використання, експертних висновків та наукових публікацій. Завдяки збереженню повних текстів у ретроспективному фонді, навіть якщо оригінальна публікація видалена з мережі, її зміст та контекст залишаються доступними для валідації.

Система інтеграції інтернет-ресурсів

Допомога Кабінет Джерела Статистика Новини проекту

Вхід Вихід

InfoStream Online

### Citation by Query: Іран

N	Who	Citate	Source	Date-Time	Url
1	Головування Кіпру зазначило:	"Цей важливий крок підтверджує, що Європейський Союз найсильніший, коли є єдиним і принциповим, відкритим для тих, хто відданий його цінностям"	1News	2026.06.13 00:00	<a href="https://1news.com.ua/s...">https://1news.com.ua/s...</a>
2	Столе Сольбаккен	"Ми всі можемо погодитися, що багато чого могло б бути інакше, але ми всі ліцемери. Ви ліцемери, ми ліцемери. Можна було б зробити по-іншому. Зараз тут проходить чемпіонат світу, і ми тут, щоб грати в футбол. Є багато інших речей, про які ми могли б поговорити, і є багато інших речей, про які я хотів би поговорити. Зараз виходить так, що в ряді ситуацій можна було вчинити інакше"	Футбол Газета	2026.06.13 00:39	<a href="http://footballgazeta...">http://footballgazeta....</a>
3	Аббас Арагчі	"Іран є переможцем війни зі США"	РБК-Україна	2026.06.13 01:45	<a href="https://www.rbc.ua/rus...">https://www.rbc.ua/rus...</a>
4	Пряма мова Арагчі:	"Іран є переможцем у війні зі США"	Українська правда	2026.06.13 03:39	<a href="https://www.pravda.com...">https://www.pravda.com...</a>
5	Пряма мова Арагчі:	"Наш меч завжди висітиме над Ормузькою протокою"	Українська правда	2026.06.13 03:39	<a href="https://www.pravda.com...">https://www.pravda.com...</a>
6	Пряма мова Арагчі:	"Іран є переможцем у війні зі США"	1News	2026.06.13 04:00	<a href="https://1news.com.ua/u...">https://1news.com.ua/u...</a>

**Аналітичний режим «Пряма мова» з верифікацією висновків через посилання на першоджерела**

### Мінімізація «галюцинацій» III

Проблема генерації неперевірених даних вирішується на кількох рівнях:

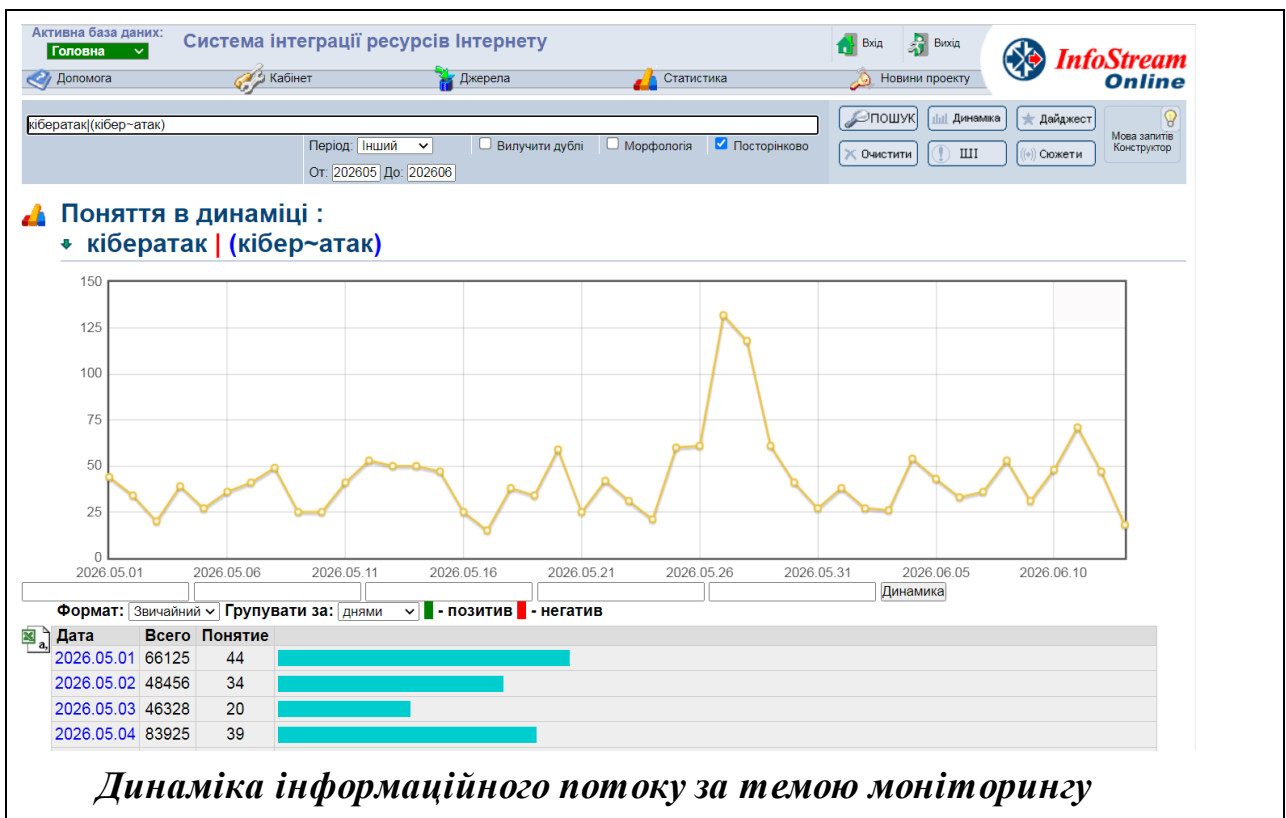
- Обмеження контексту: LLM отримує інструкцію генерувати відповідь виключно на основі наданих документів. Твердження, що виходять за межі вибірки, маркуються як «не підтверджено джерелами».
- Примусова атрибуція: Алгоритм блокує вивід фактів без прямого посилання на індексований документ. Кожна цитата або числовий показник супроводжується ідентифікатором джерела та часом публікації.
- Оцінка впевненості та маркування невизначеності: Якщо модель не знаходить достатнього підтвердження в контексті, замість вгадування відповіді вона генерує позначку «дані відсутні в доступному корпусі» або «потрібне додаткове уточнення запиту».

Поєднання класичної точності InfoReS, структурованого ретроспективного фонду та RAG-верифікації перетворює новинний моніторинг з інструменту спостереження на систему керованого семантичного аналізу з гарантованою відстежуваністю кожного кроку процесу. Це є фундаментальною умовою для переходу до прогнозного

моделювання та інтеграції аналітичних інсайтів у контур прийняття рішень, що розкривається в наступних розділах.

### 4.3. Аналіз часових рядів

Аналіз часових рядів у контексті контент-моніторингу новинних потоків є фундаментальним інструментом переходу від статичного опису подій до динамічного моделювання інформаційних процесів. Історично базовою формою візуалізації та первинного аналізу в системі InfoStream виступала гістограма «Динамика понять», що будувалася як результат пошуку за множиною запитів із фіксацією кількості публікацій за заданими тимчасовими інтервалами. Кожен стовпчик гістограми відповідав певній даті або періоду, був гіперпосиланням на відповідну вибірку документів та дозволяв оперативно відстежувати зміну згадуваності ключових термінів, персоналій чи подій. Цей класичний підхід, заснований на частотному підрахунку та темпоральній індексації, залишається незмінною основою для візуальної орієнтації аналітика. Однак сучасні вимоги до підтримки прийняття рішень вимагають виходу за межі описової статистики та переходу до аналітичного моделювання часових рядів, що включає виявлення хвильових патернів, детекцію аномалій та декомпозицію циклічності.



#### *Еволюція від описових гістограм до моделювання потоків*

Сучасний аналітичний контур InfoStream трансформує сиру послідовність дат публікацій у структурований часовий ряд шляхом

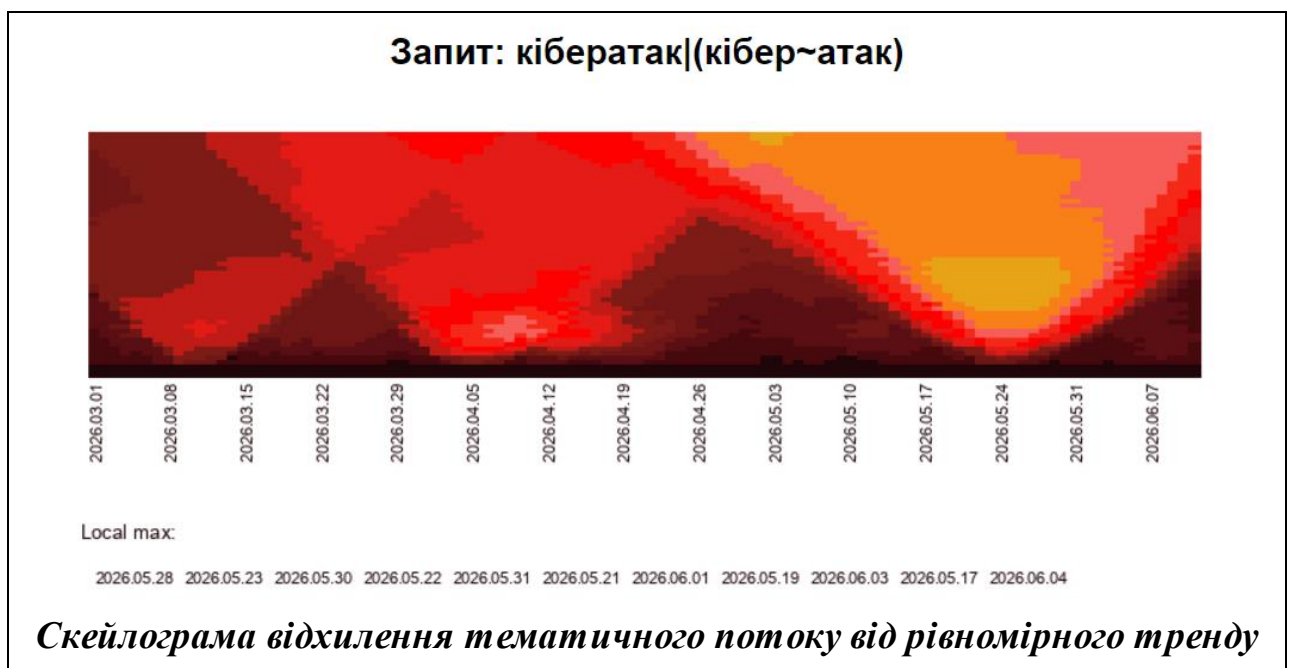
автоматичної нормалізації: усунення сезонних зсувів джерел, згладжування нерівномірностей сканування, калібрування за обсягом щоденного трафіку та нормування частотності понять до індексу медіа-інтенсивності. На цьому підґрунті застосовуються сучасні статистичні та машинні методи, що дозволяють виявляти латентні структури новинного потоку, недоступні при візуальному огляді гістограм. Кожна точка ряду супроводжується метаданими: кількість унікальних джерел, рівень цифрової насиченості, доменна приналежність, оцінка тональності, що забезпечує багатовимірність аналізу.

### ***Виявлення хвильових патернів інформаційних процесів***

Новинні потоки підпорядковуються хвильовій динаміці, де кожна інформаційна подія проходить життєвий цикл: імпульс (поява первинного повідомлення) → експоненційне зростання (масове тиражування, коментарі, аналітика) → пік резонансу → спад (завершення циклу або перехід у латентну фазу). Система автоматично ідентифікує ці фази за допомогою алгоритмів сегментації часових рядів та кластерного аналізу за темпоральною близькістю. Хвильові патерни дозволяють:

- відрізнити ізольовані події від резонансних медіа-кампаній;
- відстежувати ехо-ефекти та вторинні хвилі обговорення;
- прогнозувати момент завершення активного циклу для коригування комунікаційних стратегій;
- інтегрувати хронологію подій у сюжетні ланцюжки, де кожен сегмент відповідає окремій хвилі медіауваги.

