
INFORMATION TECHNOLOGY

УДК 004 (9+056.5)

ДМИТРО ЛАНДЕ,
АНДРІЙ БОЙЧЕНКО

МЕТОДИКА РОЗРОБЛЕННЯ СЦЕНАРІЇВ РОЗВИТКУ СИТУАЦІЇ НА ОСНОВІ АНАЛІЗУ ІНФОРМАЦІЙНОГО ПРОСТОРУ

Пропонується методика розробки сценаріїв розвитку ситуації, яка дозволяє в автоматизованому режимі вирішувати задачу формування і ранжирування сценаріїв здійснення впливів на об'єкти, які відповідають вибраним ключовим поняттям. Із масиву документів, які описують певну предметну область виділяються ключові поняття, яким можуть бути окремі слова, біграми чи триграми. Між отриманими ключовими поняттями, за допомогою вагових критеріїв, формується граф предметної області. Шляхом експертної оцінки виконується редагування графу та присвоєння числових значень його ребрам, що дозволяє побудувати мережу понять предметної області. В рамках моделі, що розглядається, зважування окремих вузлів мережі, що відповідають термінам, здійснюється за допомогою модифікованого алгоритму HITS (Hyperlink-Induced Topic Search). Модифікація алгоритму передбачає, що при розрахунку значення ваги кожного вузла враховується вага ребер, а для зниження впливу нерівномірності розподілу значень ваги на результат вага вузла множиться на монотонно-зростаючу, менш круту за лінійну функцію. Побудова онтології дозволяє врахувати основні фактори та можливі взаємозв'язків між ними, і є базою для розробки комп'ютерних сценаріїв розвитку ситуації. Для дослідження використовуються пакет візуалізації та моделювання графів Gephi, пакет для побудови онтологій Protégé та власний набір спеціально розроблених модулів на Perl. Для перевірки методики було проведено формування текстового корпусу шляхом сканування Інтернет за темою BREXIT. Візуалізація предметної області дозволяє аналітику досліджувати отримані сценарії та формувати результати у зручному для прийняття рішень вигляді. Методика може використовуватись для задач сценарного моделювання у галузях комп'ютерної безпеки, дослідження інформаційних впливів та підтримки прийняття рішень.

Ключові слова: сценарний аналіз, аналітична діяльність, інформаційні операції, мережа термінів, сценарне моделювання.

Вступ. Слід відзначити, що універсальних і бездоганних підходів до вирішення цієї проблеми на сьогодні не існує – наявні лиш спроби побудови можливих сценаріїв розвитку тих чи інших явищ [1]. Методологія сценарного аналізу була створена для подолання невизначеності, що виникає у процесі прогнозування та прийняття рішень. Сценарний аналіз використовує декомпозицію задач прогнозування з врахуванням експертних оцінок та подальше формування множини сценаріїв розвитку ситуації, які перекривають множину варіантів можливих у майбутньому подій.

Проведення сценарного дослідження для складних соціальних систем вимагає на цей час значних обчислювальних і часових ресурсів (іноді такі дослідження можуть тривати роками), залучення великої кількості висококваліфікованих експертів, аналізу гігантських обсягів інформації на всіх стадіях проекту.

Без сценарного аналізу неможливо вирішити наступні 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.

Областями застосування сценарного аналізу є економіка, інформаційна безпека, екологія, фінанси, бізнес-планування, військова справа.

У загальному випадку сценарний аналіз передбачає такі етапи [2], [3]:

1. Визначення мети написання сценарію.
2. Розробка програми сценарного аналізу, формалізація «цільової функції».
3. Введення у сценарій припущень.
4. Побудова альтернативної схеми подій.
5. Написання сценарію.
6. Аналіз сценарію з урахуванням відгалужень і поворотних моментів.
7. Формування політики (правил) для суб'єкта сценарію.
8. Розробка альтернативних стратегій поведінки суб'єкта сценарію.
9. Оцінювання альтернативних стратегій методом моделювання.

Аналіз останніх досліджень і публікацій. Тематика цього дослідження є актуальною і широко обговорюваною. Зокрема, у роботі [1] сценарний аналіз визначено як універсальну сукупність засобів і підходів, що являє собою комплекс математичних, програмних, логічних і організаційних засобів та інструментів для визначення послідовності застосування окремих методів, взаємозв'язку між ними та формування процесу передбачення в цілому. Ще у 1980-х роках Белл (Bell, 1982), Меркхофер і Кіней (Merkhofer, Keeney, 1987) розглядали можливість використання сценаріїв як базу для побудови стратегічних рішень. Портер (Porter, 1985) обґрунтував використання сценаріїв для дослідження економічних перспектив, Раубітчек (Raubitschek, 1988) – для бізнес-планування; Хасс (Huss, 1988) – прогнозування [4].

