Інформатизація та інформаційні технології

УДК 004.5

ЛАНДЕ Д.В., *доктор технічних наук,*
БОЙЧЕНКО А.В., Інститут проблем реєстрації інформації НАН України

СЦЕНАРНИЙ ПІДХІД ПРИ ДОСЛІДЖЕННІ КОНТЕНТУ ТА СТРУКТУРИ СОЦІАЛЬНИХ МЕРЕЖ

Анотація. *Розглянуто особливості інформаційних потоків у соціальних мережах, створення сценаріїв аналітичної діяльності при їх дослідженні. Наведено приклади застосування сценарного підходу.*
Ключові слова: *аналіз соціальних мереж, сценарний підхід, аналітична діяльність.*

Аннотация. *Рассмотрены особенности информационных потоков в социальных сетях, создания сценариев аналитической деятельности при их исследовании. Приведены примеры использования сценарного подхода.*
Ключевые слова: *анализ социальных сетей, сценарный подход, аналитическая деятельность.*

Summary. *The features of information flow in social networks, the features of creation of analytical activities scenario are considered. Examples of the use of the scenario approach are given.*
Keywords: socia*l network analysis, scenario approach, scenario, analytical activities.*

Постановка проблеми. Он-лайнові соціальні мережі (далі – ОСМ) є важливою складовою мережі Інтернет. Їх кількість зростає, а найбільші з них досягли гігантських розмірів. Станом на листопад 2014 року кількість користувачів у Facebook досягла 1,35 млрд. (див. Рис. 1).

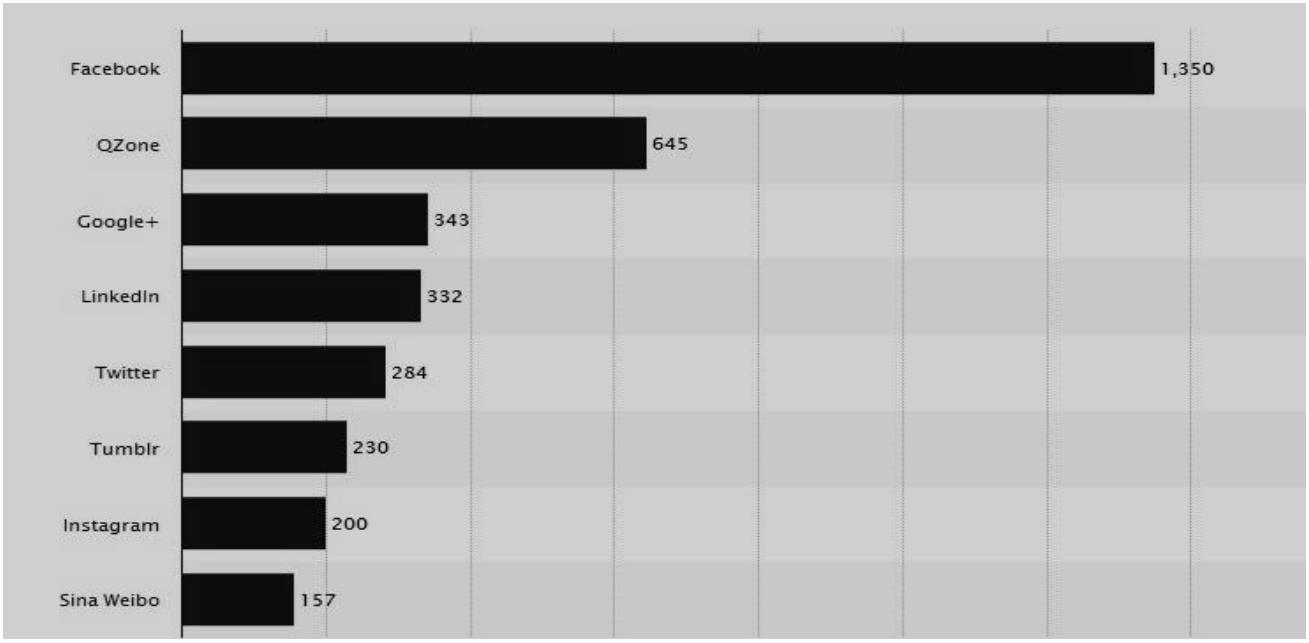

Рис. 1 – Кількість активних користувачів у основних соціальних мережах, за даними statista.com [5]

Метою статті є оцінка підходів при дослідженні контенту та структури соціальних мереж.