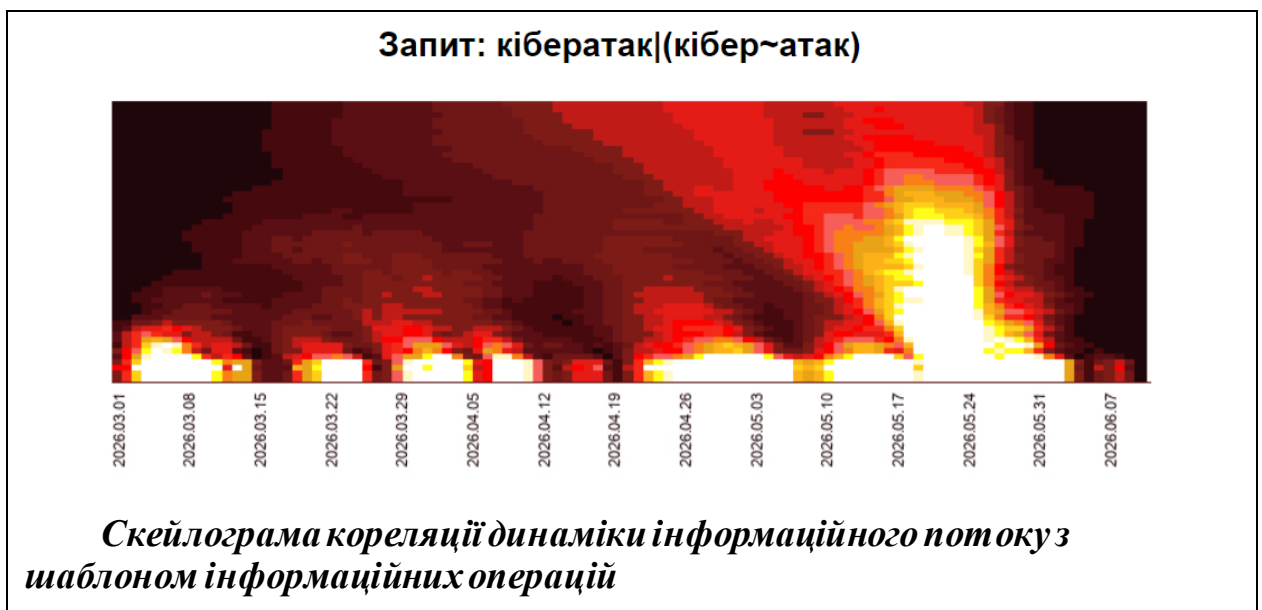
У сучасній версії системи хвильовий аналіз візуалізується на інтерактивних дашбордах.



Багаторічний ретроспективний фонд InfoStream створює унікальну емпіричну базу для виявлення довгострокових циклів. Система розкладає часові ряди на компоненти: тренд (довгостроковий напрям), сезонність (річні/квартальні патерни), циклічність (нерегулярні періодичні коливання) та шум.

Аналіз часових рядів не існує ізольовано; він є критичним компонентом гібридного аналітичного процесу:

- Зв'язок з LLM: темпоральні метадані та вектори зміни частотності понять подаються на вхід мовних моделей для семантичного прогнозування розвитку дискурсу та генерації сценаріїв подальшого реагування.
- Зв'язок з RAG: кожен графік, хвильовий сегмент або виявлена аномалія прив'язується до конкретних документів, що забезпечує відтворюваність та можливість глибокої перевірки першоджерел.
- Зв'язок з СППР: результати аналізу експортуються в корпоративні середовища, інтегруються в автоматизовані звіти та використовуються для налаштування правил оповіщення.



Інтерфейс системи надає аналітику інструменти для ручної корекції моделей, вибору горизонту прогнозу, накладання подієвих маркерів та збереження конфігурацій як шаблонів для майбутніх досліджень. Це забезпечує масштабованість аналітичних процедур та їхню адаптацію до специфіки предметної області.

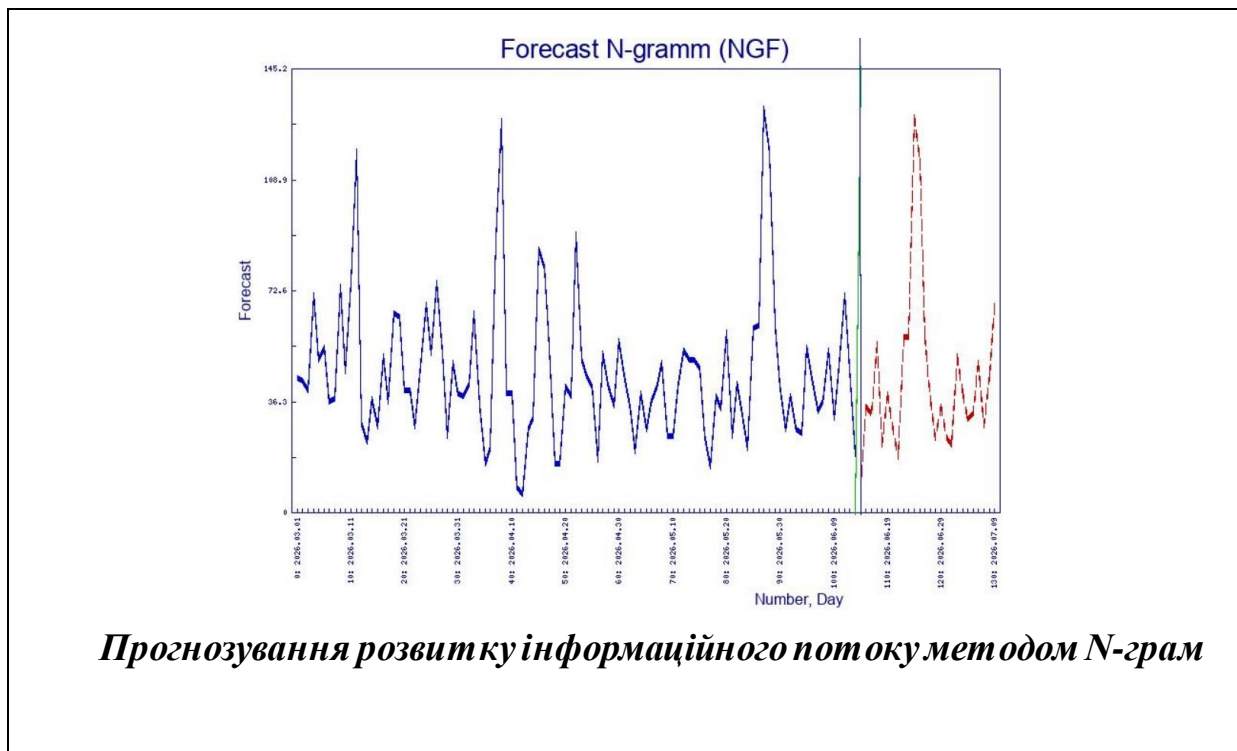
#### **4.4. Прогнозування інформаційних процесів**

Прогнозування інформаційних процесів у новинному середовищі є логічним завершенням аналітичного контуру моніторингу. Якщо попередні етапи (пошук, фільтрація, реферування, аналіз часових рядів)

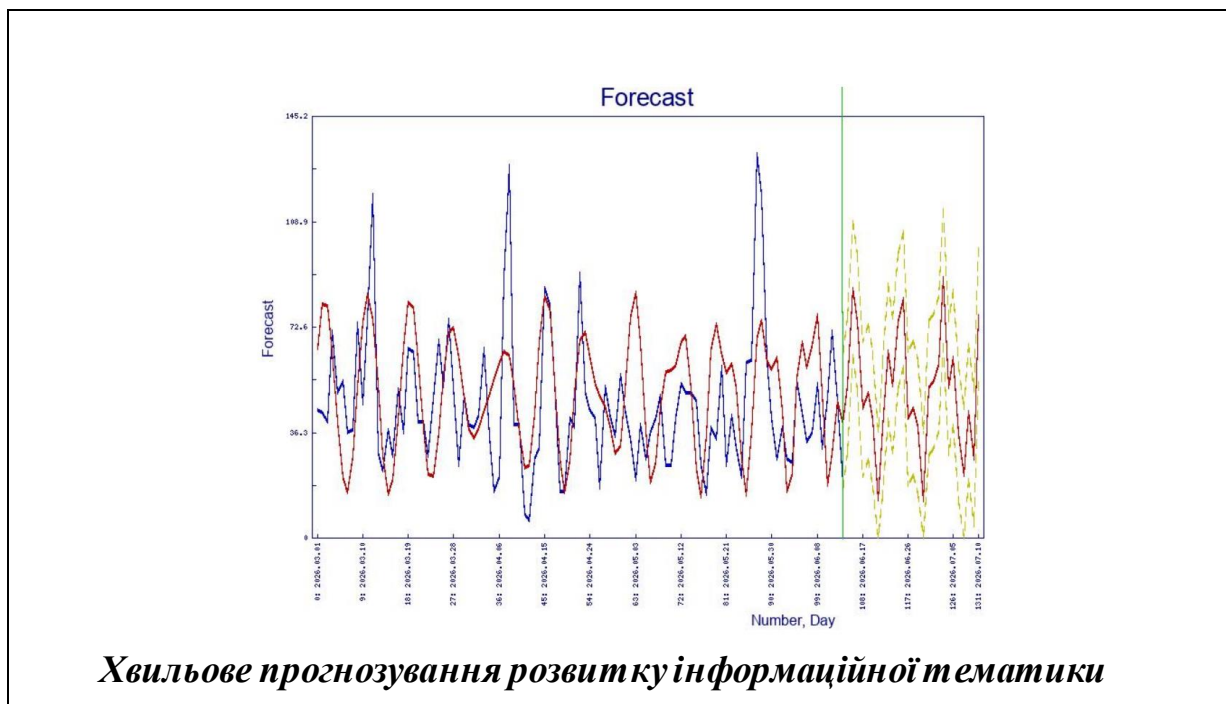
трансформували хаотичний потік даних у структуровану та верифіковану вибірку, то прогнозування переводить систему з режиму реактивного спостереження у превентивний аналітичний інструмент. У архітектурі InfoStream цей процес реалізується через двокомпонентну модель: динамічний прогноз (кількісний, базований на статистичному та машинному аналізі часових рядів) та семантичний прогноз (якісний, що використовує контекстуальні та логіко-мовні можливості великих мовних моделей). Їх інтеграція забезпечує не лише оцінку ймовірності інформаційних імпульсів, а й моделювання напрямків розвитку дискурсу, що є критично важливим для систем підтримки прийняття рішень.

### *Динамічний прогноз обсягу потоку*

Фундаментом динамічного прогнозу виступає класичний модуль «Динаміка понять», розширений сучасними методами аналізу часових рядів. Система агрегує хронологічні метадані: частотність згадувань ключових термінів, активність джерел, рівень цифрової насиченості, тональність та обсяг публікацій у заданому темпоральному інтервалі. На відміну від статичних гістограм, що відображали лише факти минулих подій, сучасний аналітичний контур застосовує декомпозицію рядів, виявлення сезонності та трендів, а також алгоритми прогнозування.



***Прогнозування розвитку інформаційного потоку методом N-грам***



### ***Семантичний прогноз розвитку тенденцій***

Якщо динамічний прогноз відповідає на запитання «скільки та коли?», то семантичний прогноз фокусується на «що саме та куди?». Великі мовні моделі, інтегровані в InfoStream, аналізують не лише частотність термінів, а й їхній контекст, причинно-наслідкові зв'язки, тональні зсуви та латентні тематичні кластери. На основі цієї обробки формується ймовірнісна картина розвитку інформаційної події або галузевого дискурсу.

#### ***Методологічні засади LLM-прогнозування:***

- ***Контекстуальне зіставлення:*** Модель порівнює поточний стан дискурсу з історичними аналогами з ретроспективного фонду, виявляючи повторювані сценарії.
- ***Генерація сценаріїв розвитку:*** LLM формує декілька ймовірних векторів з описом очікуваних подій, потенційних учасників, точок кристалізації нових тем та можливих медіа-наративів.
- ***Семантичне ранжування тенденцій:*** Модель оцінює стійкість дискурсу, ймовірність його трансформації в регуляторне рішення, ринковий імпульс або суспільний резонанс.
- ***Інтеграція з RAG-верифікацією:*** Кожен прогнозний сценарій обов'язково прив'язується до індексованих джерел, експертних коментарів або нормативних текстів, що унеможливорює вільну спекуляцію та забезпечує методологічну прозорість.

#### ***Симбіоз динамічного та семантичного прогнозування***

У практичній аналітиці ці два контури не працюють ізольовано. Їхня інтеграція реалізується через єдиний аналітичний процес:

1. **Сигнальна фаза:** Динамічна модель фіксує аномальне зростання обсягу потоку або зміну циклічності, генеруючи сповіщення про ймовірний інформаційний імпульс.
2. **Контекстуальна фаза:** LLM-модуль аналізує семантичний склад нового потоку, виявляє основні сутності, тональні зсуви та формує первинні сценарії розвитку події.
3. **Верифікаційна фаза:** RAG-архітектура перевіряє згенеровані висновки на відповідність індексованим джерелам, ретроспективним аналогам та експертним оцінкам.
4. **Інтеграційна фаза:** Результати подаються у вигляді структурованої аналітичної довідки з прогностичними метриками, сценаріями розвитку та рекомендаціями для СППР.