Сценарийний аналіз дозволяє розглядати різні альтернативи розвитку подій, визначити найбільш ймовірні сценарії [5]. Дослідження із інформаційних технологій сценарного моделювання, пов'язаних із онтологічним підходом, наведено, зокрема, в [6]. В роботі [7] наведено спеціальну нотацію для опису сценаріїв, що розробляються.

Мета цієї роботи – створення методика розробки сценаріїв розвитку ситуації, яка дозволяє в автоматизованому режимі вирішувати задачу формування і ранжирування сценаріїв здійснення впливів на об'єкти, які відповідають вибраним ключовим поняттям. Як базис для розробки сценаріїв, середовище навчання системи, мають застосовуватися документи, що знаходяться в сегменті інформаційного простору, що відповідає вибраній предметній галузі. Основними завданнями, що стоять перед розробниками сценарно-аналітичних моделей стоять наступні завдання:

- спрощення реалізації аналітичної складової сценарних моделей;
 - обґрунтування складу і структури моделі, яка розробляється, виходячи з вимог до сценаріїв, які є кінцевим результатом моделювання;
 - адаптації різноманітних програмних засобів обробки даних до спеціальних для можливості отримання цілісного моделюючого комплексу;
- скорочення часу розробки сценаріїв, протягом якого необхідно завершити процедури моделювання та підготовки до прийняття рішення.

Виклад основного матеріалу дослідження. Для вирішення вищезазначених задач сценарного аналізу необхідно визначити основні поняття, стани, переходи між якими від початкового поняття до цілі являють собою окремі сценарії. Для визначення множини опорних понять пропонується застосування методів лінгвістичного аналізу в межах предметної області.

Для цього спочатку виконується формування репрезентативного текстового корпусу із вибраного для аналізу масиву інформаційних повідомлень, після чого здійснюється виявлення термінологічної основи майбутньої сценарної моделі. Отримані терміни складаються з одного (уніграма), двох (біграма) чи трьох (триграма) слів. Ці терміни розглядаються як основа побудови відповідної мережі, яка описується за допомогою матриці інцидентності.

Зв'язки в цій мережі можуть формуватися різними шляхами, зокрема, між двома окремими термінами у даному випадку встановлюється зв'язок, якщо вони входять до одного речення (Language-network); також може будуватися мережа природних ієрархій [8].

Якщо розглядати зв'язки між окремими поняттями, що відповідають вибраним термінам, як інформаційні впливи або дії для подальшого формування їх послідовності у вигляді сценаріїв, то вони мають бути спрямовані. Напрямок зв'язків між окремими поняттями, що визначаються термінами, формуються виходячи із трьох принципів: 1) якщо один термін є складовою іншого терміна (наприклад, біграма входить у склад триграми), то напрямок зв'язку йде від першого терміна до другого; 2) якщо один термін як вузол мережі має більший ступінь (кількість суміжних ребер), то напрямок зв'язку визначається від нього; 3) лінгвістичний принцип слідування, наприклад у випадку англомовних джерел напрямок зв'язку визначається порядком появи термінів у реченнях. Вага зв'язків може відповідати частотності спільного застосування в окремих реченнях термінів, що ними зв'язуються.

В рамках моделі, що розглядається, зважування окремих понять мережі, що відповідають термінам, здійснюється за допомогою модифікованого алгоритму HITS (Hyperlink-Induced Topic Search). Для всіх понять обраховуються взаємопов'язані показники авторства (auth) та посередництва (hub) з використанням модифікованого алгоритму HITS [9]. Оскільки у отриманій мережі можуть існувати двонаправлені зв'язки між вузлами, а також ребра, що можуть мати різні показники ваги (спрощена модель передбачає тільки позитивні значення), оригінальний HITS застосовуватися напряму не може. Модифікація алгоритму передбачає при розрахунку значення ваги кожного вузла враховувати вагу ребер, а для зниження впливу нерівномірності розподілу значень ваги на результат помножити на монотонно-зростаючою та менш крутою за лінійну функцію f . Прикладом такої функції є функція логарифма.

