УДК 004.5

А. В. Бойченко, Д. В. Ланде
Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна

Формування сценаріїв управлінської діяльності на основі дослідження інформаційного простору

Розглянуто технологічні рішення, що дозволяють аналізувати масиви повнотекстових документів, і запропоновано процедуру автоматизованої побудови сценаріїв управлінської діяльності на основі аналізу інформаційного простору для вибраних предметних областей.

Ключові слова: *сценарний аналіз, управлінська діяльність, інформаційні технології, семантична мережа.*

Вступ

Прискорення розвитку інформаційних технологій вимагає від науковців невідкладного вирішення задачі підвищення ефективності управлінської діяльності. Метою цієї діяльності є задоволення інформаційних потреб суспільства за допомогою найсучасніших комп'ютерних технологій, за рахунок переробки гігантських масивів даних та отримання якісно нових знань.

Важливою складовою управлінської діяльності є сценарний аналіз, який дає можливість дослідити, наскільки істотними виявляється вплив того чи іншого показника на розвиток подій і спрогнозувати ситуації у майбутньому. Без сценарного аналізу неможливо вирішити задачі оптимізації управлінських рішень, глибокого аналізу подій як цілісного явища, розробки стратегії і тактики на кожному рівні складних соціально-економічних систем; визначення ефективності методів керування і внесення пропозицій з їхнього удосконалення, оперативної оцінки ефективності управляючих впливів у найрізноманітніших напрямках і своєчасному реагуванні на виявлені загрози.

У системах підтримки прийняття рішень [1] знаходить застосування ідея моніторингу інформаційного простору та сценарного аналізу.

Технології інтелектуального аналізу тексту (Text Mining) забезпечують отримання структурованої інформації з колекцій текстових документів і включають [2]:

— інформаційний пошук;
— витяг інформації;
— категоризацію тексту.

© А. В. Бойченко, Д. В. Ланде

Останнім часом також багато уваги приділяється дослідженням у напрямку використання онтологічного підходу. Використання онтологій надає можливість логічного виводу нових знань із семантичної мережі і забезпечує ряд переваг, зокрема, при вирішенні задач моніторингу інформаційного простору та безпеки інформації [1, 4].

Як показують недавні дослідження, видобування даних з текстових джерел і виділення з них ключових понять може використовуватись і в задачах сценарного аналізу [3, 5].

При розробці сценаріїв управлінської діяльності необхідно вирішувати ряд задач, основними з яких є:

— формування аналітичної складової сценарної моделі;

— обґрунтований вибір складу та структури моделі;

— вибір оптимального складу програмних засобів моделювання, що включає пошук і порівняння доступних моделюючих пакетів і бібліотек і реалізацію оригінальних алгоритмів шляхом створення власного програмного забезпечення;

— планування процесу розробки сценаріїв для забезпечення виконання процедури моделювання до прийняття рішення.

Методика досліджень

Для отримування сценаріїв управлінської діяльності пропонується технологія, яка передбачає аналіз масиву документів, що описують предметну область; виділення ключових понять: окремих слів, біграм (пар слів), триграм (трійок слів); формування графа зв'язків між ключовими поняттям за допомогою вагових критеріїв; експертну оцінку взаємовпливу понять та присвоєння числових значень графа предметної області; аналіз отриманої мережі понять за допомогою комплексу пакетів моделювання та розроблених програм узгодження форматів цих засобів [5].

Сформовані мережі понять розглядаються як семантичні мережі, що дозволяють застосовувати розвинені технології побудови сценаріїв управлінської діяльності. Аналіз такої мережі включає дослідження змісту складових її блоків, цільових і керуючих факторів, аналіз шляхів і циклів і взаємозв'язків між ними.

Сценарій має бути сформований із мережі, вершинами якої є факти, а дугами — зв'язки, що описують відносини спеціального типу, наприклад «причина – наслідок», «частина – ціле». Сценарії використовуються як засіб представлення проблемно-залежних каузальних знань і задаються у вигляді фреймоподібних спискових структур.

Для побудови графа сценарію пропонується використання абстрактного автомату, що описує вимоги на підставі опису аналітичної задачі.