Такий підхід трансформує моніторинг з інструменту фіксації фактів у систему передбачення інформаційних наслідків.

### ***Практичне застосування в контурі прийняття рішень***

Прогнозні модулі InfoStream інтегруються у робочі процеси через кілька стандартних сценаріїв:

- **Кризовий моніторинг:** оперативне прогнозування масштабу інформаційної події, ймовірності медіа-резонансу та рекомендації щодо пріоритетності реакції.
- **Регуляторний аналіз:** семантичне моделювання реакції стейкхолдерів на законодавчі ініціативи, прогноз термінів ухвалення рішень та очікуваних публічних наративів.
- **Ринкова та конкурентна аналітика:** динамічний прогноз медіа-активності конкурентів, семантичне оцінювання ймовірності запуску нових продуктів або зміни позиціонування.
- **Аналітичні дослідження:** використання 30-річного архіву для калібрування моделей, перевірки гіпотез щодо циклічності дискурсів та валідації прогнозованих тенденцій.

Користувач має можливість налаштовувати горизонт прогнозу, вагові коефіцієнти динамічних та семантичних компонентів, а також активувати режим «експертного затвердження», де автоматично згенеровані сценарії валідуються аналітиком перед фінальним використанням.

Попри високі аналітичні можливості, прогнозування в інформаційному середовищі має природні обмеження: непередбачувані події, раптові зміни редакційної політики джерел, маніпулятивні кампанії або технічні збої в інфраструктурі мережі. Тому система InfoStream застосовує принципи обережного прогнозування: обов'язкове подання інтервальних оцінок замість точкових прогнозів, маркування рівня впевненості моделі, прозорість логіки генерації та незмінна прив'язка до першоджерел. Прогноз розглядається не як беззаперечний факт, а як аналітичний інструмент підтримки прийняття рішень, що потребує експертної інтерпретації та постійної калібрування на основі нових даних.

# РОЗДІЛ 5. ІНТЕГРАЦІЯ В СППР: ВІД СИРИХ ДАНИХ ДО УПРАВЛІНСЬКИХ РІШЕНЬ

## БЛОК-СХЕМА ВИКЛАДУ МАТЕРІАЛУ



Якщо попередні розділи розкрили архітектурні основи системи, класичний інструментарій пошуку та сучасні аналітичні модулі (LLM, RAG, часові ряди, прогнозування), то тут фокус зміщується на прикладну інтеграцію: як перетворити відібрані, верифіковані та структуровані дані на управлінські інсайти, придатні для стратегічного планування, оперативного реагування та комунікаційного впливу.

У цифровому середовищі інформаційна цінність визначається не обсягом зібраних матеріалів, а здатністю системи трансформувати їх у відтворювані, аудировані та контекстуально обґрунтовані висновки. Саме тому розділ побудовано навколо єдиного методологічного ланцюжка: від сирих новинних потоків, через семантико-аналітичну обробку, до структурованих аналітичних довідок та їх інтеграції в корпоративні системи. Кожен етап супроводжується алгоритмами та сценаріями застосування в реальних управлінських, кризових та ринкових контекстах.

Особливу увагу приділено трансформації аналітичних результатів у формат, зрозумілий для прийняття рішень: структуровані шаблони довідок, обов'язкова прив'язка до першоджерел, розрізнення сигналів та шуму, оцінка впливу медіа-хвиль на цільові об'єкти, а також автоматизація рутинних операцій без втрати методологічної прозорості.

## **5.1. Трансформація новинного потоку в аналітичні інсайти**

Трансформація сирового новинного потоку в аналітичні інсайти є центральним процесом, що визначає ефективність будь-якої системи моніторингу в контурі підтримки прийняття рішень. У цифровому середовищі, де обсяги публікацій вимірюються сотнями тисяч документів на добу, а рівень інформаційного шуму, змістовного дублювання та маніпулятивного впливу постійно зростає, простий пошук або статистичне підсумовування вже не забезпечують управлінської цінності. Інсайт виникає лише на перетині контрольованої вибірки, семантичної обробки, верифікації джерел та прогнозного моделювання. У системі InfoStream цей процес реалізовано як багаторівневу аналітичну процедуру, де класичні алгоритми інформаційного пошуку інтегровані з сучасними модулями штучного інтелекту та аналізу часових рядів.

### ***Етап 1. Первинна нормалізація та контрольована вибірка***

Трансформація починається з формування релевантного інформаційного каналу. На цьому етапі система здійснює безперервне сканування гетерогенних джерел, нормалізує контент до єдиного текстового формату, вилучає метадані та індексує матеріали в централізованому сховищі. Основним методологічним інструментом виступає повнотекстова інформаційно-пошукова система InfoReS та формалізована мова запитів, яка гарантує, що подальшій обробці піддається не «весь Інтернет», а семантично валідна, контрольована та відтворювана вибірка. Результат етапу –

структурований масив документів з мінімальним рівнем шуму та готовий до контекстуального аналізу.

### ***Етап 2. Семантична кластеризація та інформаційне моделювання***

На цьому рівні відбувається перехід від синтаксичної відповідності до смислової інтерпретації. Класичний інструментарій InfoStream реалізує побудову «інформаційних портретів» на основі частотно-статистичного методу: система визначає ваги окремих термінів, речень, абзаців та документів, ранжуючи рубрики, мови, географічні назви, персоналії, джерела, тональність та рівень цифрової насиченості. Адаптивний «класифікатор-навігатор» та алгоритми формування сюжетних ланцюжків групують матеріали за тематичними кластерами без повторного формулювання запиту. У сучасній архітектурі цей контур розширено великими мовними моделями, які виконують синтез звітів на основі аналізу багатьох документів, витягнення сутностей, оцінку семантичної близькості та автоматичне реферування.

### ***Етап 3. Верифікація та архітектурна гарантія достовірності***

На цьому етапі інтегрується архітектура Retrieval-Augmented Generation. Замість вільної генерації текстів, система працює за принципом «спочатку пошук – потім синтез – завжди посилання». Кожен абзац, висновок або резюме у згенерованій аналітичній довідці обов'язково супроводжується гіперпосиланням на індексоване першоджерело, дату публікації та оригінальний уривок. RAG-контур виконує перехресну перевірку фактів між незалежними джерелами, маркує суперечливі дані та твердження без прямого підтвердження в оперативному чи ретроспективному фонді. Це усуває ризик «галюцинацій» ШІ, забезпечує методологічну прозорість та робить результати придатними для офіційних довідок, експертних висновків та внутрішнього аудиту.

### ***Етап 4. Динамічне моделювання та прогнозування інформаційних процесів***

Інсайт набуває стратегічної цінності, коли містить не лише опис поточного стану, а й оцінку траєкторії розвитку події. Система застосовує методи аналізу часових рядів для трансформації класичних гістограм динаміки понять у інтерактивні інструменти виявлення хвильових патернів, аномальних інформаційних спалахів, циклічності медіауваги та латентних трендів. Статистичні та машинні моделі формують динамічний прогноз: ймовірність зміни обсягу потоку, рівня резонансу або активізації джерел. Паралельно LLM, отримуючи контекст з ретроспективного фонду, генерує семантичний прогноз, моделює ймовірні напрямки дискурсу, очікувані реакції стейкхолдерів, точки кристалізації нових подій та потенційні зони ризику. Поєднання цих підходів переводить моніторинг із режиму реактивного спостереження у превентивний аналітичний контур.

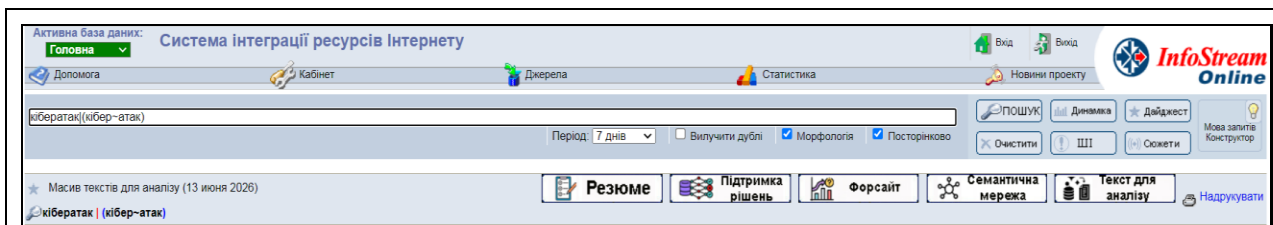
## **Етап 5. Формування аналітичного інсайту та інтеграція в контур СППР**

Фінальний етап – трансформація структурованих, верифікованих та збагачених даних у формат, придатний для прийняття рішень. Система генерує уніфіковані аналітичні довідки, кризові брифінги, медіа-репутаційні звіти та сценарні оцінки за попередньо визначеними шаблонами. Користувач має можливість налаштовувати глибину деталізації, пріоритети джерел, рівень допустимої тональної поляризації та активувати режим «експертного затвердження», де автоматично сформований матеріал проходить валідацію аналітиком перед інтеграцією в робочі процеси.

### **Методологічні принципи трансформації**

1. Відтворюваність: кожен інсайт базується на фіксованому запиті, верифікованій вибірці та збережених налаштуваннях каналу, що дозволяє незалежно відтворити результат у будь-який момент.
2. Прозорість логіки: чітке розмежування класичної фільтрації, семантичної обробки, верифікації та прогнозного моделювання виключає «чорну скриньку» з процесу прийняття рішень.
3. Відстежуваність: обов'язкова прив'язка до першоджерел, збереження історії запитів, контрольні суми архівних записів та логування змін у конфігурації каналів.
4. Мінімізація алгоритмічного зміщення: поєднання статистичних метрик з експертною валідацією та можливістю інтерактивного коригування вагових коефіцієнтів через інформаційний портрет.
5. Використання ШІ: LLM працюють виключно з релевантною вибіркою, не генерують фактів без джерельного підтвердження, а режим RAG гарантує, що кожен висновок піддається перехресній перевірці.

Таким чином, методологія трансформації новинного потоку в аналітичні інсайти в InfoStream є не сукупністю ізольованих функцій, а цілісним науково-обґрунтованим процесом, де кожен етап підпорядкований єдиній логіці: від точного відбору → до семантичної структуризації → до обов'язкової верифікації → до прогнозного моделювання → та інтеграції в управлінський контур. Саме такий підхід перетворює моніторинг з інструменту пасивного збору даних на активний компонент інтелектуальної підтримки прийняття рішень.



Активна база даних: Система інтеграції ресурсів Інтернету

Головна | Доломога | Кабінет | Джерела | Статистика | Новини проєкту | InfoStream Online

кібератака(кібер-атак) | Пошук | Динаміка | Дайджест | Мова запису | Конструктор

Період: 7 днів | Вилучити дублі | Морфологія | Тісторично

Масив текстів для аналізу (13 июня 2026)

кібератака | кібер-атак

Резюме | Підтримка рішень | Форсайт | Семантична мережа | Текст для аналізу | Надрукувати

**Інтелектуальні режими аналізу: застосування великих мовних моделей**

## 5.2. Сценарії застосування

Інтеграція системи моніторингу в контур систем підтримки прийняття рішень реалізується через конкретні прикладні сценарії, кожен із яких має власну цільову функцію, набір аналітичних метрик та алгоритмічну процедуру обробки. Кожен сценарій демонструє симбіоз класичного пошукового ядра InfoReS, інструментів динамічного уточнення та сучасних ШІ-модулів, трансформуючи сирий новинний потік у верифіковані інсайти, придатні для оперативного та стратегічного реагування.

### **Кризовий моніторинг**

**Мета:** оперативне виявлення, класифікація та відстеження інформаційних імпульсів, що несуть потенційні операційні, фінансові, регуляторні або репутаційні ризики, з метою мінімізації часу реакції та активації протоколів управління кризами.

### ***Технологічний інструментарій:***

**Класичний рівень:** режим «Актуально», «плюс-словники» сигнальних термінів, оператори контекстної близькості, фільтрація дублів, ранжування за авторитетністю джерел.

**Сучасний рівень:** LLM для миттєвої класифікації типу інциденту, витягнення сутностей (локація, відповідальні органи, постраждалі об'єкти), RAG для прив'язки кожного твердження до першоджерела, детектування аномалій у часових рядах для відокремлення органічного спалаху від координованого втручання.

### ***Алгоритм:***

1. Формування запиту з використанням сигнальних лексем, географічних/галузевих фільтрів та обмеження часовим вікном (режим реального часу).
2. Активація режиму «Актуально»: фільтрація за мінімальним часом життя документа, локальною лексичною близькістю, глобальною відмінністю від історичного фону та відсутністю дублів.
3. Автоматична кластеризація LLM: групування повідомлень за типом кризи (техногенна, регуляторна, соціальна, кіберінцидент, фінансовий шок).
4. RAG-верифікація: генерація попереднього резюме інциденту з гіперпосиланнями на URL, дати, уривки оригінальних текстів та індикаторами ступеня підтвердження фактів.
5. Хвильовий аналіз: побудова динаміки згадувань, виявлення фази росту/піку, прогноз обсягу потоку.