У результаті обчислення показники авторства та посередництва відбувається за алгоритмом, що визначається таким чином. Спочатку всім вузлам A_i , $i = \overline{1, N}$, де N – кількість вузлів мережі, призначається однакова вага авторства:

$$A_i = \frac{1}{N},$$

після чого здійснюються ітеративні обчислення за формулою:

$$hub(A_i) = \sum_{A_i \rightarrow A_j} auth(A_j) f(w_{ij}),$$

$$auth(A_i) = \sum_{A_j \rightarrow A_i} hub(A_j) f(w_{ij}).'$$

У наведених формулах w_{ij} – це вага зв'язків між вузлами A_i та A_j відповідно. На кожному кроці наведеного процесу здійснюється нормування обчислених значень:

$$hub(A_i) := \frac{hub(A_i)}{\sum_{j=1}^N hub(A_j)},$$

$$auth(A_i) := \frac{auth(A_i)}{\sum_{j=1}^N auth(A_j)}.$$

На основі обчислення показників авторства і посередництва понять, здійснюється ранжирування відповідних термінів, прийняття рішення щодо включення або виключення їх з моделі у якості вузлів мережі. На цьому етапі бажаним (але не необхідним) є залучення експертів для остаточної побудови мережі понять.

Надання власної семантики мережі та її подальше корегування виконується за допомогою Protégé – сучасного програмного пакета з відкритим вихідним кодом, призначений для розробки онтологій і систем управління знаннями. Опис об'єктів мережі використовується засобами OWL (Ontology Web Language) – стандартної мови для онтологій, побудованої на основі стандартів RDF і RDFS. OWL використовує ширшу систему типів порівняно з попередніми подібними мовами. Це дозволяє визначати унікальність та визначати еквівалентність термінів предметної області.

Надалі, розрахунок взаємного впливу вузлів мережі A_i та A_j через сценарій з номером m довжиною M здійснюється за таким принципом: вплив зворотно пропорційний відстані $d_{ij}^{(m)}$ в мережі між цими термінами і прямо пропорційний середній вазі транзитних термінів

$$\langle V_{ij} \rangle^{(m)} = \frac{1}{M} (V_{ik_1} + V_{k_1k_2} + \dots + V_{k_Mj})$$

$$S^{(m)} = \frac{1}{d_{ij}^{(m)}} \langle V_{ij} \rangle^{(m)}$$

З урахуванням цього значення здійснюється вибір найкращих сценаріїв за наступною методикою (див. рис. 1):

- визначення цілі реалізації сценарію;
- аналіз існуючих елементів мережі понять (онтології);
- визначення учасників сценарію;
- визначення послідовності інформаційної взаємодії між учасниками;
- призначення змісту і ресурсів для виконання;
- визначення умов початку і завершення кожного кроку сценарію (логічних залежностей);
- у разі необхідності повернення до початку алгоритму.