Представимо сценарій поведінки об'єкта як послідовність пар $(I(t_i), t_i)$, яка формується відповідно до правил вибору $A^{(t)}$ [7]:

$$\Re = \Re \{(I(t_i), t_i) \mid i = 0,1,..., N; t_i = 0\},$$

де N — глибина сценарію; $T = t_N$ — горизонт сценарію; $T = t_{i+1} - t$ — крок сценарію.

Побудова сценарію на момент часу $t_k \in ZT$ складається з кількох етапів:

1) відповідно до моделей M виконується оцінка початкового стану об'єкта, тобто уточнюються величини попередніх подій $\Im^{(k)}(t_k)$;

2) фіксується базовий стан метанабору — множина $M_0(t_k)$;

3) виконується оцінка ситуації $S(t_k)$, тобто уточнюються оцінки попередніх подій з точки зору поточного стану моделей M;

4) формалізується поточна обстановка шляхом фіксації $\Theta(t_k)$;

5) визначається можливість продовження формування сценарію $\Re$ у заданому напрямку залежно від поточного стану його складових елементів і стану оточуючого середовища чи необхідністю деталізації сценарію $\Re$;

6) виконується вибір множини $M(t_k)$ стратегій побудови сценарію $\Re$;

7) уточнюється правило оцінки стратегій побудови сценарію $\Re$;

8) виконується оцінка стратегій побудови сценарію $\Re$ як вибір компоненти $\Delta(t_k)$ деякої точки $\varsigma = (\vartheta, \alpha, \beta)$ множини умовних рішень $I_0 = M_0\, N_0\, B_0$;

9) уточнюється ступінь деталізації сценарію $\Re$;

10) виконується вибір чергової події і кроку сценарію в процесі реалізації обраної стратегії побудови.

Остаточний відбір вузлів-понять виконується експертами в предметній області [7].

Для пошуку сценарію пропонується використати модифікований мурашиний алгоритм, який доцільно використовувати для задач дискретної оптимізації [8]. Вихідні дані для цього алгоритму наступні:

$C = \{c_1, c_2, \ldots, c_N\}$ — множина елементів (у нашому випадку — вузлів-понять предметної області);

$L = \{l_{c_i,c_j} \mid (c_i, c_j) \in \tilde{C}\}, \mid L \mid \leq N_C^2$ — множина можливих переходів між елементами C, де $\tilde{C}$ — множина можливих з'єднань між вузлами;

$G = (C, L)$ — граф, який представляє задачу дискретної оптимізації;

$J_{c_i c_j} \equiv J(l_{c_i c_j}, t)$ — функція ваги переходу, що пов'язана з усіма $l_{c_i,c_j} \in L$;

$\Omega \equiv \Omega(C, L)$ — множина обмежень, визначених над елементами множин C та L;

$s = \langle c_i, c_j, \ldots, c_k, \ldots \rangle$ — послідовність елементів C, яка називається станом задачі, якщо $\tilde{S} \subseteq S$ при обмеженнях $\Omega(C, L)$;

Якщо задано стани $s1$ та $s2$, для яких встановлено відношення суміжності, тобто якщо при $s2$ може бути досягнутий зі стану $s1$, $c_2 \in C$, то $l_{c_1 c_2} \in L$ та $s_2 \equiv \langle s_1, c_2 \rangle$;

Ψ — рішення, якщо $\Psi \in \tilde{S}$ задовольняє всім обмеженням задачі;

$J_\Psi(L)$ — вартість рішення Ψ, $J_\Psi(L)$ — функція всіх вартостей, для всіх, що належать Ψ.

Вирішення задачі можуть бути виражені у вигляді можливих шляхів на графі G.

На рис. 1 представлено блок-схему модифікованого алгоритму, який пропонується для вибору сценаріїв у графі. Процес починається з розміщення у вершинах графа агентів, які моделюють рух мурах, потім починається їхній рух — напрям визначається ймовірнісним методом, на підставі формули:

$$P_i = \frac{l_i^q \cdot f_i^p}{\sum_{k=0}^{N} l_k^q \cdot f_k^p},$$

де P_i — імовірність переходу шляхом i; l_i^q — довжина i-го переходу; f_k^p — кількість феромонів на i-му переході; q — величина, яка визначає «жадібність» алгоритму i; p — величина, яка визначає «стадність» алгоритму i.