**Результат та інтеграція в СПІР:** структурована кризова довідка з хронологією, суб'єктами, оцінкою масштабів, верифікованими джерелами та прогнозним вектором.

## **Медіа-репутація**

Мета: оцінка тональності, інтенсивності та структури медіа-присутності компанії, бренду або персоналії; виявлення драйверів позитивного/негативного дискурсу; моніторинг ефективності комунікаційних кампаній та виявлення маніпулятивних наративів.

### ***Технологічний інструментарій:***

Класичний рівень: інформаційний портрет (ранжування джерел, тональності, географії, персоналії), сюжетні ланцюжки, матриці взаємозв'язків рубрик, ретроспективний фонд для порівнянь у часі.

Сучасний рівень: LLM для контекстуальної оцінки тональності (з урахуванням сарказму, умовних конструкцій, галузевого сленгу), RAG для формування медіа-звітів із цитуванням першоджерел, часові ряди для виявлення циклічності медіа-уваги та аномальних спалахів негативу.

### ***Алгоритм:***

1. Створення запиту з точним входженням назв брендів/персон ({}), контекстними фільтрами галузі/продукту та виключенням рекламного/технічного шуму.
2. Побудова інформаційного портрета: візуалізація розподілу за джерелами, мовами, тональністю, розміром документів, цифровою насиченістю.
3. Автоматичне формування сюжетів: семантичне групування повідомлень, виділення основних тематичних кластерів (наприклад, «якість продукту», «регуляторні перевірки», «корпоративна культура», «ринкові позиції»).
4. LLM-аналіз тональності та контексту: розрізнення фактологічних згадувань, експертних коментарів, маніпулятивних конструкцій, координованих бот-мереж.
5. RAG-генерація медіа-довідки: цитування першоджерел, хронологія основних публікацій, індекси лояльності ЗМІ, порівняння з історичними зрізами з ретроспективного фонду.

Результат та інтеграція в СППР: структурований медіа-рейтинг із динамікою тональності, індексами впливу джерел, картами тематичних кластерів та верифікованими висновками. Використовується для коригування PR-стратегій, підготовки відповідей на інформаційні атаки, оцінки ROI комунікаційних кампаній та формування рекомендацій для топ-менеджменту.

## **Ринкова аналітика**

Мета: виявлення макроекономічних, галузевих та конкурентних тенденцій; прогнозування змін попиту, регуляторного тиску або технологічних зрушень; оцінка ринкових позицій та поведінки основних гравців.

### ***Технологічний інструментарій:***

Класичний рівень: тематичні рубрики (rubrXX), багаторівневі булеві запити, динаміка понять (гістограми згадувань), таблиці взаємозв'язків рубрик, фільтрація за країнами/джерелами/мовами.

Сучасний рівень: LLM для витягнення фінансових/статистичних маркерів, семантичного прогнозування розвитку ринкових дискурсів, RAG для верифікації аналітичних звітів, статистичні моделі часових рядів (ARIMA/Prophet/LSTM) для прогнозу обсягу потоку та виявлення структурних зламів.

### ***Алгоритм:***

Конструювання галузевого запиту з використанням операторів розширення семантики, фільтрів за регіоном, типом джерел (аналітика, регулятори, біржі, експертні блоги).

Аналіз динаміки понять: побудова часових рядів згадувань ключових термінів, виявлення сезонності, трендів, точок зміни поведінки ринку.

Аналіз співзв'язності: матриця взаємозв'язків рубрик виявляє приховані асоціації (наприклад, зв'язок між «дефіцитом сировини», «змінами в тарифах» та «консолідацією ринку»).

LLM-синтез: автоматичне реферування аналітичних звітів, витягнення цифрових індикаторів, формування сценаріїв розвитку галузі.

RAG-верифікація: прив'язка кожного прогнозного висновку до першоджерела, дати, експертного контексту та історичних аналогів з ретроспективного фонду.

Результат та інтеграція в СППР: ринкова аналітична довідка з динамікою індикаторів, картами впливу факторів, сценарними прогнозами (динамічний + семантичний) та верифікованими джерелами. Інтегрується в ERP/BI-середовища, використовується для стратегічного планування, інвестиційних рішень, моніторингу конкурентів та адаптації бізнес-моделей.

**Дослідження ставлення влади, суспільства та стейкхолдерів до компанії для розробки стратегії впливу**

Мета: моніторинг регуляторних ініціатив, законодавчих змін, позицій органів влади та громадських інституцій; оцінка суспільного резонансу політичних/соціальних рішень; формування комунікаційних стратегій взаємодії з державними та громадськими зацікавленими сторонами.

### ***Технологічний інструментарій:***

Класичний рівень: офіційні портали, прес-служби, парламентські/урядові джерела, оператори точного входження (|), контекстне слідування для фіксації формулювань законів/рішень, фільтрація за географією та інституційною приналежністю.

Сучасний рівень: LLM для аналізу нормативного контексту, витягнення статей/пунктів/термінів імплементації, RAG для формування GR-довідок із цитуванням офіційних текстів, часові ряди для відстеження життєвого циклу законодавчої ініціативи (анонс → обговорення → голосування → імплементація → оцінка наслідків).

#### ***Алгоритм:***

Запит із точними назвами законопроектів, органів влади, регуляторних відомств, фільтрація за датами сесій/засідань.

Відстеження життєвого циклу події: автоматичне групування за фазами, виявлення основних рішень, реакцій експертів, громадських об'єднань, міжнародних інституцій.

LLM-аналіз семантики нормативних змін: порівняння редакцій, витягнення змінених норм, оцінка потенційного впливу на галузь/ринок.

RAG-генерація GR-звіту: цитування джерел, посилання на номери документів, дати оприлюднення, прямі уривки нормативних актів.

Прогнозування: динамічний прогноз активності дискусії, семантичний прогноз ймовірних коригувань або супутніх ініціатив.

Результат та інтеграція в СППР: структурована GR/PR-довідка з хронологією, основними суб'єктами, аналізом нормативних змін, оцінкою суспільного резонансу та верифікованими посиланнями. Використовується для підготовки позиційних документів, комунікаційних відповідей, участі в громадських обговореннях, формування стратегій регуляторного взаємодії та мінімізації політичних/репутаційних ризиків.

#### **Інтеграція сценаріїв у контур СППР**

Усі чотири сценарії базуються на єдиному технологічному принципі: контрольована вибірка → верифікована Ш-обробка → прогнозне моделювання → інтеграція в управлінський цикл. Класичне ядро InfoReS гарантує точність, відтворюваність та мінімізацію інформаційного шуму. LLM забезпечує семантичну глибину та та узагальнення інформації з множини документів. RAG виконує роль методологічного стандарту аудиту, усуваючи ризик неперевіраних генерацій. Аналіз часових рядів трансформує хронологічні дані у прогностичні індикатори. Ретроспективний фонд (понад 30 років) забезпечує історичну валідацію моделей та калібрування прогнозів.

### **5.3. Алгоритми побудови аналітичних довідок**

Аналітична довідка є фінальним продуктом інформаційно-аналітичного процесу InfoStream, що трансформує структуровану вибірку новинних документів у верифікований, контрольований та готовий до інтеграції в системи підтримки прийняття рішень інструмент. Її побудова не є довільною генерацією тексту, а реалізується через детермінований алгоритмічний

контур, який поєднує класичні методи повнотекстового пошуку, частотно-статистичної обробки та лінгво-аналітичного кластерного аналізу з сучасною архітектурою RAG та семантичними моделями LLM.

### **Алгоритмічний контур побудови довідки**

Процес формування довідки реалізується у чотири послідовні етапи, кожен з яких підпорядкований вимогам прозорості та контрольованості:

1. Формування контексту вибірки. Користувач конструює пошуковий запит мовою InfoReS. Класичне ядро системи виконує первинний відбір, гарантуючи релевантність та відтворюваність масиву документів.
2. Семантико-аналітична обробка. Відібрана вибірка передається в аналітичний контур, де LLM-модулі виконують узагальнення множини документів.
3. Верифікація та контекстуалізація. Кожний згенерований висновок проходить процедуру прив'язки до першоджерела.
4. Структурування, експорт та інтеграція. Система генерує уніфікований шаблон довідки з розділами, метаданими та гіперпосиланнями.

### **Структурні компоненти аналітичної довідки**

Типова довідка (резюме), сформована в InfoStream, містить наступні методологічно обґрунтовані блоки:

- Метадані: тема моніторингу, часовий інтервал, кількість оброблених документів, мовна та джерельна репрезентативність та час генерації.
- Хронологія та сюжетні ланцюжки: стислий перебіг розвитку події з акцентом на точки кристалізації, офіційні заяви, реакції регуляторів або ринку. Сюжети формуються автоматично на основі семантичного ранжування та об'єднують розрізнені публікації у єдиний наратив.
- Ключові суб'єкти, об'єкти та взаємозв'язки: персоналії, організації, географія, фінансові/статистичні маркери.
- Динаміка та хвильовий аналіз: інтерактивні графіки появи понять, виявлення аномалій, оцінка циклічності медіауваги та прогнозні показники.
- Верифіковані висновки: структуровані тези, кожен з яких супроводжується посиланням на джерело, датою публікації та уривком оригінального тексту.
- Маркери для СППР: потенційні ризики, точки впливу, сценарії реагування, індикатори достовірності та рівень експертної валідації.

Поєднання класичного пошукового ядра, багаторівневої фільтрації, RAG-верифікації та цитування першоджерел перетворює аналітичну довідку з технічного звіту на науково обґрунтований інструмент підтримки рішень.

★ Масив текстів для аналізу (13 июня 2026)

фейк

Резюме Підтримка рішень Форсайт Семантична мережа Текст для аналізу Надрукувати

Ранг	Ланцюжок	Розширене пояснення ланцюжка	Шляхи
15	Проблема фейк - Відсутність інформації - Дезінформація - Фейкові новини - Вирішення проблеми, обумовленої фейк	Починаючи від проблеми фейк, відсутність інформації може призвести до дезінформації та розповсюдження фейкових новин. Для вирішення цієї проблеми необхідно забезпечити доступ до достовірної інформації та навчити людей критично оцінювати інформацію.	Забезпечити доступ до достовірної інформації, навчати людей критично оцінювати інформацію, проводити інформаційні кампанії щодо підвищення медіаграмотності.
24	Проблема фейк - Довіра до джерел - Перевірка інформації - Відслідковування - Фейкові новини - Вирішення проблеми, обумовленої фейк	Якщо людям бракує інформації, вона може довіряти сумнівним джерелам. Тому необхідно перевіряти інформацію та відслідковувати її. Це допоможе зменшити вплив фейкових новин.	Навчати людей перевіряти інформацію, використовувати лише достовірні джерела, проводити фактчекінг.
33	Проблема фейк - Медіаграмотність - Критичне мислення - Розпізнавання фейків - Вирішення проблеми, обумовленої фейк	Люди повинні вміти критично оцінювати інформацію та розпізнавати фейки. Медіаграмотність та критичне мислення допоможуть людям орієнтуватися в інформаційному просторі.	Навчати людей медіаграмотності, проводити інформаційні кампанії щодо підвищення критичного мислення.
42	Проблема фейк - Законодавство - Регулювання інформації - Відповідальність за фейки - Вирішення проблеми, обумовленої фейк	Існує потреба у законодавстві, яке регулює інформацію та передбачає відповідальність за фейки. Це допоможе зменшити розповсюдження фейкових новин.	Розробити законодавство щодо регулювання інформації, передбачити відповідальність за фейки, створити органи контролю за інформаційним простором.
51	Проблема фейк - Соціальна свідомість - Єдність суспільства - Подолання фейків - Вирішення проблеми, обумовленої фейк	Соціальна свідомість та єдність суспільства допоможуть подолати фейки. Якщо суспільство єдине у своєму ставленні до фейків, воно зможе ефективно протистояти їм.	Проводити інформаційні кампанії щодо підвищення соціальної свідомості, сприяти єдності суспільства, виховувати патріотизм та громадянську активність.

Поновлення бази даних здійснюється кожні 15 хвилин

InfoStream® - моніторинг новин © 2000-2026 EVIsti Information Center

Наверх

## Формування рекомендацій для систем підтримки прийняття рішень

У методології InfoStream корисність реалізується через інтерактивні дашборди СППР, де візуалізуються метрики ефективності каналів, прогнози, тональні зрушення та матриці взаємозв'язків. Користувач може налаштовувати вагові коефіцієнти метрик відповідно до пріоритетів конкретного управлінського сценарію: у кризовому моніторингу пріоритетом є швидкість та точність, у стратегічному плануванні – повнота та прогноз, у комунікаційному впливі – корисність та інтегрованість у корпоративні системи.