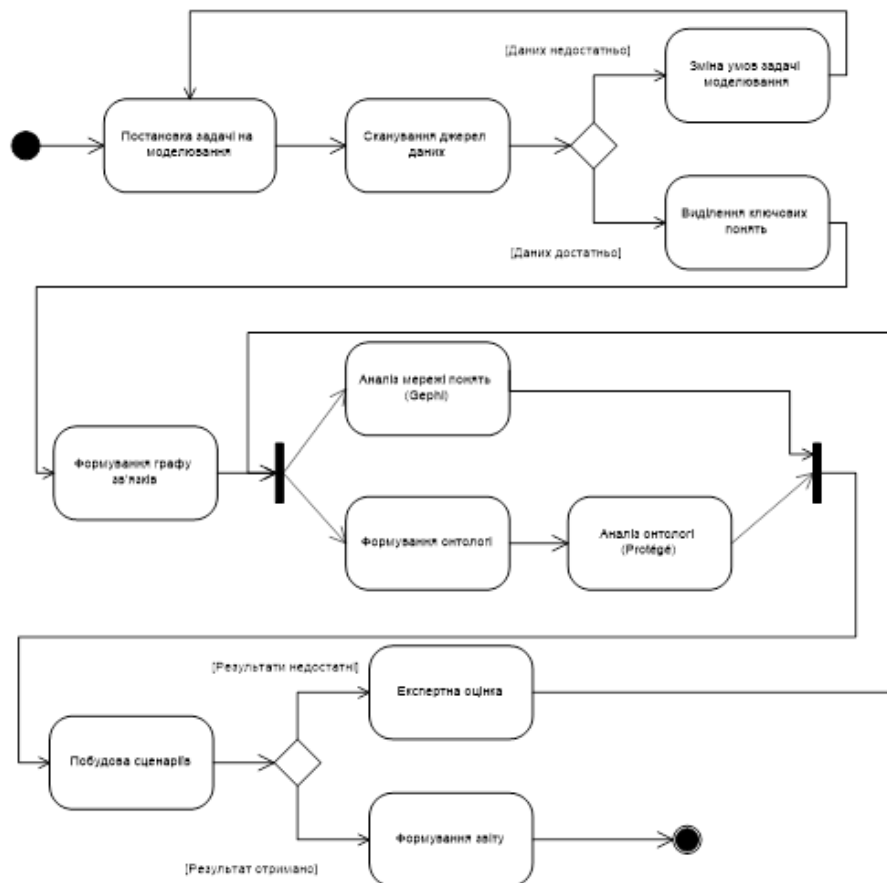


Рисунок 1 – Представлення методики діаграмою діяльності

Модель предметної області. Відповідну модель предметної області (онтологію або когнітивна карта) можна представити у вигляді множини:

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

- де C – підмножина класів, що описують поняття деякої предметної або проблемної області;
 S – 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.

Отримана онтологія являє собою знаковий орієнтований граф: $G = \langle A, E \rangle$, де A – множина вершин $A_i \in A, i = 1, 2, \dots, M$, які є елементами досліджуваної системи; E – множина дуг $e_{ij} \in E, i, j = 1, 2, \dots, N$, які відображують взаємозв'язок між вершинами A_i та A_j (див. рис. 2).

Практична реалізація. Для перевірки методики було проведено формування текстового корпусу з інформаційних Internet за темою BREXIT. Дослідження було обмежене лише австралійськими сайтами і склало загалом 257 документів.

За результатами досліджень було виявлено терміни з більшим значенням auth, зокрема: “minister theresa may”, “british prime minister”, “comey confirms fbi”, “minister david cameron”, “leader nigel Farage”, “pull brexit trigger”;

Терміни з найбільшим значенням hub: “brexit”, “eu”, “uk”, “may”, “european”, “minister”.

За результатами описаних досліджень побудовано сценарну модель інформаційного простору (див. рис 3). При формуванні сценаріїв експерт вибирає, які вершини в отриманій моделі слід виділити, та які зв'язки між вершинами враховувати, а які – ні.

