

*Фурсова Н.А., ст. викл.,
Кочережко С.С., студент
Полтавський національний технічний університет
імені Юрія Кондратюка*

АРХІТЕКТУРА ПРОГРАМНОЇ СИСТЕМИ ЗБОРУ ТА ОЦІНЮВАННЯ РЕЛЕВАНТНОСТІ ПУБЛІКАЦІЙ В МЕРЕЖІ ІНТЕРНЕТ

У статті представлено архітектуру програмної системи збору та оцінювання релевантності публікацій в мережі Інтернет. Розкрито функції та принципи роботи основних компонентів архітектури системи.

Ключові слова: програмна система, релевантна публікація, мікросервіс.

Аналіз досліджень релевантності та впливу публікацій в соціальних мережах висуває сьогодні нові вимоги – нестандартний набір даних, обмеження публічного прикладного програмного інтерфейсу та непридатність застосування механізму крос-посилань, який використовується пошуковим сервісом Google. У сучасному світі широка аудиторія користувачів створила з соціальних мереж ключового гравця на інформаційному ринку. Разом з цим проблема визначення релевантних публікацій, які можуть потенційно вплинути на широку аудиторію стрімко набуває актуальності.

Встановлено, що релевантність – показник, який визначає придатність отриманих даних зазначеному предметній області у пошуку, а потенційний вплив вказує на охоплення аудиторії у соціальних мережах певною публікацією. Ці показники перебувають у нерозривному зв'язку, оскільки нерелевантна для користувача інформація не здійснює значного впливу на цільову аудиторію [1]. Зазначимо, що проблемі кількісної оцінки активності користувачів присвячено змістовне дослідження Якоба Нільсена “Participation Inequality: Encouraging

More Users to Contribute” [2], де визначено, що 90% публікацій надходить від 1% відсотка користувачів. Аннаук Вельхьойз у своїй роботі “Twitter and Pareto Principle”[3] визначив, що 20% користувачів з тестової вибірки генерують 84% контенту, а інші публікації створюються малоактивною групою користувачів, що налічує 9 % від загальної кількості.

Враховуючи стрімке поширення інформації та соціальну активність користувачів, розроблено програмну систему «Sonar Service» для збору та оцінювання релевантності публікацій в мережі Інтернет. Система оцінювання інтернет-публікацій опрацьовує дві групи публікацій – публікації з видачі соціальних мереж та публікацій з інформаційної видачі пошукових движків.

Виходячи із цього, отримано дві основоположні підсистеми – підсистему оцінювання публікацій з видачі пошукових движків та підсистему роботи з API (Application Programming Interface) соціальних мереж. Оскільки блоги є одним із пріоритетних джерел отримання публікацій, їх враховано в якості окремої підсистеми.

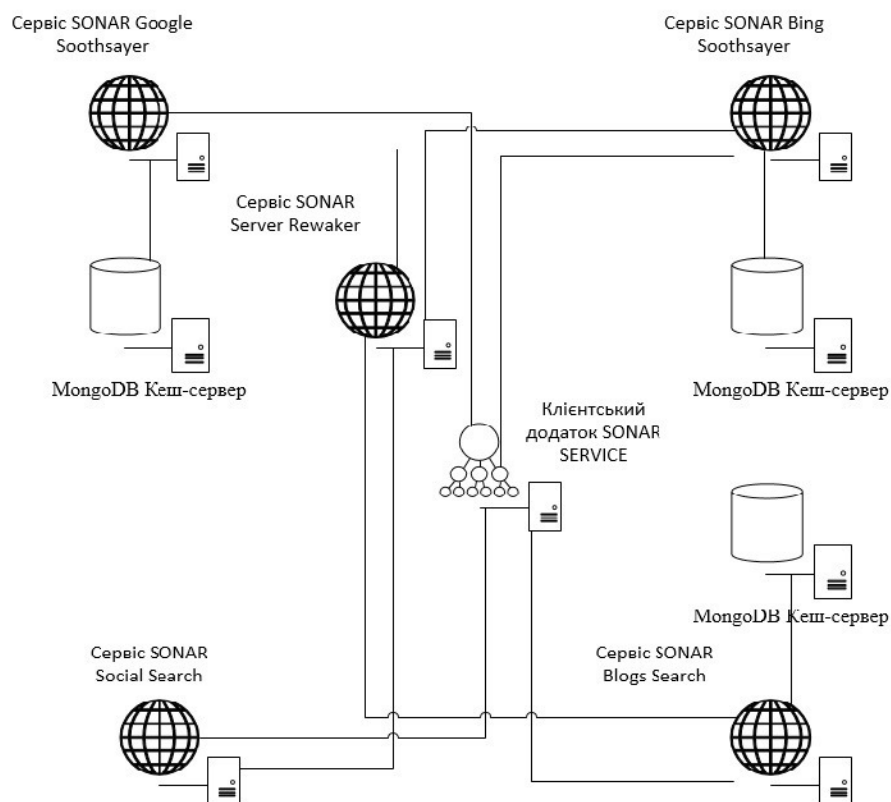


Рис. 1 – Архітектура програмної системи «Sonar Service»

