

СИСТЕМА АГРЕГАЦІЇ СОЦІАЛЬНИХ МЕДІА З ПИТАНЬ КІБЕРБЕЗПЕКИ

Д. В. Ланде¹, Н. А. Кальян¹, О. Т. Матіїшин¹

¹Національний технічний університет України,
«Київський політехнічний інститут імені Ігоря Сікорського»,
Інститут спеціального зв'язку і захисту інформації

Анотація

В роботі наведено архітектуру корпоративної інформаційно-аналітичної системи контент-моніторингу та агрегування інформації із соціальних медіа. На базі технології OSINT побудовано діючий макет системи із максимальним застосуванням компонент відкритого доступу. Запропоновані сервіси контент-моніторингу і агрегування інформації із соціальних мереж містять сучасні можливості персоналізації, надання доступу до баз даних в режимі он-лайн, у тому числі з мобільних пристроїв, для чого широко застосовуються можливості форматів RSS.

Ключові слова: розділи, стаття, тези

Вступ

На цей час при підготовці рішень на багатьох рівнях керування виникають питання урахування інформації, що з'являється у соціальних медіа. Зокрема, відомо, що спеціальні інформаційні потоки є компонентом інформаційної зброї, спрямованої на реалізацію попередньо спланованих інформаційно-психологічних впливів на аудиторію для досягнення заздалегідь визначених цілей [1]. Інформація із соціальних мереж, з одного боку, містить багато «шуму», і навіть дезінформації, а, з іншого боку, є самою оперативною (повідомлення в мережі Twitter у середньому на 6 годин випереджують повідомлення на веб-сайтах). Спроби розробки технологій агрегації інформації із соціальних мереж здійснюються вже близько десятиріччя, але гальмуються великою вартістю відповідних систем, обчислювальними труднощами, особливо в разі потреби опису динаміки даних великих обсягів [2]. Звичайно, на цей час існує досить багато можливостей для добування із соціальних мереж і комп'ютерної обробки даних, що дозволяє, з одного боку, готувати набори вхідних даних на підставі аналізу результатів досліджень інформаційного простору, а з іншого – вирішувати формалізовані завдання з достатнім ступенем точності і в допустимий час. Однак сьогодні серверні рішення і в більшості випадків не доступні для придбання, не існує також бюджетних рішень на основі OSINT (Open-Source Intelligence) [3] для обробки інформації із соціальних мереж в межах корпорації.

На цей час залишається актуальним створення технологічних засад та визначення інструментальних засобів контент-моніторингу і агрегування інформації із соціальних мереж (Social Media Aggregation – SMA) на базі технології OSINT, побудова діючого макету корпоративної інформаційно-аналітичної системи агрегування інформації із соціальних медіа на

основі моніторингу соціальних мереж із максимальним застосуванням компонент відкритого доступу.

Новим рішенням, що пропонується у цій роботі для розв'язання проблеми створення корпоративної системи контент-моніторингу і агрегування інформації із соціальних медіа, є архітектура системи, до складу яких входять лише загальнодоступні компоненти інформаційного пошуку і агрегування інформаційних потоків, реалізовані як програми з відкритим кодом.

При створенні системи контент-моніторингу необхідно вирішити низку задач:

- 1) Аналіз існуючих підходів до агрегації тематичних потоків.
- 2) Запропонувати та обґрунтувати підходи до побудови корпоративної системи контент-моніторингу соціальних мереж.
- 3) Створити комплекс інструментальних засобів збирання інформації із соціальних мереж з вибраних замовником питань.
- 4) Адаптувати серверні додатки, що реалізують функції пошуку і аналізу інформації.
- 5) Адаптувати додатки користувачів, що реалізують персоналізацію інформаційного забезпечення.

1. Розв'язок поставлених задач

Сучасні методи агрегування і контент-моніторингу інформації – це адаптація концепції глибинного аналізу текстів (Text Mining) і класичних методів контент-аналізу до умов формування і розвитку динамічних інформаційних масивів, зокрема, потоків інформації в мережі Інтернет. Соціальні медіа являють собою сукупність онлайн-сервісів та інтернет-додатків, які дозволяють користувачам спілкуватися один з одним в тому числі, і в режимі реального часу.

Функціонування систем, що розглядаються, містить такі кроки:

- 1) знаходження релевантних загальній широкій тематиці повідомлень із соціальних мереж – формування інформаційного потоку із тематичних повідомлень;
- 2) завантаження отриманого потоку у повнотекстові бази даних;
- 3) налаштування інформаційно-пошукової системи на сервері на вузькі тематичні запити окремих користувачів, формування запитів на формування динамічних вузько тематичних інформаційних потоків, зокрема, у форматі синдикації новин RSS;
- 4) підключення і застосування систем аналізу і прогнозування динаміки інформаційних потоків;
- 5) підключення і застосування систем агрегації інформаційних каналів на робочих місцях користувачів.