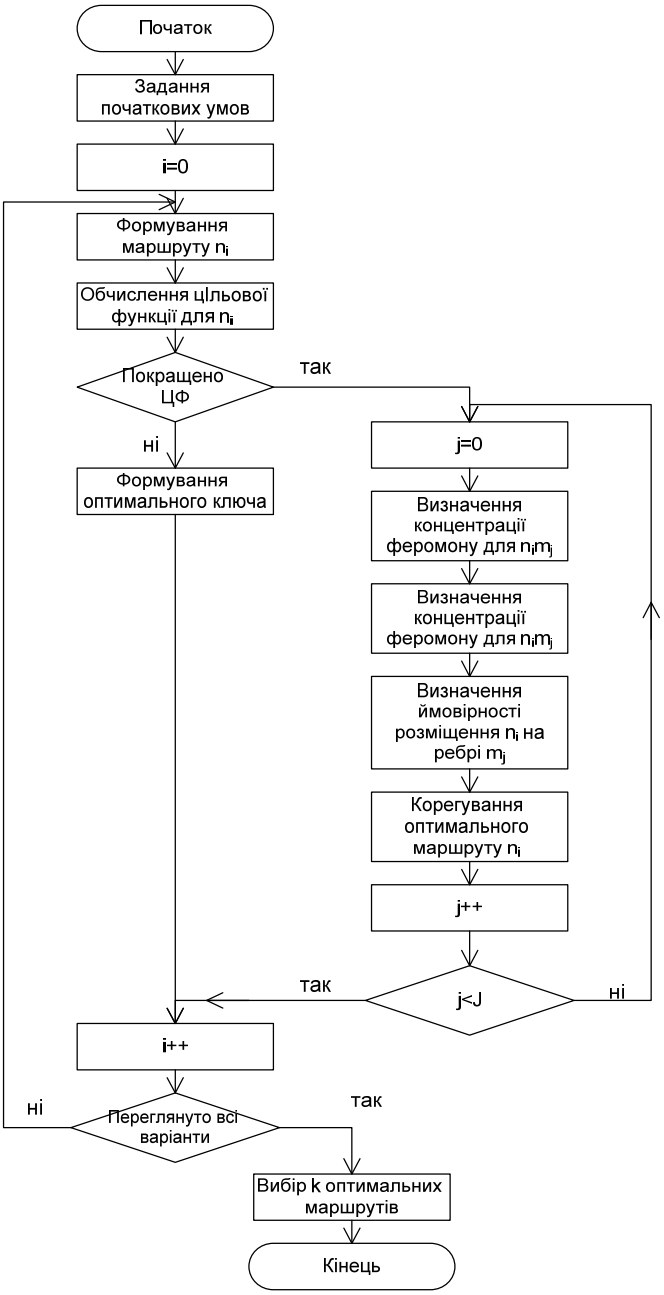

Рис. 1. Модифікований мурашиний алгоритм

У рамках наведеної методики було проведено формування текстових корпусів з інформаційних сайтів мережі Інтернет за темами: «Brexit», «Mass Shuting», «Catalan referendum», «Saudi Arabian purge», «Zimbabwe coup d'etat», «Syrian civil war».

Подальше дослідження отриманої мережі здійснюється за допомогою програмних засобів моделювання та візуалізації Gepfi та Protégé, які належать до вільного програмного забезпечення. Також використовуються оригінальні програмні модулі, що написані мовою програмування Perl з використанням модулів LWP, HTML, Lingua.

На рис. 2 показано граф, який використовувався для побудови сценарію, вершини якого відображають терми, що отримані з текстового корпусу «Syrian civil war», а дуги відповідають взаємодіям між ними. Програмний пакет Gephi дозволяє виконати кластеризацію графа та вирахувати значення показників портальності (hub) та авторства (auth), для вершин використовується алгоритм HITS (Hyperlink Induced Topic Search).