Для опрацювання пошукової видачі публікацій із потенційно можливих джерел спроектовано архітектуру програмної системи «Sonar Service», що містить мікросервіси «SONAR Google Soothsayer», «SONAR Bing Soothsayer», «SONAR Social Search», «SONAR Blogs Search» (рис. 1).

Мікросервіси програмної системи «Sonar Service» реалізовано на мові програмування JavaScript стандарту ECMAScript 2015 з використанням можливостей платформи Node.js версії 6.1.0 та фреймворку Express.js. В якості вхідного повідомлення мікросервіси використовують JSON-об'єкт, стандартний для всіх мікросервісів програмної системи SONAR. В результаті проведення аналізу мікросервіси генерують масив JSON-об'єктів, що репрезентуює окрему релевантну та впливову публікацію з видачі пошукового сервісу Google.

Функція мікросервісу «SONAR Google Soothsayer» полягає у обробці пошукової видачі сервісу Google. Для проведення аналізу клієнтський додаток з'єднується з мікросервісом за DDP-протоколом, викликаючи метод «google».

Першим етапом функціонування програмної системи є отримання релевантних публікацій – вилучення посилань на публікації з пошукової видачі Google. На другому етапі, відповідно до кожного посилання, проводиться перевірка публікацій на їх наявність у базі даних. У випадку повторного проведення аналізу, результати вилучаються з бази даних і конвертуються у стандартизований формат елементів видачі. Діаграма активності на рис. 2 описує принципи роботи мікросервісу SONAR Google Soothsayer.

Мікросервіс «SONAR Bing Soothsayer» відповідає також за обробку пошукової видачі, однак пошукового сервісу Bing від Microsoft. В результаті проведення аналізу мікросервіс генерує масив JSON-об'єктів, що репрезентуює окрему релевантну та впливову публікацію з видачі пошукового сервісу Google. Для проведення аналізу клієнтський додаток з'єднується з мікросервісом за DDP-протоколом та викликає метод «bing».

Мікросервіс «SONAR Blogs Search» відповідає за знаходження релевантних та впливових публікацій, розміщених на популярних блог-платформах – Medium, Blogger, Ghost, Postagon, Posthaven та Postach.io. Для

проведення аналізу клієнтський додаток під'єднується до мікросервісу за DDP-протоколом та викликає один з методів-аналізаторів – «blogger», «medium», «ghost», «postagon», «posthaven» чи «postach», передаючи йому в якості аргументу стандартний JSON-об'єкт.