Виклад основного матеріалу.

Сучасні підходи до дослідження соціальних мереж.

Для дослідження соціальних мереж використовують теорію графів, теорію великих систем, вейвлет-аналіз [7 – 9].

У загальному вигляді модель комп’ютерної мережі (в тому числі ОСМ) являє собою випадковий граф, закон взаєморозміщення ребер і вершин для якого задається розподілом ймовірностей. У даний час сформульовано чотири основних підходи до моделювання складних мереж [9]:

- випадкові пуасонівські графи та узагальнені випадкові графи;
- марковські випадкові графи і модель блукання по “графові графів” з ймовірностями, які пропорційні бажаним властивостям;
- модель “тісного світу” Ватса і Строґатса і її узагальнення, еволюційна модель росту мережі Барабаші і Альберта;
- модель Прайса.

Особливої ваги набирають дослідження безпеки ОСМ. Моделі просочування (percolation) і зараження (contagion), являють собою популярний спосіб вивчення поширення інформації в соціальній мережі. Блоггер (людина, яка веде блог – мережевий щоденник) може прочитати блог одного (сприйнятливий), присвячений деякої теми, а потім може і сам написати про цю тему (інфікований), і пізніше повернутися до неї (сприйнятливий). Для соціальних мереж ключовим показником є “епідемічний поріг” – критична ймовірність зараження сусіда, при перевищенні якої “інфекція” поширюється по всій мережі. Епідемічний поріг залежить від властивостей графа соціальної мережі, наприклад: числа вершин, розподілу зв’язків, коефіцієнта кластеризації. Тому поширення інфекції сильно залежить від обраної моделі графа мережі [4].

Також для дослідження інформаційної безпеки ОСМ успішно використовуються наступні моделі:

– Розгалужені процеси Маркова (ймовірність зараження комп’ютера користувача соціальної мережі).

– Математична модель процесу зараження комп’ютерів користувачів соціальних мереж на основі теорії випадкових графів.

– Математична модель поширення шкідливого програмного забезпечення в соціальних мережах на основі теорії ланцюгів Маркова.

– Імовірнісні моделі інформаційно-психологічного впливу на користувачів соціальних мереж.

– Ризик-моделі інформаційно-психологічних впливів на користувачів соціальних мереж.

Інформаційні потоки та інформаційне протиборство у соціальних мережах.

Соціальну мережу можна розглядати як множину об’єктів, які в процесі взаємодії утворюють інформаційні потоки.

В результаті аналізу численних діаграм поведінки типових інформаційних операцій, були виявлені найбільш типові, базові профілі їх поведінки. Запропоновані моделі повністю відповідають реальним даним, які екстрагуються системами контент-моніторингу ОСМ. Тому наведені залежності можуть бути використані як шаблони, наприклад, для виявлення інформаційних операцій – як шляхом аналізу ретроспективного фонду мережевих публікацій, так і для оперативного моніторингу

появи деяких їхніх ознак в реальному часі. Зокрема, для виявлення інформаційних операцій слід уважно стежити за динамікою публікацій за цільовою темою і, якщо є можливість, користуватися доступними аналітичними засобами, засобами цифрової обробки даних і розпізнавання образів, наприклад, вейвлет-аналізом [7].

Основними проблемами, що підлягають вирішенню під час інформаційного протиборства, є:

- виявлення існуючих джерел уразливості в позиціях супротивника щодо спірних питань, а також у системі їх інформаційної підтримки;
- пошук потенційних джерел уразливості на основі прогнозних оцінок можливої реакції супротивної сторони на певні інформаційні впливи з метою їх подальшого використання для відстоювання власної позиції;
- планування та оцінка ефективності з точки зору поставлених цілей і наявних обмежень комплексу цілеспрямованих АІК і ВДІВ по відображенню інформаційних загроз власним позиціям з одночасним посиленням інтенсивності та глибини моніторингу розвитку ситуації.

Формалізація структури сценарію дослідження соціальних мереж.

Сценарії або сценарні моделі є одним з видів логіко-лінгвістичних моделей, призначених для відображення розгорнутих у часі послідовностей взаємопов’язаних подій, операцій або процесів. Сценарії можуть мати структуру, в якій встановлені умови переходу до тієї чи іншої приватної стратегії, або просто відображені можливі альтернативи без вказівки умов. Вимога взаємопов’язаності стосовно сценарних моделей не є строгою і має досить умовний характер, оскільки встановлюється на основі суб’єктивних суджень експертів, а також визначається специфікою формулювання цілей діяльності.

