

УДК 004.5

**А. В. Бойченко**

Інститут проблем реєстрації інформації НАН України  
вул. М. Шпака, 2, 03113 Київ, Україна

## **Побудова онтологічної моделі для задач сценарного аналізу**

*Розглянуто використання онтологічної моделі при організації сценарного моделювання предметних областей. Запропонована технологія дозволяє в автоматизованому режимі, на базі аналізу вхідного пакета документів, вирішувати задачу розробки та дослідження сценаріїв інформаційного впливу на об'єкти, які відповідають вибраним ключовим поняттям.*

**Ключові слова:** сценарний аналіз, онтології, когнітивні карти, аналітична діяльність.

Сценарний аналіз найбільш повно відповідає завданням дослідження та прогнозування поведінки складних процесів, до яких належить і інформаційна безпека.

Відповідно до сценарної методології сценарієм поведінки систем є послідовність розширених фазових станів (виділених подій) та умов його функціонування. Сценарій описує процес зміни його параметрів моделі, послідовно відображуючи моменти переходу моделі до нового якісного стану.

Інформаційні технології сценарного моделювання включають: формування структури масивів вхідних і вихідних даних, методику ітеративного процесу моделювання предметної області, онтології певної предметної області та оригінальне програмне забезпечення.

У рамках зазначених напрямків слід виділити дослідження Д.В. Ланде, А.А. Снарського, О.Г. Додонова, Н.Г. Загоруйко, Е.А. Рабчевського, В.Ш. Рубашкина, Bart Kosko, Lotfi Aliasker Zadeh, T. Gruber, S. Staab, R. Studer, N.F. Noy, A. Gomez-Perez [1–5].

Тому актуальними є дослідження процесу створення сценарних моделей, систематизація теоретичних підходів до побудови сценаріїв, напрацювання рекомендацій щодо методів і програмно-технічних засобів побудови онтологій предметних областей, підхід до формування когнітивних карт та організації моделювання. Науково обґрунтовані методики формування інформаційних та аналітичних ресурсів дозволять значно скоротити терміни сценарного моделювання, підвищити ефективність інформаційної та аналітичної підтримки процесів напрацювання управлінських рішень.

© А. В. Бойченко

## Рішення, що пропонується

Для вирішення проблеми пропонується технологія моделювання різноманітних предметних областей шляхом ітераційного підходу до розробки та дослідження сценаріїв поведінки систем (рис. 1).

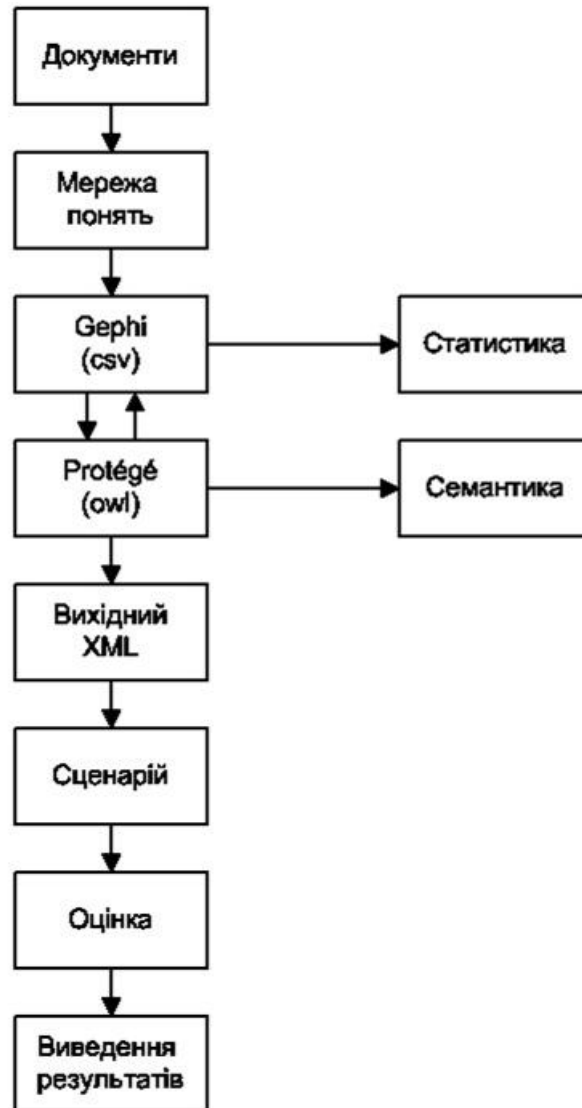


Рис. 1. Процес розробки сценаріїв

Отримати масив документів для дослідження предметної області можна різними шляхами. Для цього може використовуватись як сканування паперової документації, підбір аналітиками файлів у ручному режимі, так і використання системи контент-моніторингу, яка охоплює достатній обсяг інформації, що відноситься до досліджуваної предметної області [2].