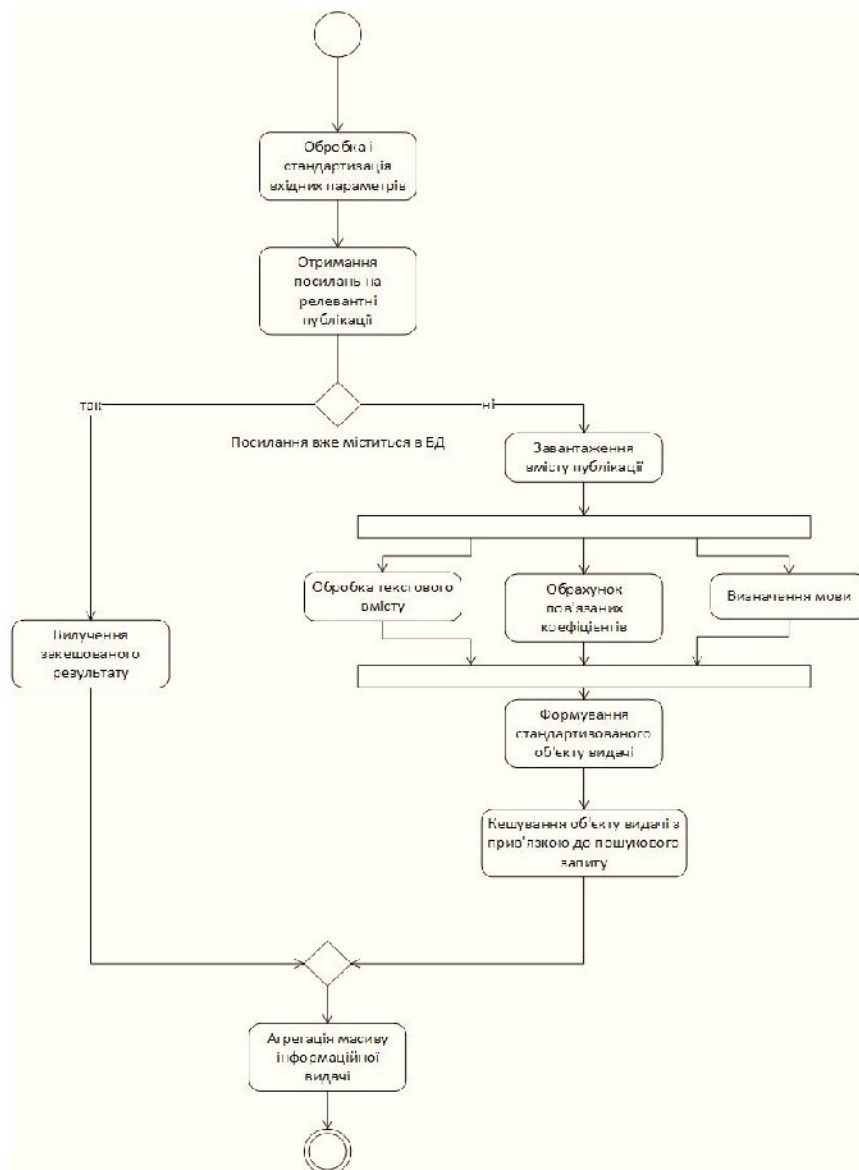


Рис. 2 – Діаграма активності мікросервісу «SONAR Google Soothsayer»

Мікросервіс «SONAR Social Search» відповідає за обробку та оцінювання публікацій, розміщених у соціальних мережах Facebook, Twitter та Google+. Для проведення аналізу клієнтський додаток з'єднується з мікросервісом за допомогою DDP-протоколу та викликає один з трьох методів – «facebook», «twitter» чи «gplus», які репрезентують соціальні мережі Facebook, Twitter та Google+, відповідно передаючи йому в якості аргументу стандартний JSON-

об'єкт. Під час аналізу публікацій із соціальної мережі Facebook система знаходить необхідні тематичні джерела – сторінки та групи користувачів, розмежовуючи їх на неактивні та нерелевантні. Джерела, які пройшли фільтрацію, попадають під вилучення публікацій зі стрічки новин. Кількість публікацій до вилучення регламентується 3 місяцями, у випадку повного сканування - 24 годинами.

Логіка роботи «SONAR Social Search» відрізняється від роботи мікросервісів «SONAR Google Soothsayer», «SONAR Bing Soothsayer», «SONAR Blogs Search», за рахунок одержання посилання на релевантні канали та публікації саме від публічних API соціальних мереж, не враховуючи проміжні запити до пошукових сервісів. Цей мікросервіс не підтримує кешування, швидка зміна даних робить це недоцільним.

Архітектура програмної системи оцінювання релевантності публікацій представлена сукупністю окремих спеціалізованих мікросервісів, призначених для обробки кожного джерела публікацій. Компоненти системи містять базу даних, що зберігає попередні результати аналізу, нівелює затримки роботи системи та скорочує часові витрати на завантаження тих публікацій, які вже опрацьовані. Таким чином, програмна система оцінювання впливу публікацій може бути використана для реалізації бренд-моніторингу та ранжування інформаційного потоку у мережі Інтернет.

Література

1. Фурсова Н.А., Кочережко С. Чинники зростання впливу інтернет-публікацій в соціальних мережах на цільову аудиторію. *Інформаційні технології: теорія, інновації, практика: матеріали Міжнародної науково-практичної Інтернет-конференції, 15 – 18 грудня 2015 року / Полтавський національний технічний університет імені Юрія Кондратюка. – Полтава. – 2015. – 121-122 с.*
2. Nielsen, Jakob. "Participation Inequality: Encouraging More Users to Contribute" (2006). [Електронний ресурс] - Режим доступу: http://www.useit.com/alertbox/participation_inequality.html
3. Welhuis, Annouck. "Twitter and Pareto Principle", (2012). [Електронний ресурс] - Режим доступу: <http://www.annouckwelhuis.nl/twitter-and-the-pareto-principle-2/>