★ Масив текстів для аналізу (13 июня 2026)

фейк

Резюме Підтримка рішень Форсайт Семантична мережа Текст для аналізу Надрукувати

### Прогноз розвитку фейк на майбутнє

Номер	Імовірність реалізації	Ланцюжок понять зі зв'язками	Пояснення
1	0,15	Поточний стан фейк - Шахрайські листи - Віруси та шкідливі ПЗ - Кібербезпека - Фейкові повідомлення	Шахраї використовують фейкові листи для розповсюдження вірусів та шкідливого ПЗ, що може призвести до проблем з кібербезпекою.
2	0,12	Поточний стан фейк - Біолабораторії - Фейкові звинувачення - Російська пропаганда - Дезінформація	Російська пропаганда використовує фейкові звинувачення щодо біолабораторій для розповсюдження дезінформації.
3	0,10	Поточний стан фейк - Арешт майна - Російський закон - Порушення прав людини - Європейський суд	Прийняття російського закону про арешт майна може призвести до порушень прав людини та розгляду справ у Європейському суді.
4	0,09	Поточний стан фейк - Війна в Україні - Фейкові новини - Масові маніпуляції - Демократичні інститути	Фейкові новини про війну в Україні можуть використовуватися для проведення масових маніпуляцій та підірвання демократичних інститутів.
5	0,08	Поточний стан фейк - Міжнародні відносини - Фейкові угоди - Дипломатичні скандали - Глобальна безпека	Фейкові угоди можуть призвести до дипломатичних скандалів та підірвати глобальну безпеку.
6	0,07	Поточний стан фейк - Біологічна зброя - Фейкові повідомлення - Паніка та страх - Населення	Фейкові повідомлення про біологічну зброю можуть викликати паніку та страх серед населення.
7	0,06	Поточний стан фейк - Кібератаки - Фейкові дані - Кібербезпека - Інформаційна війна	Кібератаки з використанням фейкових даних можуть бути частиною інформаційної війни.
8	0,05	Поточний стан фейк - Пропаганда - Фейкові новини - Суспільна думка - Влада	Пропаганда та фейкові новини можуть використовуватися для впливу на суспільну думку та підтримки влади.
9	0,04	Поточний стан фейк - Інформаційна війна - Фейкові повідомлення - Психологічна війна - Населення	Інформаційна війна з використанням фейкових повідомлень може бути частиною психологічної війни проти населення.
10	0,03	Поточний стан фейк - Шпигунство - Фейкові дані - Кібербезпека - Національна безпека	Шпигунство з використанням фейкових даних може підірвати кібербезпеку та національну безпеку.

**Узагальнене пояснення**

У сучасному світі фейки стали частиною інформаційного простору. Вони можуть використовуватися для впливу на суспільну думку, підірвання демократичних інститутів, проведення масових маніпуляцій та інших шкідливих цілей. Цей прогноз розвитку фейк на майбутнє визначає можливі шляхи їх розповсюдження та впливу на різні аспекти життя суспільства, включаючи кібербезпеку, міжнародні відносини, біологічну безпеку та національну безпеку. Для протидії фейкам необхідно розробляти та впроваджувати ефективні заходи щодо їх виявлення, запобігання та ліквідації наслідків.

## Форсайт-аналіз та прогнозування розвитку ситуації

## 5.4. Побудова семантичних карт на основі великих мовних моделей

Одним із сучасних напрямів розвитку інформаційно-аналітичних систем є автоматичне формування семантичних карт (мереж) предметних областей. На відміну від традиційних статистичних методів аналізу спільної зустрічальності термінів, сучасні великі мовні моделі дозволяють виявляти

не лише формальні лексичні зв'язки, а й приховані смислові відношення між поняттями, організаціями, персоналіями, подіями та процесами.

У системі InfoStream семантичні карти будуються на основі документів, відібраних за пошуковим запитом користувача. Після формування релевантної вибірки текстові матеріали передаються до спеціалізованого модуля семантичного аналізу, який за допомогою LLM виконує виділення сутностей, нормалізацію назв, визначення тематичних категорій та встановлення смислових взаємозв'язків між ними.

Результатом роботи є графова структура, у якій вузли відповідають інформаційним об'єктам (поняттям, організаціям, персоналіям, країнам, технологіям, подіям тощо), а ребра відображають виявлені семантичні зв'язки. Вага зв'язку визначається частотою його появи в документах, силою семантичної близькості або експертною оцінкою LLM.

На відміну від класичних статичних схем, семантична карта в InfoStream є інтерактивною. Кожний вузол мережі функціонує як активне посилання на відповідний інформаційний масив системи. При виборі вузла автоматично формується пошуковий запит до бази даних InfoStream, що дозволяє миттєво перейти від узагальненого представлення предметної області до конкретних документів, які стали підставою для формування даного концепту.

Таким чином реалізується принцип пояснюваної аналітики (Explainable Analytics): користувач може не лише бачити структуру інформаційного поля, але й оперативно перевіряти джерела, що підтверджують існування відповідних зв'язків.

### ***Алгоритм побудови семантичної карти***

Формування семантичної мережі виконується у декілька етапів:

1. Формування тематичної вибірки документів за допомогою пошукового ядра InfoReS.
2. Автоматичне вилучення сутностей та ключових понять із застосуванням LLM.
3. Нормалізація та об'єднання еквівалентних назв.
4. Визначення семантичних зв'язків між об'єктами.
5. Розрахунок ваг вузлів та зв'язків.
6. Побудова графової структури.
7. Візуалізація мережі та створення інтерактивних посилань на результати пошуку.



- дослідження тематичних екосистем;
- аналізу конкурентного середовища;
- побудови карт зацікавлених сторін;
- виявлення центрів впливу в інформаційному просторі;
- підтримки аналітичних та управлінських рішень.

### ***Використання семантичних мереж у СППР***

Інтеграція семантичних карт у контур систем підтримки прийняття рішень дозволяє перейти від аналізу окремих документів до аналізу структури предметної області в цілому. Керівник або аналітик отримує можливість бачити не лише інформаційні факти, але й систему взаємозв'язків між ними.

Поєднання класичного пошукового ядра InfoReS, багаторічного ретроспективного фонду даних та сучасних великих мовних моделей перетворює семантичну карту на динамічний інструмент дослідження інформаційного простору. У такому представленні кожний вузол мережі стає точкою входу до відповідного інформаційного масиву, а сама мережа — інтерактивною картою знань, що забезпечує швидкий перехід від узагальненого бачення ситуації до першоджерел і підтверджувальних документів.

# РОЗДІЛ 6. СЕРВІСНІ МОДУЛІ ТА ЕКСПЛУАТАЦІЯ СИСТЕМИ

## БЛОК-СХЕМА ВИКЛАДУ МАТЕРІАЛУ



## 6.1. Онлайн-доступ до оперативних та ретроспективних баз даних

Онлайн-доступ до оперативних та ретроспективних баз даних є основним режимом системи InfoStream, який забезпечує безпосередню взаємодію користувача з індексованими масивами новинної інформації. Цей режим реалізує концепцію «єдиного вікна» доступу до тисяч веб-ресурсів, поєднуючи інтуїтивну навігацію з розвиненим інструментарієм пошуку, динамічного уточнення вибірок та миттєвої аналітичної обробки результатів.

### *Архітектура інтерфейсу*

Робочий простір онлайн-доступу структуровано за модульним принципом і поділено на дві основні зони:

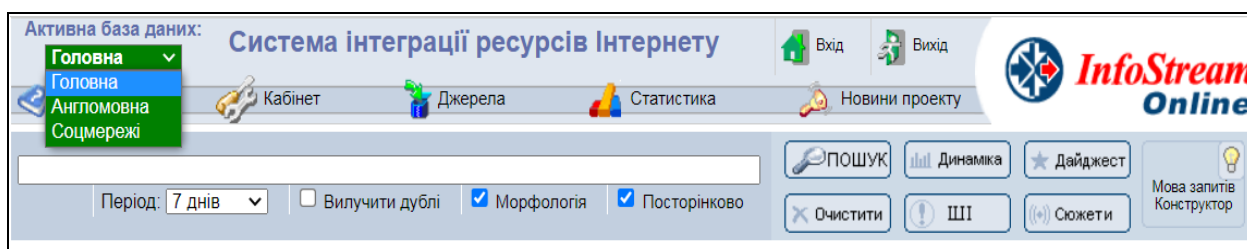
1. Верхнє системне меню – фіксована панель навігації, що містить селектор баз даних, розділи «Довідка», «Кабінет користувача», «Каталог джерел», «Статистика поступлень» та кнопку виходу. Меню залишається незмінним під час роботи, забезпечуючи стабільний доступ до системних функцій.
2. Робоча область – динамічний контейнер, який адаптується залежно від обраного контексту. У його центрі розташована форма пошукового запиту з полями для введення термінів, логічних операторів, перемикачів морфології та фільтрів дублювання. Залежно від активної бази, нижче або збоку відображаються допоміжні панелі: тематичні рубрики, перелік основних сюжетів, персональні інформаційні канали, персональні джерела або календар вибору інтервалу. Кожен елемент інтерфейсу є інтерактивним: клік на рубрику, джерело або заголовок миттєво відкриває відповідну вибірку або повний текст документа.

Перемикання між інформаційними контурами InfoStream реалізується через селектор баз даних, розташований у лівому верхньому куті інтерфейсу. Система підтримує шість основних контекстів:

- Оперативна база – містить документи, що надійшли за останні 7 днів. Оновлюється в режимі, близькому до реального часу. Інтерфейс активує панелі сюжетів, персональних каналів та джерел, орієнтуючись на моніторинг поточних подій.
- Ретроспективна база – архівне сховище, що формується понад 30 років. Інтерфейс підключає календар вибору інтервалу дат, дозволяючи

формувати зрізи за довільними періодами (від конкретної доби до повного діапазону).

- Англomовна база – фокус на зарубіжних ЗМІ та міжнародних агентствах. Структура аналогічна оперативній, але з пріоритетом на мовні та регіональні фільтри країн ЄС, США та глобальних агентств.
- Англomовна/ретроспективна – поєднує лінгвістичний контекст, що формується понад 30 років, із календарним інтерфейсом для досліджень іноземних дискурсів.
- Оперативна база соціальних медіа – містить повідомлення, що надійшли з соціальних медіа (Youtube, Telegram, Reddit, форуми, тощо) за останні 7 днів.
- Ретроспективна база соціальних медіа – архівне сховище, що формується понад 10 років.



При зміні контексту інтерфейс автоматично реконфігурує доступні інструменти, приховуючи непотрібні елементи та акцентуючи увагу на функціях, релевантних обраному режиму.

### ***Багаторівнева система фільтрації***

Фільтрація в InfoStream реалізується на трьох взаємодоповнюючих рівнях, що дозволяють поступово трансформувати широкий пошук у високоточну, аналітично придатну вибірку:

1. ***Базовий рівень (форма пошукового запиту)***. Користувач формує первинний запит за допомогою мови InfoReS: терміни природної мови, праві скорочення, логічні оператори (&, |, !, +, ,), контекстне слідування (~), контекстна близькість (@), точне входження (|), круглі дужки для багаторівневих конструкцій. Додатково доступні перемикачі:
  - Морфологія – враховувати/ігнорувати закінчення слів для розширення або звуження пошуку.
  - Дублі – показувати/приховувати змістовно ідентичні або семантично близькі документи.

- Діапазон дат – обмеження вибірки за періодом публікації (активується в ретроспективному контексті).
2. **Динамічне уточнення (Інформаційний портрет).** Після виконання пошуку система генерує частотно-статистичну модель отриманої вибірки. Портрет відображає ранжировані характеристики масиву: тематичні рубрики, мови, розмір повідомлень (малий/середній/великий), цифрову насиченість (низька/середня/висока), країни джерел, географічні назви, імена джерел, персоналії, тональність документів та характерні терміни. Найвагоміші елементи позначаються символом \*. Клік на будь-яку характеристику миттєво додає фільтр до запиту. Можливе комбінування кількох фільтрів через логічні зв'язки AND (обов'язкова наявність) або NOT (виключення). Цей рівень усуває необхідність повторного введення складних запитів, замінюючи їх інтерактивним уточненням.
  3. **Контекстна навігація (Класифікатор-навігатор).** Реалізує принцип «пошукових папок». На основі алгоритмів семантичної близькості система будує адаптивне дерево уточнень результатів первинного пошуку. Наприклад, запит бензин автоматично пропонує гілки: ціни, експорт, пальне, ринки, акцизи тощо. Користувач може послідовно звужувати вибірку, переходячи від загального терміна до конкретних підтем без зміни синтаксису запиту. Навігатор особливо ефективний для розвідки нових тематичних полів та виявлення латентних зв'язків у великих масивах.