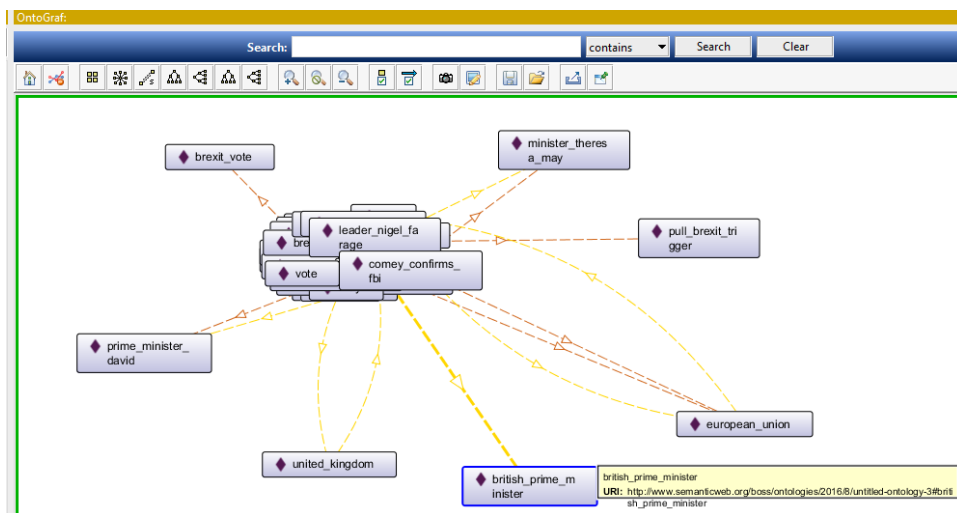


Рисунок 2 – Приклад отриманої онтології у середовищі Protégé

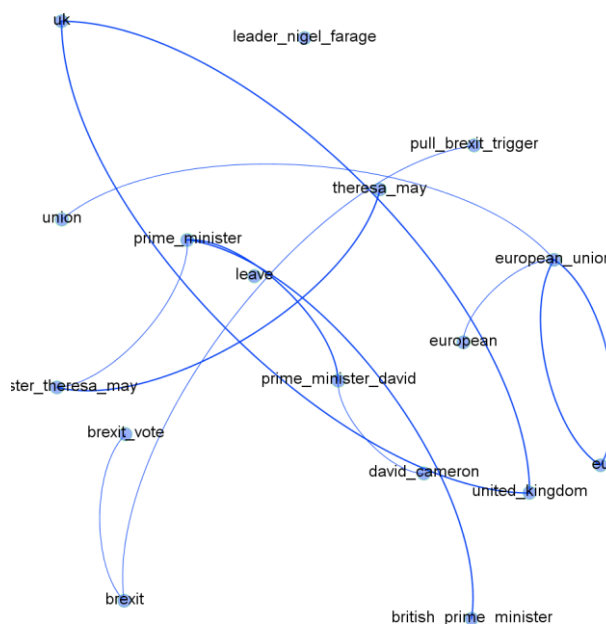


Рисунок 3 – Приклад отриманого графа (Gerphi)

Доцільність подальшого розвитку запропонованої методики визначається її здатністю значно прискорити аналіз великих (потенційно необмежених) масивів складної інформації, високою гнучкістю та відносною простотою реалізації.

Висновки. Запропонована вище методика дозволяє в автоматизованому режимі на базі аналізу вхідного пакету документів вирішити задачу формування та ранжирування сценаріїв інформаційного впливу на об'єкти, які відповідають вибраним ключовим поняттям. Візуалізація предметної області дозволяє аналітику досліджувати отримані сценарії та формувати результати у зручному для прийняття рішень вигляді. Методика може використовуватись для задач сценарного моделювання у галузях комп'ютерної безпеки, дослідження інформаційних впливів та підтримки прийняття рішень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] М.З. Згуровский, та Н.Д. Панкратова, *Технологическое предвидение*. Київ, Україна: Видавництво "Політехніка", 2005.
- [2] А.Г. Додонов, Д.В. Ландэ, та Т.В. Коваленко, "Модели предметных областей в системах поддержки принятия решений на основе мониторинга информационного пространства", на *VI Междунар. науч.-техн. конф. Открытые семантические технологии проектирования интеллектуальных систем*, Минск, 2016, с. 171-176.
- [3] М.З. Згуровский, та Н.Д. Панкратова, *Основи системного аналізу*. Київ, Україна: Видавнича група BHV, 2007.
- [4] P.J.H. Schoemaker, "Multiple scenario development: its conceptual and behavioral foundation", *Strategic Management Journal*, Vol. 14, Iss. 3, pp. 193-213, 1993.
doi: 10.1002/smj.4250140304.
- [5] B. Hassani, *Scenario analysis in risk management*. Gewerbestrasse, Switzerland: Springer International Publishing, 2016.
doi: 10.1007/978-3-319-25056-4.
- [6] В.Л. Шульц, В.В. Кульба, А.Б. Шелков, и В.В. Чернов, "Сценарный анализ эффективности управления информационной поддержкой государственной политики России в Арктике", *Национальная безопасность / nota bene*, № 6 (17), с. 104-137, 2011.
- [7] D. Amyot, and A. Eberlein, "An evaluation of scenario notations and construction approaches for telecommunication", *Telecommunications Systems Journal*, Vol. 24, Iss. 1, pp. 61-94, September 2003.
- [8] Д.В. Ландэ, и А.А. Снарский, "Подход к созданию терминологических онтологий", *Онтология проектирования*, № 2 (12), с. 83-91, 2014.
- [9] J. Kleinberg, "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, Vol. 46, Iss. 5, pp.604-632.
doi: 10.1145/324133.324140.

Стаття надійшла до редакції 21 серпня 2017 року.

REFERENCE

- [1] M.Z. Zgurovsky, and N.D. Pankratova, *Technological foresight*. Kyiv, Ukraine: Politekhnik Publishing House, 2005.
- [2] A.G. Dodonov, D.V. Lande, and T.V. Kovalenko, "Models of subject domains in decision-making support system on the basis of information space monitoring", in *Proc. VIth international conference Open Semantic Technologies for Intelligent Systems*, Minsk, 2016, pp. 171-176.
- [3] M.Z. Zgurovsky, and N.D. Pankratova, *The foundations of system analysis*, Kyiv, Ukraine: Publishing group BHV, 2007.
- P.J.H. Schoemaker, "Multiple scenario development: its conceptual and behavioral foundation", *Strategic Management Journal*, Vol. 14, Iss. 3, pp. 193-213, 1993.
doi: 10.1002/smj.4250140304.