Слід зазначити, що в ряді випадків важко провести грань між сценарною моделлю і алгоритмом. Однак між сценарною моделлю і алгоритмом існує досить велика різниця, яка полягає в тому, що алгоритм – це сукупність інструкцій, виконання яких повинно призвести до деякого результату, в той час як сценарна модель – це не обов’язково алгоритм, наприклад, вона може являти собою протокол подій, повторення яких в тій же послідовності не обов’язково призведе до тієї ж ситуації, що і в попередній раз. Тобто, поняття сценарної моделі – це більш широке поняття, ніж поняття алгоритму. Сценарна модель накладає менш суворі обмеження на характер причинно-наслідкових відносин.

Сценарій аналітичної діяльності складається з наступних елементів:

- учасники сценарію;
- короткий опис суті відпрацьовуємо у сценарії функціонального завдання;
- етапи виконання функціональної задачі ФЗ;
- кроки САД (включаючи тимчасові вимоги);
- зміст дій учасників на кожному кроці сценарію;
- посилання на екранні форми, що ілюструють зміст дій;
- опис задіяних БД;
- опис умов виконання кроку сценарію;
- опис обмежень на виконання;
- посилання на ділянки проектної документації, що роз’яснюють зміст кроку сценарію;
- коментарі.

Програма виконання сценарію може бути описана у вигляді алгоритму, і в свою чергу, включає виконання певних елементарних операцій.

Структура САД задається множиною трійок, які ставляться у відповідність діям:

$$AM_r = <\{X_i\}_{i_r=1}^{n_r}, A_r, X_{j_r}, >, \quad (1)$$

де: Ar – дія; $\{X_i\}_{i_r=1}^{n_r}$ – множина вхідних станів AMr для дій; X_{j_r} – вихідний стан AMr , тобто стан, після успішного виконання дії; AM – множина всіх дій.

Складний сценарій описується наступним чином:

$$SM_r = <\{X_i\}_{i_r=1}^{n_r}, S_r, X_{j_r}, >, \quad (2)$$

де: SMr – сценарій; $\{X_i\}_{i_r=1}^{n_r}$ – множина вхідних станів SMr; X_{j_r} – вихідний стан сценарію, тобто стан, після успішного виконання дії; AM – множина всіх дій.

SMr ∈ { SMk } = SM – множина всіх можливих сценаріїв розвитку ситуації у РКС.

Визначимо SV = AM ∪ SM, тобто множину всіх моделей розвитку ситуації і відповідних сценаріїв.

Позначимо X множину всіх станів (проміжних і кінцевих), $X = \bigcup_{r=1}^{S} \bigcup_{i_r}^{n_r} \{X_i\}$.

Відношення – UAX на множині SV визначається декартовому добутку SV × X. В ньому для кожної успішної дії Ar ∈ SV наступним є тільки один елемент із множини X, який є результатом дії.

Відношення – YAX на множині SV визначається декартовому добутку X × SV. В ньому для кожної успішної дії Ar ∈ SV попереднім є тільки один елемент із множини X, який є попередником дії.

Таким чином, модель САД у РКС можна представити у виді:

$$SM\ SKB = <SV, X, U_{AX_r}, Y_{XA}>, \quad (3)$$

де: SV – множина дій (в тому числі елементарних операцій); X – множина станів; UAX – відношення наступності; YAX – відношення передування на множині станів.

Формально етап створення компонентів можна описати математичним виразом:

$$Com^{pr}\{Sp_n^{\Pr}, F_n^{\Pr}, Q_n^{out}, St^{cias}, U^{cias}, Rs^{IT}\}$$

Кінцевим результатом процесу Com^{Pr} є ПЗ вузлів обробки та перетворення даних (з використанням визначених інформаційних технології, наприклад, U_{dw} у сполученні з U_{olap}) або програмна реалізація різноманітних додатків, які формують функціональні АРМи:

$$C_n^{\Pr}\{F_n^{\Pr}, Sn_n^{\Pr}, Q_n^{out}\},$$

де: $Sn_n^{\Pr}$ – множина сценаріїв аналітичної обробки, які реалізують визначену функціональність $F_n^{\Pr} \bigcup F^{pd}$ *n*-го додатку; Q_n^{out} – певна множина вихідних форм (результатів розвитку ситуації користувачів), яка має утворюватися n-им додатком у процесі аналітичної обробки інформаційних ресурсів РКС.