Додатково доступні фільтри за персональними джерелами (попередньо відібрані через «Каталог джерел»), рівнем цифрової насиченості (корисний для пошуку аналітики, таблиць, рейтингів), розміром повідомлень та мовною версією. Усі фільтри зберігаються разом із запитом при конфігурації персонального інформаційного каналу.

Активна база даних: Система інтеграції ресурсів Інтернету

Головна | Допомога | Кабінет | Джерела | Статистика | Новини проекту | Вхід | Вихід | InfoStream Online

### Каталог джерел

Рубрики каталогу	Кількість
Інформаційні агентства	252
Теле-радіо канали	192
Газети	190
Щотижневики та журнали	179
Офіційні джерела	436
Асоціації, компанії	553
Інтернет-видання	2107
Інші	2
Архиви	6828

Всього джерел у системі: 10739  
Відібрано користувачем: 7  
Активність джерел: 3455  
Загальна: 98%

Пошук

Інформаційні агентства	Теле-радіо канали	Газети	Щотижневики та журнали	Офіційні джерела	Асоціації, компанії	Інтернет-видання	Інші	Архиви
Інтернет-видання								
www.UZ	RUS	ED	2026.06.13	11:27				
www.RU	RUS	EW	2026.06.11	12:11				
www.MD	RUS	ED	2026.06.13	16:11				
www.UA	UKR	ED	2026.06.13	16:21				
www.UA	UKR	ED	2026.06.13	15:14				
www.UA	UKR	EW	2026.06.13	16:00				
www.UA	UKR	ED	2026.06.13	16:00				
www.UA	UKR	EW	2026.06.13	06:00				
www.UA	UKR	EM	2026.06.12	15:00				
www.UA	UKR	ED	2026.06.13	16:00				
www.UA	UKR	ED	2026.06.13	14:00				
www.UA	UKR	EW	2026.06.12	15:00				
www.UA	UKR	EW	2026.06.12	21:00				
www.UA	UKR	ED	2026.06.12	18:10				
www.UA	UKR	EW	2026.06.12	21:00				
www.UA	UKR	EW	2026.06.12	20:10				
www.UA	UKR	EM	2026.06.13	16:00				
www.UA	UKR	EW	2026.06.12	21:00				
www.UA	UKR	ED	2026.06.13	16:00				
www.UA	RUS	EW	2026.06.12	21:00				
www.UA	UKR	ED	2026.06.13	16:00				

0312.ua Ужгород  
032.ua Львів  
0332.ua Луцьк  
0342.ua Івано-Франківськ  
0352.ua Тернопіль  
0362.ua Рівне  
0372.ua Чернівці  
0382.ua Хмельницький  
0412.ua Житомир  
04141.com.ua Звягель  
0432.ua Вінниця  
04563.com.ua Біла Церква  
04597.com.ua Ірпінь  
0462.ua Чернігів  
0472.ua Черкаси  
048.ua Одеса  
04868.com.ua Черноморськ  
0512.com.ua Миколаїв

online.infostream.ua/source.php

### Каталог джерел

### Інтеграція з аналітичними модулями

У сучасній архітектурі InfoStream фільтрована вибірка є готовим вхідним контуром для подальшої обробки. Без зміни контексту користувач може активувати:

- інформаційний пошук;
- автоматичне реферування та побудову дайджестів;
- семантичне групування в сюжетні ланцюжки;
- діаграми динаміки понять та хвильового аналізу;
- інтеграцію з ШІ, як працює виключно з релевантним, очищеним від дублів та шуму масивом, що суттєво підвищує стабільність генерацій та зменшує ризик контекстуальних спотворень.

Ефективна робота в онлайн-режимі базується на ітеративному підході:

1. Починати з мінімального набору термінів, використовувати морфологію для розширення охоплення.
2. Аналізувати інформаційний портрет першої вибірки, відключати дублі, фільтрувати за рубриками, мовами або країнами.
3. Використовувати класифікатор-навігатор для швидкого звуження без переформулювання запиту.

4. Зберігати оптимальну конфігурацію як персональний канал через «Кабінет користувача».
5. Регулярно оновлювати список персональних джерел та перевіряти статистику поступлень для корекції «сліпих зон».

## **6.2. Автоматизована доставка контенту**

Автоматизована доставка інформаційних вибірок є завершальною ланкою технологічного процесу моніторингу, що трансформує оброблені, верифіковані та структуровані дані в інструмент безперервного інформування та інтеграції в робочі процеси. У системі InfoStream реалізовано модульну архітектуру трансляції контенту, яка поєднує перевірені методи поширення інформації (електронна пошта, RSS) із сучасними програмними інтерфейсами, веб-віджетами. Кожен канал налаштовується індивідуально під аналітичні задачі, зберігаючи методологічну прозорість, прив'язку до першоджерел та контроль над логікою фільтрації на всіх етапах передачі даних.

### ***Email-розсилки: конфігурування каналів та структура звітів***

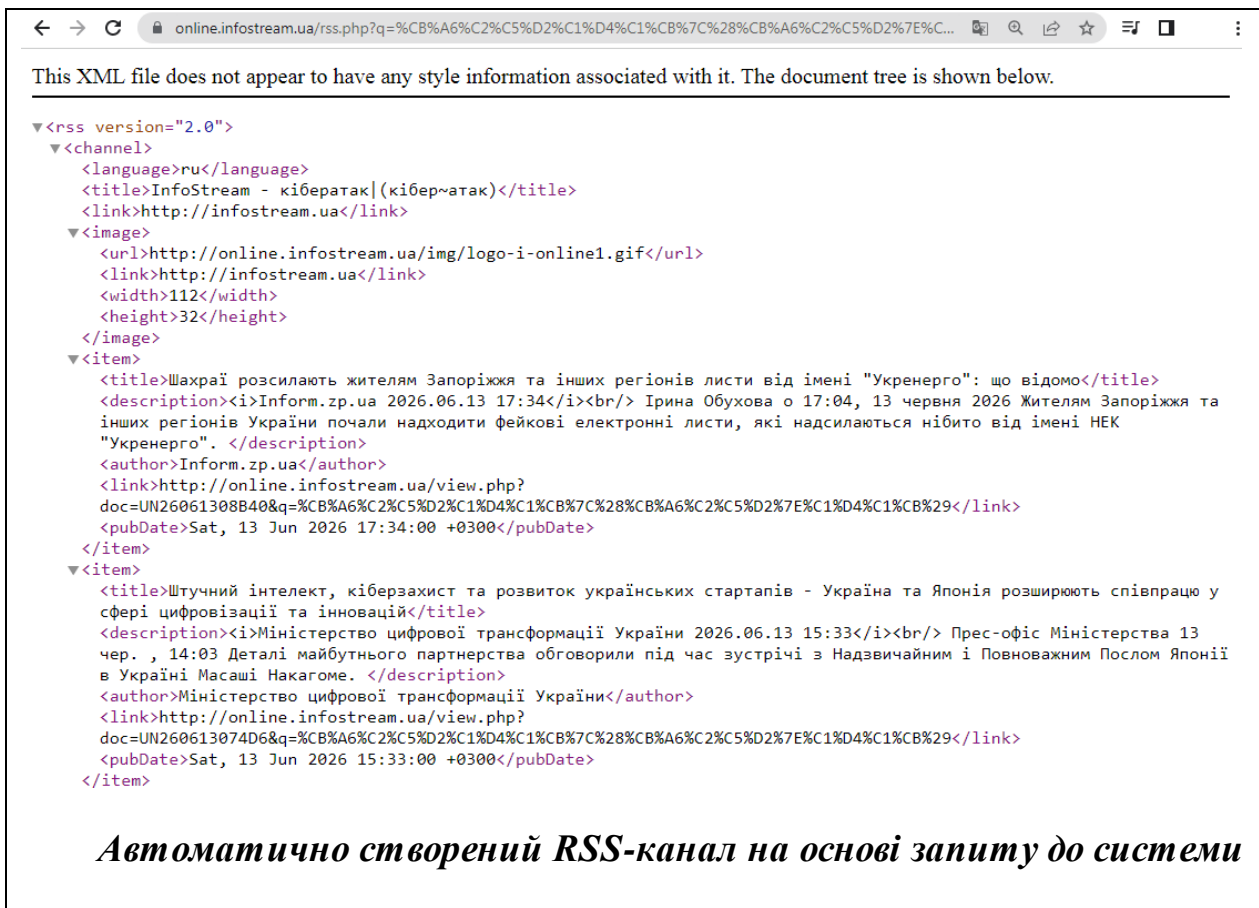
Електронна пошта залишається універсальним інструментом доставки аналітичних матеріалів, особливо в сценаріях, що вимагають періодичного ознайомлення, архівування або маршрутизації довідок у корпоративному середовищі. В InfoStream налаштування email-розсилки базується на збереженому пошуковому запиті, до якого застосовуються фільтри за джерелами, мовами, тематичними рубриками, діапазоном дат та рівнем цифрової насиченості. Періодичність доставки варіюється від пакетного формування (щоденні/щотижневі дайджести) до режиму, наближеного до реального часу, коли матеріали надходять у міру індексації нових документів.

Структура звітів гнучко адаптується до потреб користувача: від базового переліку заголовків з анотаціями до розгорнутих сюжетних ланцюжків, автоматично реферованих дайджестів та RAG-верифікованих аналітичних резюме. Кожен фрагмент звіту супроводжується гіперпосиланням на індексований оригінал, датою публікації та індикатором тональності/достовірності.

### ***RSS-потоки: синдикація та персоналізація стрічок***

Технологія RSS (Really Simple Syndication) реалізована в InfoStream як механізм машинно-зчитуваної трансляції новинних потоків у форматі XML.

Формування потоку відбувається динамічно на основі інформаційно-пошукового запиту InfoReS, який передається як параметр URL-адреси. Це дозволяє аналітикам створювати необмежену кількість тематичних каналів (фідів), кожен з яких оновлюється згідно з графіком індексації системи. RSS-потоки інтегруються з агрегаторами та сучасними браузерами, забезпечуючи кросплатформний доступ без необхідності постійного входу в веб-інтерфейс.