Схема взаємозв'язку компонент системи контент-моніторингу, що розглядається, складається із трьох головних частин – системного забезпечення, ядра системи і додатків користувачів (див. рис 1). До системного забезпечення входять:

- Апаратне забезпечення (сервер, телекомунікаційне обладнання);
- Операційна система (FreeBSD);
- Мови програмування і відповідні бібліотеки (Shell, PHP, JavaScript, Perl, R);
- Веб-сервер (Apache).

До ядра системи входять:

- Засоби збирання даних із соціальних мереж – на діючому макеті реалізовані засоби збирання із таких соціальних мереж: Twitter (сервіс мікроблогів), Youtube (відеосервіс, текстовий супровід), Telegram (месенджер).
- Засоби створення та супроводження баз даних (на діючому макеті реалізовані бази даних, що дозволяють зберігати повнотекстову інформацію у вигляді XML-файлів).
- Засоби повнотекстового пошуку і генерації потоків відповідей у форматі RSS (на діючому макеті система Sphinx Search і написані у рамках цього проекту програмні засоби генерації даних у форматі RSS).
- Засоби аналітики і прогнозу, що базуються на дослідженні мереж, статистики/динаміки тематичних інформаційних потоків (оригінальні розробки, Gephi, Matlab). До додатків користувачів входять:
- Веб-браузери із встановленими плагінами для роботи із даними у форматі RSS.
- RSS-агрегатори (наприклад, FeedDemon 3.5, FeedReader 3.14, RSS Guard 3.4.1), що реалізують можливості доступу до баз даних системи контент-моніторингу та агрегування інформації із соціальних мереж, а також можливість персоналізації (ведення персональних баз даних).
- Офісні додатки, систем ком'ютерної математики, інші прикладні застосування.

За результатами аналізу сучасного стану систем контент-моніторингу соціальних мереж за вибраною предметною областю було встановлено, що впровадження повнофункціональної корпоративної системи такого типу – складна і витратна проблема. Одно-

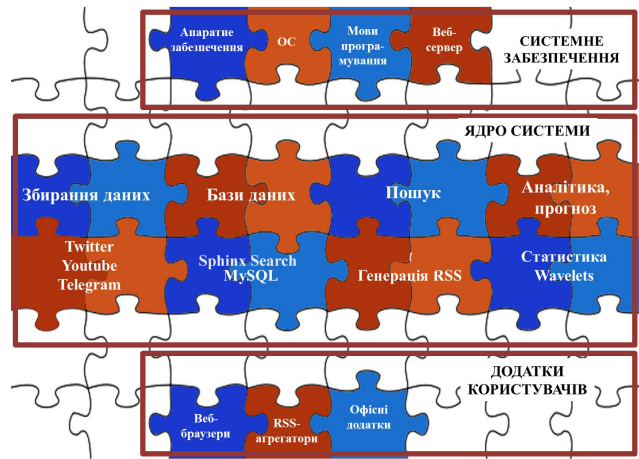


Рис. 1. Основні компоненти системи контент-моніторингу та агрегування інформації із соціальних медіа

часно існує декілька готових до застосування компонентів, можливість створення власних компонентів на базі існуючих бібліотек, системна інтеграція яких дозволяє створити повнофункціональну діючу модель системи контент-моніторингу соціальних мереж. Окремі компоненти можуть бути сполучені між собою, завдяки чому реалізуються функціональні можливості контент-моніторингу у повному обсязі.

Висновки

Запропоновані сервіси побудовані на базі OSINT, сервіси контент-моніторингу і агрегування інформації із соціальних мереж містять сучасні можливості персоналізації, надання доступу до баз даних в режимі online, у тому числі з мобільних пристроїв, для чого широко застосовуються можливості форматів RSS. Обґрунтовано вибір «готових» програмних компонентів, описані засоби власної розробки (сканери соціальних мереж, засоби формування динамічних RSS-каналів), наведені результати їх інтеграції у єдиний програмно-апаратний комплекс.

Практичне значення отриманих результатів полягає в створенні діючого макету системи контент-моніторингу і агрегування інформації із соціальних мереж, готової до застосування компоненти систем підтримки прийняття рішень, зокрема, щодо інформаційної і кібернетичної безпеки.

Перелік використаних джерел

1. Додонов А. Г., Ландэ Д. В., Цыганок В. В. Распознавание информационных операций. — К.: Инжиниринг, 2017. — 282 с. — ISBN: 978-966-2344-60-8.
2. Конкурентная разведка в компьютерных сетях / А. Г. Додонов, Д. В. Ландэ, В. В. Прищеп, В. Г. Путятин. — К.: ИПРИ НАН Украины, 2013. — 248 с. — ISBN: 978-966-2944-96-9.
3. Додонов А. Г., Ландэ Д. В., Путятин В. Г. Конкурентная разведка в компьютерных сетях // Реєстрація, зберігання і обробка даних. Щорічна підсумкова наукова конференція 17 – 18 травня 2018 року. — 2018. — С. 110–112.