Вибір цільової функції для оцінки ваги понять предметної області — вузлів онтології — здійснюється експертами.

Отримана когнітивна карта являє собою знаковий орієнтований граф:  $G = \langle V, E \rangle$ , де  $V$  — множина вершин  $V_i \in V, i = 1, 2, \dots, k$ , які є елементами досліджуваної системи;  $E$  — множина дуг  $e_{ij} \in E, i, j = 1, 2, \dots, N$ , які відображають взаємозв'язок між вершинами  $V_i$  та  $V_j$ . Вплив  $V_i$  та  $V_j$  може бути позитивним, коли збільшення (зменшення) одного фактора призводить до збільшення (зменшення) іншого; негативним, коли збільшення (зменшення) одного фактора веде до зменшення іншого, чи бути відсутнім (0).

Аналіз когнітивної карти включає дослідження змісту складових її блоків, цільових і керуючих факторів, аналіз шляхів і циклів. При формуванні сценаріїв експерт вибирає, які вершини в отриманій моделі слід виділити, та які зв'язки між вершинами враховувати, а які — ні. З отриманої таким чином множини можливих сценаріїв поведінки моделі обираються ті, що забезпечують прийнятний рівень захищеності.

### Аналіз графа понять та отримання статистики

Ребра графа мають ваги +1 або -1, скорочено позначаються знаками «+» чи «-». Знак «+» позначає позитивний зв'язок, знак «-» позначає негативний вплив. Вага шляху дорівнює добутку ваг його ребер, тобто позитивний, якщо число негативних ребер у ньому парне, і негативний, якщо це число непарне. При позитивному зв'язку зростання чинника-причини призводить до зростання фактора-наслідку, а при негативному зв'язку зростання чинника-причини призводить до зменшення фактора-наслідку. Якщо ж від вершини  $a_i$  до вершини  $b_j$  ведуть як позитивні, так і негативні шляхи, то питання про характер впливу фактора  $a_i$  на фактор  $b_j$  залишається невизначеним.

Імпульсний метод, який найчастіше використовується для обчислення когнітивних карт має ряд недоліків, зокрема існує проблема сходження ряду, що досліджується.

K-метод, що запропонований у роботі [7], позбавлений недоліків імпульсного методу.

Для обчислення ваги вузлів в отриманому графі пропонується алгоритм, подібний до відомого алгоритму HITS (Hyperlink-Induced Topic Search), що запропонований Дж. Клейнбергом для пошуку найбільш вагомих веб-сторінок відповідно до запиту користувача на основі інформації, яка визначається множиною гіперпосилань [8].

Згідно з модифікованим алгоритмом HITS для кожного вузла мережі обрховуються взаємопов'язані показники авторства (*auth*) та портальності (*hub*) з урахуванням додаткового множника — деякої невід'ємної неспадної функції  $f$  від ваги ребра:

$$\begin{aligned} \text{hub}(A_i) &= \sum_{A_i \rightarrow A_j} \text{auth}(A_j) f(w_{ij}), \\ \text{auth}(A_i) &= \sum_{A_j \rightarrow A_i} \text{auth}(A_j) f(w_{ij}). \end{aligned}$$

Але оскільки ваги ребер орієнтованого графа можуть приймати від'ємні значення, задача обчислення ваги вузлів вимагає додаткової модифікації алгоритму HITS.

На рис. 2 представлено граф, що отриманий шляхом вибору біграм з документації по організації комп'ютерної безпеки.

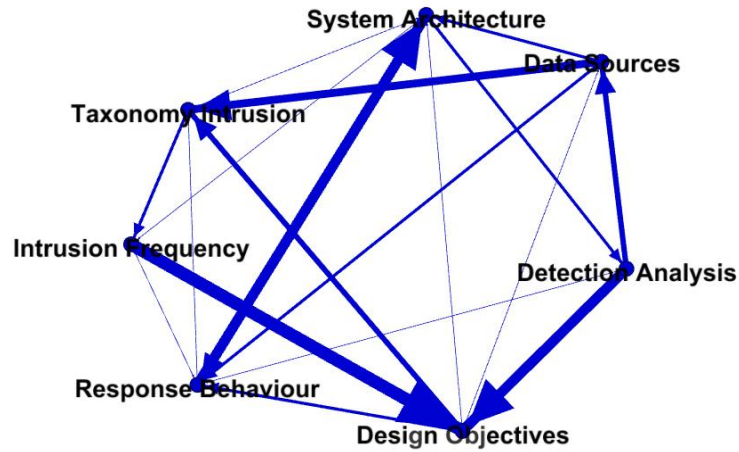


Рис. 2. Отриманий граф у програмі Gephi

### Побудова онтології та аналіз семантики

Процес побудови онтології складається з наступних блоків:

- визначення концептів (поняття, класи, сутності, категорії);
- опису властивості концептів (слоти, атрибути, ролі);
- знаходження зв'язків між концептами (форми зв'язку, залежності, функції);
- внесення додаткових обмежень (аксіоми, фацет).