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
  <channel>
    <language>ru</language>
    <title>InfoStream - кібератак|(кібер-атак)</title>
    <link>http://infostream.ua</link>
    <image>
      <url>http://online.infostream.ua/img/logo-i-online1.gif</url>
      <link>http://infostream.ua</link>
      <width>112</width>
      <height>32</height>
    </image>
    <item>
      <title>Шахраї розсилають жителям Запоріжжя та інших регіонів листи від імені "Укренерго": що відомо</title>
      <description><i>Inform.zp.ua 2026.06.13 17:34</i><br/> Ірина Обухова о 17:04, 13 червня 2026 Жителям Запоріжжя та інших регіонів України почали надходити фейкові електронні листи, які надсилаються нібито від імені НЕК "Укренерго". </description>
      <author>Inform.zp.ua</author>
      <link>http://online.infostream.ua/view.php?doc=UN26061308B40&q=%CB%A6%C2%5D2%C1%D4%C1%CB%7C%28%CB%A6%C2%5D2%7E%1%D4%C1%CB%29</link>
      <pubDate>Sat, 13 Jun 2026 17:34:00 +0300</pubDate>
    </item>
    <item>
      <title>Штучний інтелект, кіберзахист та розвиток українських стартапів - Україна та Японія розширюють співпрацю у сфері цифровізації та інновацій</title>
      <description><i>Міністерство цифрової трансформації України 2026.06.13 15:33</i><br/> Прес-офіс Міністерства 13 чер. , 14:03 Деталі майбутнього партнерства обговорили під час зустрічі з Надзвичайним і Повноважним Послом Японії в Україні Масаші Накагоме. </description>
      <author>Міністерство цифрової трансформації України</author>
      <link>http://online.infostream.ua/view.php?doc=UN260613074D6&q=%CB%A6%C2%5D2%C1%D4%C1%CB%7C%28%CB%A6%C2%5D2%7E%1%D4%C1%CB%29</link>
      <pubDate>Sat, 13 Jun 2026 15:33:00 +0300</pubDate>
    </item>
  </channel>
</rss>
```

**Автоматично створений RSS-канал на основі запиту до системи**

## Веб-віджети (WebStream)

Сервіс веб-віджетів реалізує можливість експорту відібраних новинних потоків безпосередньо на зовнішні або внутрішні веб-ресурси організації за допомогою вбудовуваного JavaScript-додатку. Віджети відображають заголовки, анонси або повні тексти публікацій, синхронізуючись з основною базою даних у режимі, наближеному до реального часу.

Архітектура віджетів передбачає розширені можливості модерації: адміністратор порталу може налаштовувати фільтри за рубриками, виключати конкретні джерела, задавати пріоритетність відображення та активувати режим попереднього затвердження публікацій перед їхнім виведенням на сайт. Це особливо важливо в сценаріях внутрішніх інформаційних порталів, де необхідно забезпечити відповідність контенту

корпоративним стандартам та уникнути інформаційного шуму. Візуальне оформлення віджетів адаптується до стилю цільового ресурсу, а механізми кешування та пагінації гарантують стабільну роботу навіть при високому навантаженні.

### 6.3. Кабінет користувача

Кабінет користувача в архітектурі InfoStream виступає центральним інтерфейсно-методологічним вузлом. На відміну від статичних налаштувань, що фіксують лише початкові параметри пошуку, сучасний кабінет реалізує динамічний контур управління інформаційними каналами, де кожен параметр піддається моніторингу, оптимізації та аудиту.

**Кабінет користувача**

Настройки

- Персональні канали (укр)
- Персональні канали (англ)
- Персональні джерела (укр)
- Персональні джерела (англ)
- Статистика
- Останні дії

Персональні канали

- Мониторинг
- Архів
- Антиплагиат
- Правова наука
- Захист персональних даних
- Електронне декларування
- кібератака
- LLM
- ШІ

Видалити

Поновлення бази даних здійснюється кожні 15 хвилин

InfoStream® - моніторинг новин  
© 2000-2026 EVisti Information Center

**Інтерфейс кабінету користувача**

#### ***Управління персональними інформаційними каналами***

Персональний інформаційний канал у InfoStream є збереженою конфігурацією пошуково-аналітичного процесу. Його створення базується на класичному алгоритмі:

1. Формування первинного запиту мовою InfoReS (терми, логічні оператори, контекстна близькість/слідування, дужки).
2. Аналіз результатів через «Інформаційний портрет» (ранжування рубрик, мов, джерел, географії, тональності, розміру повідомлень, цифрової насиченості, ключових слів).

### 3. Збереження конфігурації під унікальною назвою з фіксацією всіх критеріїв.

Керування каналами здійснюється через єдину панель.

Фільтри в InfoStream реалізують багаторівневу селекцію потоку на етапі індексації, видачі та аналітичної обробки. Класичний набір включає:

- Лінгвістичні: морфологія (враховувати/не враховувати), точне входження (I), праві скорочення, нечутливість до регістру.
- Структурні: розмір повідомлень (малий/середній/великий), цифрова насиченість (numb.small/medium/large), відображення/приховування змістовних дублів.
- Джерельно-географічні: рубрики, мови, країни (country.ua тощо), конкретні домени або групи джерел, періодичність оновлення.
- Темпоральні: інтервали дат (формат РРРР.ММ.ДД), календарні вікна для ретроспективи, фільтри «Нових подій».

У сучасному контурі ці фільтри доповнено динамічним уточненням.

Статистичний модуль кабінету трансформує класичні метрики активності у комплексну оцінку якості моніторингу та ефективності аналітичного контуру. До базових показників, закладених у документації розробника, належать:

- кількість переглянутих сторінок та документів;
- кількість унікальних документів;
- дата першого та останнього фіксування події за звітний період.

Усі показники візуалізуються у вигляді інтерактивних дашбордів з можливістю експорту та автоматичного формування щомісячних звітів про ефективність інформаційного каналу.

#### **6.4. Каталог джерел**

Каталог джерел є фундаментальним сервісним модулем InfoStream, що трансформує хаотичний масив відкритих веб-ресурсів у структурований, верифікований та керований реєстр. Він слугує методологічною основою для всіх пошукових, аналітичних та прогнозних операцій системи, гарантуючи прозорість походження даних, контроль якості інформаційного потоку та відтворюваність результатів у контурі систем підтримки прийняття рішень.

**Кабінет користувача: персональні джерела**

Архітектура каталогу базується на ієрархічній рубрикації, що відображає тип редакційної політики та функціональне призначення ресурсу. Джерела групуються за категоріями: інформаційні агентства, телерадіомовні компанії, друковані та онлайн-видання, офіційні державні та регуляторні портали, корпоративні прес-центри, експертні спільноти, тематичні блоги, міжнародні інтегратори та архівні сховища. Кожен запис містить уніфікований набір метаданих: унікальний ідентифікатор, назву, URL, країну реєстрації, мову контенту, періодичність оновлення, дату останнього сканування та індикатор активності.

Каталог підтримує гнучкий механізм фільтрації, який інтегрується в конструктор запитів, «Інформаційний портрет» та налаштування персональних каналів:

- Фільтр за країною: використовує стандартизовані дволітерні коди (country.ua, country.pl, country.us тощо) для географічного сегментування потоку. Дозволяє ізолювати локальні дискурси, відстежувати міжнародне резонування подій або формувати крос-кордонні вибірки.
- Мовний фільтр: забезпечує розділення мовних версій подій, крос-мовний моніторинг та виключення контенту, що не відповідає лінгвістичним параметрам запиту. Важливо для аналізу міжнародних комунікацій та виявлення перекладених маніпулятивних наративів.
- Фільтр за періодичністю: класифікує джерела за частотою оновлення (реал-тайм/щогодинно/щоденно/тижнево/архівно). Цей параметр калібрує очікування щодо оперативності, дозволяє налаштувати пріоритети доставки та автоматично адаптувати інтервали сканування під редакційний цикл ресурсу.

Усі фільтри зберігаються як метадані джерела та використовуються системою для автоматичного зважування результатів пошуку, що забезпечує високу релевантність та мінімізацію тематичного дрейфу.

### *Інтерфейс експлуатації та персоналізація*

Веб-інтерфейс каталогу забезпечує інтуїтивну навігацію та розширені функції управління:

- пошук за назвою, доменом або категорією;
- перегляд повного списку індексованих документів окремого джерела;
- миттєвий перехід до оригінального веб-сайту ресурсу;
- візуальне маркування за країною (кольорові індикатори) та статусом активності;
- додавання нових джерел за запитом користувача.

Користувач має можливість формувати списки «Персональних джерел», комбінувати їх із логічними операторами, зберігати конфігурації як шаблони та автоматизувати доставку контенту без втрати контролю над якістю вхідного потоку.

Каталог джерел InfoStream перетворює збір інформації з пасивного сканування мережі на керований, контрольований та верифікований процес. Поєднання класичної таксономії, моніторингу активності на основі часових рядів, багаторівневої фільтрації та інтеграції з LLM-контуром забезпечує методологічну прозорість, мінімізує ризики роботи з ненадійними ресурсами та формує надійну основу для прийняття обґрунтованих рішень.

The screenshot displays the 'Кабінет користувача' (User Cabinet) interface. At the top, there is a navigation bar with 'Головна' (Home) and 'Система інтеграції ресурсів Інтернету' (Internet Resource Integration System). Below this is a search bar with options for 'Період: 7 днів', 'Вилучити дублі', 'Морфологія', and 'Посторінково'. The main content area is titled 'Кабінет користувача' and includes a sidebar with 'Настройки' (Settings) and 'Останні дії' (Recent Actions). The 'Останні дії' section shows a table of the last 500 events.

Лог останніх 500 подій				
N	Документ / Запит	URL	Час	IP
1.	Кабінет	URL	2026-06-13 18:39:13	176.111.183.104
2.	Кабінет	URL	2026-06-13 18:39:10	176.111.183.104
3.	Кабінет	URL	2026-06-13 18:39:07	176.111.183.104
4.	Дайджест	URL	2026-06-13 18:39:03	176.111.183.104
5.	Сюжети	URL	2026-06-13 18:39:00	176.111.183.104
6.	Документ	URL	2026-06-13 18:38:54	176.111.183.104
7.	фейк	URL	2026-06-13 18:38:51	176.111.183.104
8.	Кабінет	URL	2026-06-13 18:38:35	176.111.183.104
9.	Кабінет	URL	2026-06-13 18:38:32	176.111.183.104
10.	Кабінет	URL	2026-06-13 18:38:30	176.111.183.104
11.	моніторинг (internet   web   інтернет   веб) (разведк   контент   телеком   мобиль   безопасн)	URL	2026-06-13 18:38:28	176.111.183.104

Below the table, the text **Кабінет користувача: Останні дії** is displayed.

## **ВИСНОВКИ**

У цьому науково-методичному посібнику систематизовано теоретичні, технологічні та практичні засади сучасного контент-моніторингу новинних потоків на прикладі системи InfoStream. Викладений матеріал демонструє, як поєднання перевіреної класичної архітектури повнотекстового пошуку, тридцятирічного ретроспективного фонду та сучасних аналітичних модулів на базі великих мовних моделей та методів аналізу часових рядів трансформує моніторинг з інструменту пасивного збору даних у активний компонент інтелектуальної підтримки прийняття рішень.

### ***1. Симбіоз класичних алгоритмів та ШІ-аналітики***

Основним методологічним висновком є доцільність поєднання детермінованих пошукових механізмів із генеративними моделями. Формалізована мова запитів InfoReS, конструктор пошукових сценаріїв, механізми контекстного слідування та близькості, а також автоматична виявлення змістовних дублів забезпечують контрольовану, відтворювану та семантично валідну первинну вибірку. Саме ця класична основа стає надійним фільтром для подальшої обробки LLM-модулями, які виконують семантичний кластерний аналіз, екстрагування сутностей та автоматичне реферування.

### ***2. Верифікація, відтворюваність та архітектура RAG***

Інтеграція Retrieval-Augmented Generation вирішує фундаментальну проблему довіри до ШІ-генерацій у моніторингових системах. Обов'язкова прив'язка кожного аналітичного висновку, резюме сюжету чи прогнозу до першоджерела, дати публікації та оригінального фрагменту робить результати придатними для офіційних довідок, експертизи та звітності. Поєднання цього підходу з лінгво-статистичними алгоритмами фільтрації забезпечує надійне відокремлення природного інформаційного циклу від штучного втручання, інформаційних операцій, атак.

### ***3. Аналіз часових рядів та двокомпонентне прогнозування***

Перехід від ретроспективного опису до превентивного моделювання реалізується через інтеграцію класичних гістограм динаміки понять із сучасними методами виявлення аномалій, хвильового моделювання та прогнозування. Виявлення структурних зламів, циклічності медіауваги, латентних трендів та кореляцій між рубриками дозволяє не лише фіксувати події, а й оцінювати їх потенціал, точки кристалізації нових сюжетів та

ймовірні сценарії розвитку. Динамічний прогноз (обсяги тематичних інформаційних потоків) та семантичний прогноз (розвиток дискурсу за допомогою LLM) формують єдиний аналітичний контур, орієнтований на прийняття обґрунтованих рішень на випередження.

#### ***4. Інтеграція в контур СППР та автоматизація процесів***

Технологічна цінність моніторингу розкривається повною мірою лише при його вбудовуванні в корпоративні інформаційні інфраструктури. Автоматизація процедур від пошуку до генерації структурованих довідок скорочує цикл «виявлення → аналіз → рішення → дія», мінімізує людський фактор на рутинних етапах та забезпечує стандартизацію вихідних документів без втрати методологічної прозорості.

#### ***5. Методологічна прозорість, навчання та аудит якості***

Ефективне використання системи вимагає диференційованого підходу до формування компетенцій та налагодження процедур верифікації. Поєднання базових принципів конструювання запитів, просунутих знань у налагодженні LLM-промптів та навичок інтерпретації часових рядів і прогнозних моделей формує повноцінний аналітичний цикл, де кожний висновок підлягає обґрунтуванню. Регулярний аудит релевантності, повноти, оперативності та точності ШІ-генерацій гарантує стабільність системи при зміні медіа-дискурсу чи появи нових змістовних патернів.