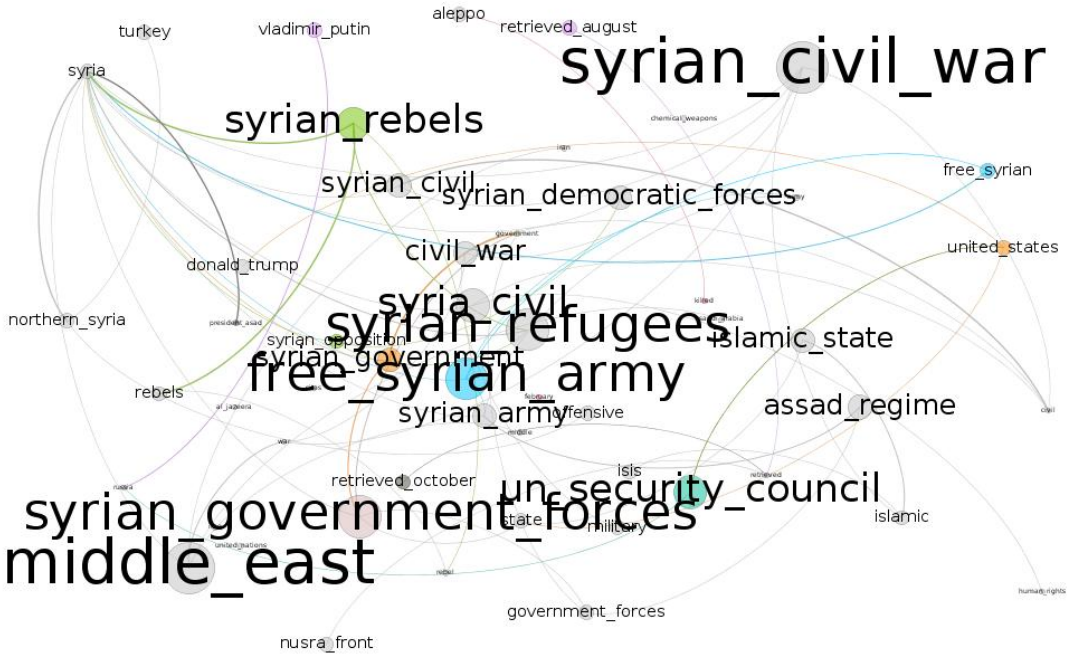

Рис. 2. Отриманий граф у програмі Gephi

Gephi також дозволяє виконати візуалізацію мережі та експортувати результати у формат CSV, який у подальшому за допомогою розробленого інтерфейсу дозволяє отримати файл у форматі OWL — Web Ontology Language, стандарту W3C, мові для семантичних тверджень, розроблений як розширення RDF та RDFS. Згідно розширеної специфікації мови OWL2 профіль OWL2 EL передбачає використання 19-ти типів даних, більшість з яких виразно в рамках специфікації XSD — мови опису структури XML-документів.

На рис. 3 показано фрагмент мережі у процесі дослідження засобами програмного пакету Protégé. Protégé — це редактор онтологій з відкритим кодом і

фреймворк для побудови баз знань, який підтримує два основних способи моделювання онтологій за допомогою редакторів Protégé-Frames і Protégé-OWL. Онтології, що побудовані в Protégé, можуть бути експортовані до форматів RDF, OWL і XML Schema. Побудова онтології дозволяє відобразити основні фактори та можливі взаємозв'язків між ними і є базою для побудови більш детальних комп'ютерних сценаріїв розвитку ситуації. Відкрита архітектура проекту дозволяє підключати велику кількість модулів розширення функціональності, написаних сторонніми розробниками а також створювати власні. Таким чином, застосування онтологій при реалізації сценарного підходу дозволяє значно підвищити ефективність аналітичної діяльності.