Типові приклади сценаріїв.

Розглянемо наступні приклади сценаріїв.

Як приклад сценарію аналітичної діяльності можна розглядати методику, запропоновану у роботі [7]. Сценарій аналізу включає наступні етапи:

- вибір системи інтеграції Інтернет-документів;
- формування запиту в середовищі обраної системи;

• знаходження тематичних публікацій за запитом за допомогою систем контент-моніторингу;

• визначення динаміки тематичних публікацій за запитом;

• визначення критичних точок в динаміці тематичних публікацій;

• визначення основних подій в критичних точках;

• виявлення об’єктів моніторингу;

• виявлення і візуалізація взаємозв’язків;

• прогноз розвитку подій.

В даний час існує декілька відкритих інформаційних сервісів, в рамках яких можна спостерігати тимчасову динаміку обсягів публікацій за тематиками, визначеним запитами. Так Google books Ngram Viewer (http://ngrams.googlelabs.com), сайт Національного корпусу російської мови (НКРЯ) (http://www.ruscorpora.ru/ngram.html), які виконують візуалізацію динаміки кількості книг, в яких згадуються слова.

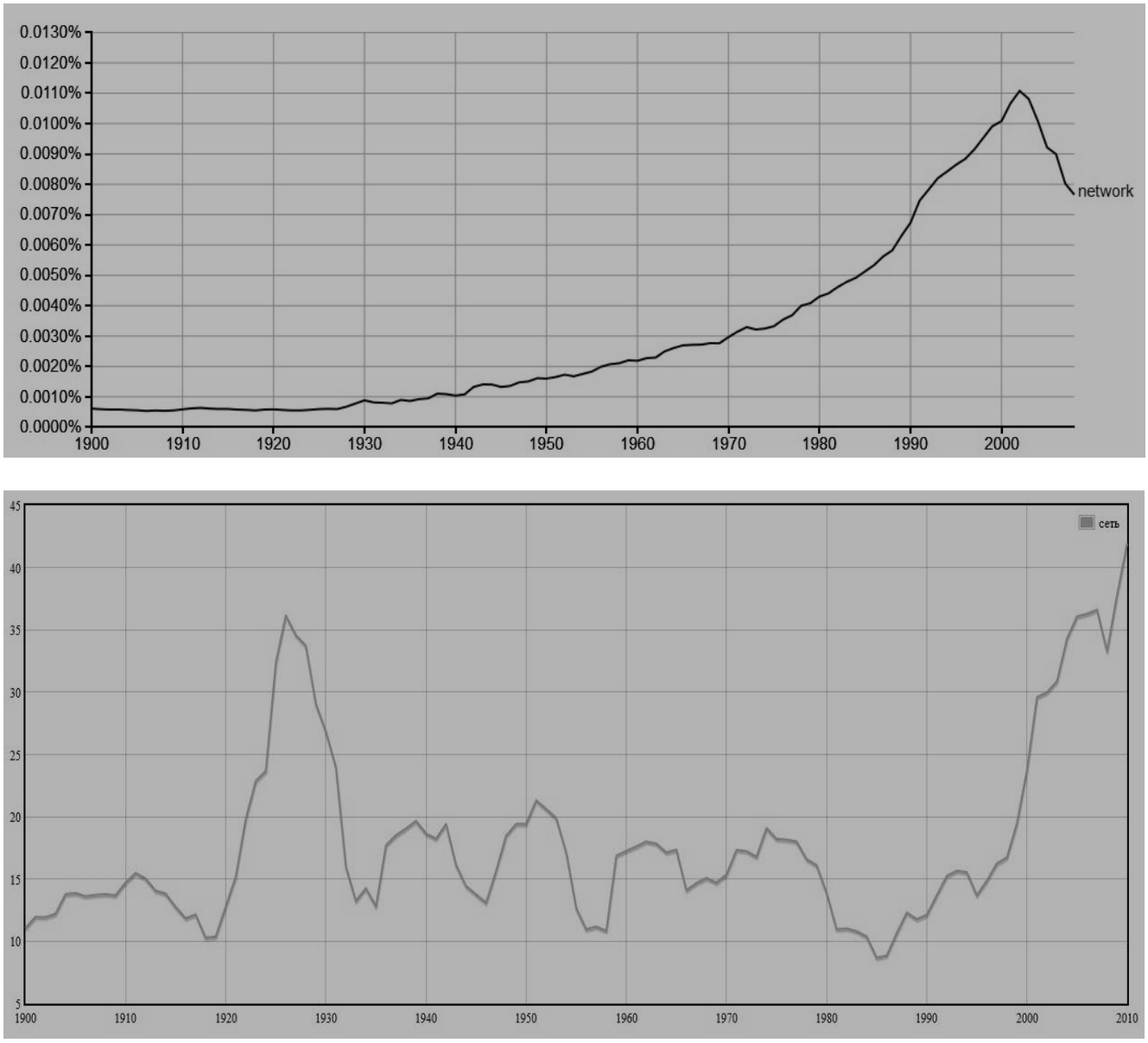

Рис. 2 – Динаміка обсягів публікацій за запитами “network” та “сеть” на сайтах відповідно //www.ngrams.googlelabs.com та //www.ruscorpora.ru

Типовий сценарій дослідження такої динаміки містить такі кроки:

1. Вибір ресурсу для побудови динаміки.
2. Введення слова, для якого будується динаміка.
3. Завдання періоду (наприклад 1900 – 2010 роки).
4. Вказування коефіцієнту згладжування.
5. Побудова графіка.

Якщо результат задовольняє поставлену задачу – експорт отриманих даних до аналітичного документу,

Якщо ні – повернення до п. 1.

Висновки.

Таким чином, застосування сценарного підходу при дослідженні контенту та структури соціальних мереж дозволяє значно підвищити ефективність аналітичної діяльності та забезпечити успішність інформаційного протиборства в он-лайн соціальних мережах.

Застосування підходу, на погляд авторів, дозволить розв’язати сформульовану проблему щодо надійного забезпечення. В подальшому на основі отриманих результатів планується розробка відповідного моделюючого комплексу.