Онтологію можна представити у вигляді множини

$$\langle C, A, T, D, R, F \rangle,$$

де  $C$  — підмножина класів, що описують поняття деякої предметної або проблемної області;  $A$  — підмножина атрибутів (слотів), що описують властивості понять;  $T$  — підмножина типів значень атрибутів;  $D$  — підмножина доменів;  $R$  — підмножина відносин, заданих на класах (поняттях);  $F$  — підмножина обмежень на значення атрибутів.

Повторне використання існуючих онтологій доцільно, якщо системі потрібно взаємодіяти з іншими додатками, які вже увійшли в окремі онтології або контрольовані словники. Багато онтологій (метаонтології та онтології-аналоги) вже доступні в електронному вигляді та можуть бути імпортовані до використовуваної середі проектування онтології.

Побудова онтології передбачає розробку глосарію, що, в свою чергу, включає вивчення предметної області, виявлення основних понять предметної області та підбір для них несуперечливих визначень.

Ключовою фазою при розробці онтології є визначення класів і створення їхньої ієрархії. Доцільним є застосування комбінованого підходу до розбиття на

класи, який поєднує можливості як спадного так і висхідного. Процес низхідної розробки починається з визначення найбільш загальних понять предметної області з подальшою їхньою конкретизацією. Процес висхідної розробки починається з визначення самих конкретних класів, листя ієрархії з подальшим угрупуванням цих класів у більш загальні поняття.

Визначення властивостей класів передбачає опис внутрішньої структури понять предметної області. При цьому, для всіх класів онтології потрібно обов'язково описати тип значення, дозволені значення, число значень, які може приймати слот.

Аналіз і діагностика розробленої онтології включає застосування аналітичних і різних діагностичних інструментальних засобів аналізу онтологій. Останній аналіз може включати як перевірку логічної вірності онтології, так і діагностику типових помилок при проектуванні онтологій.

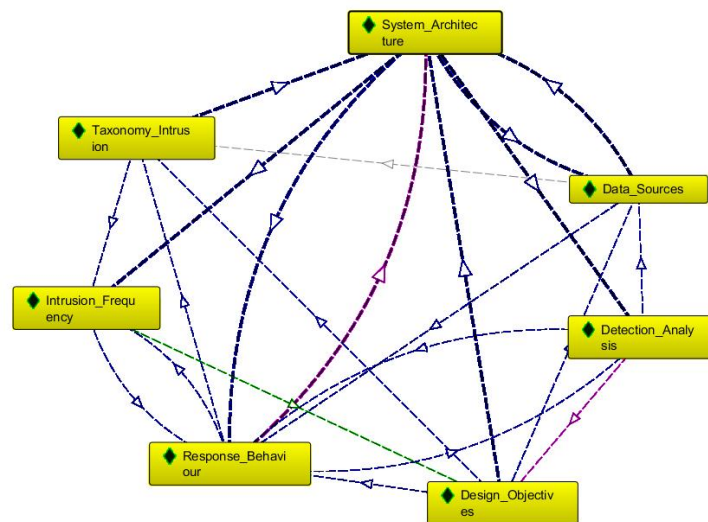


Рис. 3. Приклад отриманої онтології у програмі Protégé

Для розробки програмного забезпечення, яке забезпечує взаємодію між програмними комплексами дослідження орієнтованих графів та онтологій і формування вхідних і вихідних файлів у потрібному форматі було обрано мову програмування Perl. Цей засіб розповсюджується вільно, доступний для основних операційних систем, має потужні можливості для обробки тексту без довільних обмежень на довжину даних і маніпуляції текстовими файлами.

### Формування сценаріїв впливу

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

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

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

де  $A_r$  — дія;  $\{X_i\}_{i=1}^{n_r}$  — множина вхідних станів  $AM_r$  для дій;  $X_{j_r}$  — вихідний стан  $AM_r$ , тобто стан після успішного виконання дії;  $AM$  — множина всіх дій.

Складний сценарій описується наступним чином:

$$SM_r = \langle \{X_i\}_{i=1}^{n_r}, S_r, X_{j_r} \rangle, \quad (2)$$

де  $SM_r$  — сценарій;  $\{X_i\}_{i=1}^{n_r}$  — множина вхідних станів  $SM_r$ ;  $X_{j_r}$  — вихідний стан сценарію, тобто стан після успішного виконання дії;  $SM$  — множина всіх дій;  $SM_r \in \{SM_k\} = SM$  — множина всіх можливих сценаріїв розвитку ситуації у РКС.