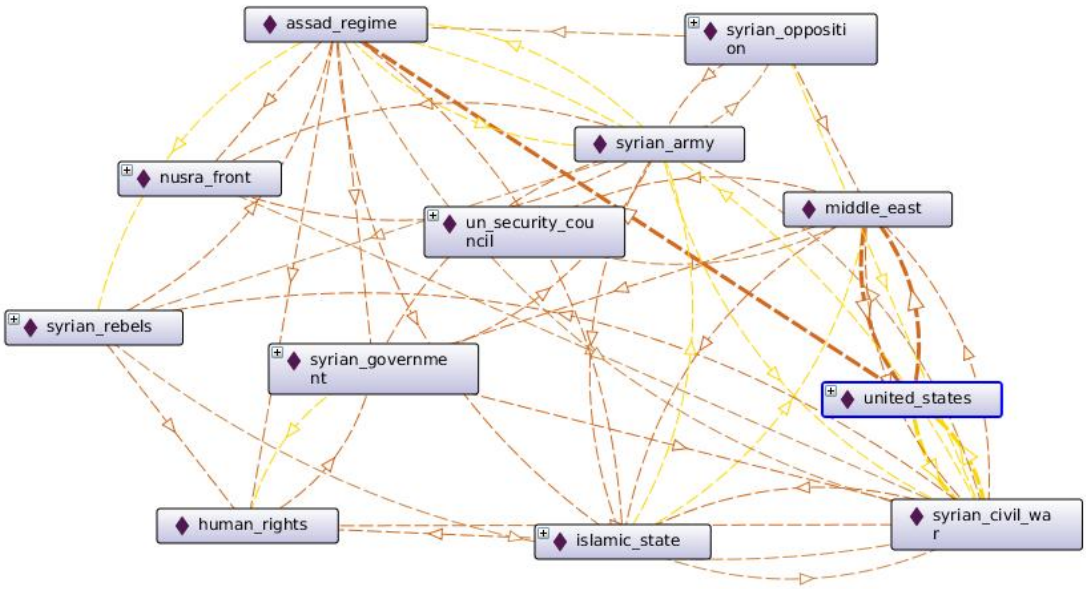

Рис. 3. Фрагмент онтології у програмому комплексі Protégé

Характеристики отриманих сценаріїв перевіряються експертами у даній галузі, що дозволяє уточнити значення параметрів для сценаріїв і виконати деякі (далеко не повні) перевірки коректності їхніх результатів. Оскільки ті самі типи даних використаються багаторазово, автоматизована генерація сценаріїв дозволяє істотно знизити витрати на етапі моделювання.

Висновки

За результатами описаних досліджень одержано такі основні результати:

— показано необхідність застосування сценарних методів для побудови мережі понять на основі аналізу існуючих підходів і засобів автоматизації управлінської діяльності;

— розроблено та досліджено математичну модель якісної оцінки сценаріїв управлінської діяльності для предметної області;

— розроблено інформаційну технологію автоматичного формування моделі предметної області на основі сканування текстових ресурсів і створення мережі понять;

— розроблено інформаційну технологію інтеграції інструментальних програмних засобів дослідження графів та онтології для дослідження текстового інформаційного простору;

— розроблена технологія може використовуватися в інформаційно-аналітичних системах органів державної влади та науково-освітніх закладах.

Запропонована технологія дозволяє в автоматизованому режимі на базі аналізу вхідного пакету документів вирішувати задачу розробки та дослідження сценаріїв здійснення впливів на об’єкти, які відповідають вибраним ключовим поняттям.

1. Додонов А.Г., Ландэ Д.В., Коваленко Т.В. Модели предметных областей в системах поддержки принятия решений на основе мониторинга информационного пространства. Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016): Материалы VI междунар. науч.-техн. конф. (18–20 февраля 2016 г., Минск). Минск: БГУИР, 2016. С. 171–176.

2. Srivastava Ashok N., Mehran Sahami. Text Mining: Classification, Clustering, and Applicatios. Chapman and Hall/CRC, 2009. 328 p.

3. Kayser Victoria, Shala Erduana. Generating Futures from Text: Scenario Development using Text Mining. 5th International Conference on Future-Oriented Technology Analysis (FTA) — Engage today to shape tomorrow. Brussels, Belgium, 2014.

4. Zobia Rehman., Claudiu V. Kifor. An Ontology to Support Semantic Management of FMEA Knowledge. *International Journal of Computers Communications &Control*. 2016. **11**(4).

5. Ланде Д.В., Бойченко А.В. Побудова моделі розвитку ситуації на основі аналізу інформаційного простору. *Актуальні проблеми управління інформаційною безпекою держави*: зб. матеріалів наук.-практ. конф. (24 трав. 2017 р., Київ). Електрон. дані. Київ: Нац. акад. СБУ, 2017. С. 61–62.

6. Ланде Д.В., Фурашев В.М. Термінологічна мережева модель як відображення процесів денцентралізації влади в Україні. *Інформація і право*. 2017. № 2(21). С. 66–71.

7. Ланде Д.В., Бойченко А.В. Організація аналітичної діяльності на основі сценарного підходу. *Реєстрація, зберігання і оброб. даних*. 2015. Т. 17. № 1. С. 68–76.

8. Dorigo M. and Di Caro G. The Ant Colony Optimization Meta-Heuristic. In D. Corne, M. Dorigo, and F. Glover, editors. New Ideas in Optimization. McGraw-Hill, 1999. P. 11–32.

Надійшла до редакції 15.12.2017