Сучасний моніторинг новинних потоків – це не ізольована технологічна утиліта, а цілісна інформаційно-аналітична інфраструктура. Симбіоз класичної точності, верифікованої ШІ-аналітики, глибини ретроспективного фонду та інтеграції з СППР забезпечує організаціям не лише оперативну обізнаність, а й стратегічну стійкість у умовах високої динаміки інформаційного простору. Описані методичні підходи, алгоритми та практичні сценарії можуть бути адаптовані до різних галузей, формуючи універсальну основу для розвитку професійних компетенцій у сфері цифрової аналітики, інформаційної безпеки, медіа-досліджень та підтримки управлінських рішень нового покоління.

## ДОДАТОК А. Довідник мови запитів InfoReS

Мова запитів інформаційно-пошукової системи InfoReS є формалізованим інструментом, що забезпечує точне, відтворюване та багаторівневе конструювання інформаційних каналів у системі InfoStream. Нижче наведено структурований методичний довідник синтаксичних правил, таблицю пріоритетів операторів, опис модифікаторів та приклади їх практичного застосування для аналітичних та управлінських задач.

### 1. Базові правила синтаксису

<i>Правило</i>	<i>Опис</i>	<i>Приклад</i>
Регістр	Система не чутлива до регістру літер	банк = Банк = БАНК
Терми	Слова природної мови або їх праві скорочення (мінімум 2 літери)	інформ, монітор
За замовчуванням	Пошук за правою основою слова (автоматичне відкидання закінчень). Зірочка * залишається для сумісності з іншими системами, але може не використовуватись	завод → знайде: завод, заводи, заводу, заводський
Точне входження	Додавання символу ] після слова вимикає морфологічне розширення	фон] → знайде тільки фон, але не фонди, фоновий
Дужки	Використовуються для групування термінів, зміни пріоритетів та побудови багаторівневих конструкцій	(кава чай)&напій

### 2. Пріоритети операторів

Оператори обробляються у такій послідовності (від найвищого пріоритету до найнижчого):

<i>Пріоритет</i>	<i>Символ / Назва</i>	<i>Функція</i>	<i>Обмеження</i>
1	~, ADJ	Контекстне слідування (порядок важливий)	Макс. 3 операнди; беруться в дужки. Операндом не може бути вираз у дужках
2	@	Контекстна близькість (порядок не важливий)	Аналогічно ~
3	!, ^, NOT	Логічне НІ (I-NI)	Виключає документи з указаним термом
4	&, +, пробіл, AND	Логічне І	Вимагає наявності всіх операндів
5	, кома, OR	Логічне АБО	Вимагає наявності всіх операндів

Примітка: Відстань між термами у контекстних операторах задається конструкцією /n/, де n – максимальна кількість слів між ними (відсутність параметра) означає сусідні слова.

### 3. Спеціальні конструкції та фільтри метаданих

Окрім текстових термів, мова запитів підтримує вбудовані директиви для фільтрації за структурованими метаданими:

<i>Тип фільтра</i>	<i>Синтаксис</i>	<i>Приклад</i>
Дати	PPPP.MM.ДД або праві скорочення	2026.05.12 (точна дата) 2026.05. (весь травень) 2026.05.2 (20–29 травня)
Країни джерел	country.XX (ISO-код)	country.ua, country.us, country.de
Джерела (домени)	Частина доменного імена або повна адреса	elvisti.com, 4vlada.net, www.lenta
Розмір документа	leng.large, leng.medium, leng.small	leng.medium (1000-10 000 символів)
Тематичні рубрики	rubrXX (внутрішній код класифікатора)	rubr02 (фінанси/банки), rubr05 (енергетика)
Цифрова насиченість	numb.large, numb.medium, numb.small	numb.large (>10% цифрових символів)
Морфологія	Перемикач в інтерфейсі (враховувати / не враховувати)	При вимкненні – пошук строго за правою основою
Дублі / Подібні	Керуються прапорцями інтерфейсу або API	показувати дублі / не показувати дублі

### 4. Приклади складних конструкцій

#### **Приклад 1. Галузевий моніторинг з виключенням комерційного шуму**

(енергет | електрич | генераці) & (тариф | ціна | ринок | регулюван) ! (акці | знижк | подарунок | розпродаж)

Мета: відібрати аналітичні та новинні матеріали про енергетичний ринок, виключивши рекламні публікації.

#### **Приклад 2. Відстеження законодавчої ініціативи з контекстною близькістю**

(законопроект~ /2/ цифров | закон~ /1/ діджитал) & (парламент | комітет | голосуван | читанн) & country.ua & 2026.05.

Мета: знайти матеріали про ухвалення або обговорення цифрових законів у травні 2026 р., де ключові терми розташовані поруч, що підвищує семантичну точність.

### ***Приклад 3. Точний моніторинг суб'єкта з фільтрацією за насиченістю***

("Національний банк України" | "НБУ" | нбу) & (ставка | інфляці | ліцензі | штраф) & numb.medium

Мета: відокремити офіційні комунікації та аналітичні огляди про НБУ, мінімізуючи короткі новинні замітки та прес-релізи з низькою інформаційною щільністю.

### ***Приклад 4. Багаторівневий кризовий сценарій***

((авар | пожеж | відключ | затримк) & (енергет | інфраструктур | логістик)) | ((санкці | конфіск | розслідуван) & (фінанс | банк | компанії)) & ! (реклам | ваканс | тендер)

Мета: сформувати канал раннього попередження про інфраструктурні інциденти або регуляторні тиски на фінансовий сектор, виключаючи нерелевантні комерційні події.

### ***5. Методичні рекомендації щодо конструювання запитів***

- Принцип «від загального до конкретного»: починати з широкої булевої конструкції, після чого звужувати вибірку через NOT, контекстні оператори та фільтри метаданих.
- Контроль семантичного дрейфу: використовувати ~ та @ із явно заданою відстанню (/1/, /2/) для утримання термінів у межах одного смислового блоку.
- Верифікація через Інформаційний портрет: після первинного пошуку аналізувати ранжировані характеристики вибірки (рубрики, джерела, географія, тональність) та інтерактивно коригувати запит через AND/NOT-маркери портрета.
- Збереження конфігурацій: успішні конструкції зберігати як персональні інформаційні канали. Параметри записуються разом із налаштуваннями морфології, фільтрами дублів та діапазоном дат, що гарантує повну відтворюваність результатів.
- Уникнення надмірної складності: уникайте вкладеності дужок понад 3 рівні. У разі потреби розбийте запит зі складною логікою на декілька послідовних запитів або використайте режим «Сюжети» / «Дайджести» для автоматичного групування.

## ДОДАТОК Б. Глосарій термінів та скорочень

**API (Application Programming Interface)** – програмний інтерфейс застосунків, що забезпечує стандартизовану взаємодію між компонентами програмного забезпечення або зовнішніми системами.

**Дайджест** – автоматично сформований структурований звіт, що містить найбільш інформативно насичені фрагменти новинного масиву. Будується на основі частотно-статистичного реферування та LLM-синтезу, з посиланням на джерело, дату та оригінальний уривок.

**Дедублікація (детекція змістовних дублів)** – алгоритмічний процес виявлення та маркування документів, ідентичних або семантично близьких за змістом. Забезпечує мінімізацію інформаційного шуму та підвищення співвідношення «сигнал/шум» у релевантній вибірці.

**Grounding (верифікація III)** – процедура обов'язкової прив'язки кожного згенерованого штучним інтелектом висновку до індексованого першоджерела, дати публікації та оригінального тексту.

**InfoReS (IIC InfoReS)** – повнотекстова інформаційно-пошукова система, програмно-технологічне ядро InfoStream. Забезпечує індексацію, морфологічну обробку, контекстний пошук, булеву логіку, фільтрацію метаданих та автоматичну детекцію дублів.

**InfoStream** – система контент-моніторингу новин з Інтернету та соціальних мереж, що поєднує класичні алгоритми інформаційного пошуку, ретроспективний фонд (понад 30 років) та сучасні аналітичні модулі (LLM, RAG, часові ряди) для підтримки прийняття управлінських, комунікаційних та наукових рішень.

**Класифікатор-навігатор** – інтерактивний інструмент динамічного уточнення пошуку, що реалізує принцип «пошукових папок». Дозволяє кластеризувати результати за семантичною близькістю без повторного формулювання запиту, автоматично адаптуючись до лексичного складу вибірки.

**LLM (Large Language Model / Велика мовна модель)** – нейронна мережа, навчена на масштабних текстових корпусах, здатна виконувати завдання розуміння мови, генерації, реферування та витягнення сутностей. В InfoStream використовується виключно на попередньо відібраній та фільтрованій вибірці, що гарантує стабільність результатів.

**Metadata / Метадані** – структуровані описові дані про документ: URL, джерело, дата/час публікації та індексації, мова, країна, рубрика, витягнені сутності, ідентифікатор дубля, контрольні суми. Дозволяють відстежити джерела даних, провести аудит і повністю відтворити хід аналізу для обґрунтування висновків..

**NLP (Natural Language Processing / Обробка природної мови)** – розділ штучного інтелекту та комп'ютерної лінгвістики, що вивчає взаємодію між машинами та людською мовою. У системі охоплює морфологічний аналіз, визначення тональності, NER (розпізнавання іменованих сутностей), семантичну кластеризацію та частотно-статистичне зважування термінів.

**OSINT (Open Source Intelligence)** – розвідка та аналітика на основі відкритих джерел. InfoStream реалізує технологічний контур OSINT для безперервного сканування веб-ресурсів, офіційних порталів, експертних блогів та соціальних мереж.

**Precision (Точність)** – метрика оцінки релевантності пошукової або аналітичної вибірки, що визначає частку дійсно релевантних документів серед усіх матеріалів, відібраних системою за заданим запитом чи критерієм фільтрації. Розраховується як відношення кількості правильно відібраних релевантних документів (істинно позитивних результатів) до загальної кількості відібраних документів (включаючи хибно позитивні, тобто тематично невідповідні або шумові публікації).

**RAG (Retrieval-Augmented Generation / Генерація з розширенням пошуку)** – архітектурний підхід, при якому генерація тексту III відбувається на основі контексту, отриманого з індексованої бази даних. Забезпечує цитування джерел, мінімізує ризик неперевірених тверджень та робить результати придатними для офіційних довідок.

**Recall (Повнота)** – метрика оцінки охоплення, що визначає частку релевантних документів, які система здатна знайти, від загальної кількості таких документів у джерельній базі. Залежить від широти сканування, налаштування морфології та якості каталогу джерел.

**Ретроспективний фонд** – архів новинних публікацій, що безперервно формується в InfoStream понад 30 років. Забезпечує глибину для досліджень у часі, калібрування аналітичних моделей, верифікації історичних зрізів та збереження матеріалів, видалених з оригінальних веб-сайтів.

**Семантична кластеризація** – автоматичне групування документів за смисловою близькістю, а не лише за лексичним збігом ключових слів. Реалізується за допомогою векторних представлень та LLM, що дозволяє виявляти приховані тематичні зв'язки та латентні тренди.

**Сюжетні ланцюжки (Основні сюжети)** – автоматично сформована послідовність публікацій, об'єднаних спільною темою, хронологією та причинно-наслідковими зв'язками. Ранжується за кількістю повідомлень, часом публікації та медіа-резонансом.

**СППР (Система підтримки прийняття рішень / Decision Support System)** – інформаційна система, що інтегрує дані, аналітичні моделі та інтерфейси для допомоги фахівцям у прийнятті управлінських, комунікаційних або наукових рішень. InfoStream виступає аналітичним та

інформаційним контуром СППР, трансформуючи сирий новинний потік у верифіковані інсайти.

**Text Mining (Глибинний аналіз текстів)** – сукупність методів автоматичного виявлення структури, патернів та знань у неструктурованих текстових масивах. Включає частотно-статистичний аналіз, кластеризацію, витягнення сутностей, побудову онтологічних схем та автоматичне реферування.

**Тональність (Sentiment Analysis)** – оцінка емоційного забарвлення тексту або ступеня підтримки/критики щодо конкретного об'єкта (позитивна, нейтральна, негативна, агресивна, маніпулятивна). Визначається за допомогою NLP-моделей та LLM, використовується в інформаційних портретах та хвильовому аналізі.

**Часові ряди (Time Series)** – послідовність даних, впорядкованих у часі, що відображають динаміку появи публікацій, частотність понять, зміни тональності або активність джерел.

**Хвильовий аналіз / Аналіз інформаційних хвиль** – метод виявлення патернів зростання, піку та затухання медіа-уваги до події. Дозволяє відрізнити природний життєвий цикл інформаційного імпульсу від штучного нагнітання, координованих наративів або бот-активності.

**Частотно-статистичний метод** – алгоритмічна основа інформаційних портретів та дайджестів, що базується на розрахунку ваги термінів, речень або документів у межах вибірки. Використовується для автоматичного ранжування, фільтрації та побудови структурованих резюме.

**Інформаційний портрет** – візуально-аналітична модель релевантної вибірки, що відображає ранжовані характеристики: рубрики, мови, джерела, географію, тональність, розмір документів, цифрову насиченість та ключові терміни. Використовується для динамічного уточнення запиту через AND/NOT-маркери.]

**Прогнозування (динамічне / семантичне)** – двокомпонентна модель передбачення інформаційних процесів, у якій динамічний компонент оцінює кількісні зміни новинного потоку, ймовірність медіаспалахів або затухання тем на основі статистичного аналізу часових рядів, а семантичний компонент моделює якісний розвиток дискурсу, виявляє точки кристалізації нових подій та прогнозує ймовірні реакції в регуляторному, експертному та медіа-середовищі із використанням великих мовних моделей; їхнє поєднання трансформує контент-моніторинг з інструменту реактивного фіксування фактів у систему випереджувального аналізу, придатну для сценарного планування та обґрунтованої підтримки управлінських рішень.