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

Позначимо через  $X$  множину всіх станів (проміжних і кінцевих):

$$X = \bigcup_{r=1}^S \bigcup_{i_r}^{n_r} \{X_{i_r}\}.$$

Відношення  $U_{AX}$  на множині  $SV$  визначається декартовим добутком  $SV \times X$ . У ньому для кожної успішної дії  $A_r \in SV$  наступним є тільки один елемент із множини  $X$ , який є результатом дії.

Відношення  $Y_{AX}$  на множині  $SV$  визначається декартовим добутком  $Y \times SV$ . У ньому для кожної успішної дії  $A_r \in SV$  попереднім є тільки один елемент із множини  $X$ , який є попередником дії.

Таким чином, модель сценарію можна представити у вигляді:

$$SM \ SKB = \langle SV, X, U_{AX}, Y_{AX} \rangle, \quad (3)$$

де  $SV$  — множина дій (у тому числі елементарних операцій);  $X$  — множина станів;  $U_{AX}$  — відношення наступності;  $Y_{AX}$  — відношення передування на множині станів.

Отримані в процесі моделювання онтології розглядаються як когнітивні карти, що дозволяють застосовувати розвинені технології побудови сценаріїв здійснення впливів.

Основними параметрами, які відслідковуються у процесі моніторингу виконання сценаріїв:

- дотримання порядку виконання кроків сценарію;
- помилки, що виникають у процесі виконання;
- попередження зациклювань при виконанні сценарію;
- цілісність і повнота даних, які зберігаються в процесі виконання.

## Висновки

Запропоновано технологію розробки сценаріїв розвитку ситуації, засновану на онтологіях, що отримані з масиву документів, які описують певну тематичну область. Технологія включає наступні етапи:

- 1) аналіз масиву документів, які описують предметну область;
- 2) виділення ключових понять: окремих слів, біграм (пар слів), триграм (трійок слів);

3) формування графа зв'язків між ключовими поняттями за допомогою вагових критеріїв. Як такий може розглядатися граф входження понять у близькі фрагменти тексту;

4) експертну оцінку взаємовпливу понять і присвоєння числових значень графу предметної області;

5) аналіз отриманої мережі понять за допомогою комплексу пакетів моделювання (Gephi, Protégé) та розробленого мовою Perl програмного комплексу, що дозволяє інтегрувати засоби моделювання.

Показано, що побудова онтології дозволяє відобразити основні фактори та можливі взаємозв'язки між ними і є базою для побудови більш детальних комп'ютерних сценаріїв розвитку ситуації. Таким чином, застосування онтологій при реалізації сценарного підходу дозволяє значно підвищити ефективність аналітичної діяльності.

1. Додонов А.Г., Ландэ Д.В. Архитектура системы информационной поддержки на основе мониторинга информационного пространства и метода сценарного анализа. *Інформаційні технології та спеціальна безпека*. 2015. № 1. С. 3–14.

2. Додонов А.Г., Ландэ Д.В., Бойченко А.В. Сценарный подход при исследовании динамики информационных потоков в сети Интернет. Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2015): материалы V междунар. науч.-техн. конф. (19–21 фев. 2015, г. Минск). Минск, 2015. С. 225–230.

3. Шульц В.Л., Кульба В.В., Шелков А.Б., Чернов И.В. Сценарный анализ эффективности управления информационной поддержкой государственной политики России в Арктике. *Национальная безопасность/nota bene*. 2011. № 6. С. 104–137.

4. Снарский А.А., Ландэ Д.В. Метод выделения подсетей в каузальных сетях в задачах сценарного анализа. «Информационные технологии и безопасность». Материалы XV Международной научно-практической конференции ИТБ-2015. ИПРИ НАН Украины. Киев, 2015. С. 212–215.

5. Ландэ Д.В., Бойченко А.В. Використання моделей предметних областей у задачах сценарного аналізу. Міжнародна науково-практична конференція «Інтелектуальні технології лінгвістичного аналізу». Тези доповідей. Київ: НАУ, 2016. С. 9.

6. Kleinberg J. Authoritative sources in a hyperlinked environment. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, 1998. P. 668–677.

7. Ландэ Д.В., Снарский А.А. Подход к созданию терминологических онтологий. *Онтология проектирования*. 2014. № 2(12). С. 83–91.

8. Снарский А.А., Зоринец Д.И., Ландэ Д.В., Левченко А.Ю. Метод расчета парных связей в когнитивных картах. XVI Международная научная конференция имени Т.А. Таран «Интеллектуальный анализ информации» ИАИ-2016 (18–20 мая 2016, г. Киев). Киев, 2016. С. 227–234.

Надійшла до редакції 20.12.2016