Використана література

1. А.Г. Додонов, Д.В. Ландэ. Выявление понятий и их взаимосвязей в рамках технологии контент-мониторинга // Реєстрація, зберігання і обробка даних. – 2006. – № 4. – С. 8.

2. Губанов Д.А. Социальные сети : модели информационного влияния, управления и противоборства / Д.А. Губанов, Д.А. Новиков, Д.А. Чхартишвили. – М. : Издательство физ.-мат. литературы, 2010.

3. Кононов Д.А. Сценарный анализ динамики поведения социально-экономических систем. Финансовая математика / Д.А. Кононов, С.А. Косяченко, В.В. Кульба ; под ред. Ю.М. Осипова, М.В. Грачевой, Р.М. Нижегородцева, Е.С. Зотовой. – М. : ТЕИС. 2001. – С. 7–53.

4. Информационные риски в социальных сетях : монографія / [Г.А. Остапенко, Л.В. Паринова, В.И. Белоножкин, И.Л. Батаронов, К.В. Симонов] ; под ред. чл.-корр. РАН Д.А. Новикова. – Воронеж : “Научная книга”, 2013. – 160 с.

5. – Режим доступа : //www.statista.com

6. Toriumi F., Okada I., Yamamoto H., Suwa H., Izumi K., Hashimoto Y. Classification of Social Network Sites based on Network Indexes and Communication Patterns // International Workshop on Social Web Mining Co-located with IJCAI2011, 2011. – Режим доступа : //www.users.cecs.anu.edu.au/~sguo/swm2011_submission_10.pdf

7. Додонов А.Г., Ландэ Д.В. Методика аналитического исследования динамики событий на основе мониторинга веб-ресурсов сети Интернет // Информационные технологии и безопасность : основы обеспечения информационной безопасности. – Вып. 14. – К., 2014.

8. Usui S., Toriumi F., Matsuo M., Hirayama T., Mase K. Greedy Network Growth Model of Social Network Service // Computational Intelligence and Intelligent Informatics. – 2014.– № 4. – P. 590-597. – Режим доступа : //www.fujipress.jp/finder/xslt.php?mode=present&inputfile=JACII001800040015.xml

9. Пасічник В.В., Іванущак Н.М. Моделювання і візуалізація засобами PROCESSING комп’ютерних мереж // Математичні машини і системи. – 2013. – № 1.

~~~~~~~~~~~~~ * * * ~~~~~~~~~~~~~
~~~~~~~~~~~~~