- [5] B. Hassani, *Scenario analysis in risk management*. Gewerbestrasse, Switzerland: Springer International Publishing, 2016.
doi: 10.1007/978-3-319-25056-4.
- [6] V.L. Shulz, V.V. Kulba, A.B. Shelkov, and V.V. Chernov, "Scenario analysis of the efficiency of management of the information support of the state policy of Russia in the Arctic territory", *National Security / nota bene*, № 6 (17), pp. 104-137, 2011.
- [7] D. Amyot, and A. Eberlein, "An evaluation of scenario notations and construction approaches for telecommunication", *Telecommunications Systems Journal*, Vol. 24, Iss. 1, pp. 61-94, September 2003.
- [8] D.V. Lande, and A.A. Snarskij, "Approach to the creation of terminological ontologies", *Ontology of Designing*, № 2 (12), pp. 83-91, 2014.
- [9] J. Kleinberg, "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, Vol. 46, Iss. 5, pp.604-632.
doi: 10.1145/324133.324140.

ДМИТРИЙ ЛАНДЭ,
АНДРЕЙ БОЙЧЕНКО

ИССЛЕДОВАНИЕ СЦЕНАРИЕВ РАЗВИТИЯ СИТУАЦИИ НА ОСНОВЕ АНАЛИЗА ИНФОРМАЦИОННОГО ПРОСТРАНСТВА

Предлагается методика разработки сценариев развития ситуации, которая позволяет в автоматизированном режиме решать задачу формирования и ранжирования сценариев осуществления воздействий на объекты, которые соответствуют выбранным ключевым понятием. Из массива документов, относящихся к определенной предметной области выделяются ключевые понятия, которыми могут быть отдельные слова, биграммы или триграммы. Между полученными ключевыми понятиями, с помощью весовых критериев, формируется граф предметной области. Путем экспертной оценки выполняется редактирование графа и присвоение числовых значений его ребрам, что позволяет построить сеть понятий предметной области. В рамках рассматриваемой модели взвешивание отдельных узлов сети, соответствующих терминам, осуществляется с помощью модифицированного алгоритма HITS (Hyperlink-Induced Topic Search). Модификация алгоритма предусматривает, что при расчете значения веса каждого узла учитывается вес ребер, а для уменьшения влияния неравномерности распределения значений веса на результат, вес узла умножается на монотонно-возрастающую, менее крутую, чем линейная, функцию. Построение онтологии позволяет учесть основные факторы и возможные взаимосвязи между ними, и является базой для разработки компьютерных сценариев развития ситуации. Для исследования используются пакет визуализации и моделирования графов Gephi, пакет для построения онтологий Protégé и собственный набор специально разработанных модулей на Perl. Для проверки методики было проведено формирование текстового корпуса путем сканирования Интернет по теме BREXIT. Визуализация предметной области позволяет аналитику исследовать полученные сценарии и формировать результаты в удобном для принятия решений виде. Методика может использоваться для задач сценарного моделирования компьютерной безопасности, исследования информационных воздействий и поддержки принятия решений.

Ключевые слова: сценарный анализ, аналитическая деятельность, информационные операции, сеть терминов, сценарное моделирование.

DMYTRO LANDE,
ANDRII BOICHENKO

SCENARIO RESEARCH BASED ON ANALYSIS OF INFORMATION SPACE

A method to developing scenarios for the development of a situation is proposed, which allows solving a problem of formation and ranking scenarios for implementing impacts on objects that correspond to the selected key concept in the automatic processing mode. From the array of

documents belonging to a certain subject area, identified key concepts, which can be separate words, bigrams or trigrams. The subject area graph formed, between the received key concepts, using weight criteria. By expert evaluation, the graph edited and numerical values assigned to his edges, to form a network of subject domain concepts. Within the framework of the model, the weighing of individual network nodes corresponding to the terms is done using a modified algorithm HITS (Hyperlink-Induced Topic Search). Modification of the algorithm provides that when calculating the weight value of each node, the weight of the edges taken into account. To reduce the effect of the uneven distribution of weight values on the result, the weight of the node is multiplied by a monotonically increasing, less steep than a linear function. Building an ontology allows taking into account general factors and interrelations between them and develop several computer scenarios considering alternative possible outcomes. Open-source network analysis software package Gephi, ontology development environment Protégé and an original package of specially developed Perl modules are used for simulation and visualization. To test the method a network of domain concepts was formed by scanning the Internet on the topic BREXIT and then analyzed. Visualization of the subject domain allows the analyst to explore the resulting scenarios and generate results in a convenient form for decision-making. The method can be used in scenario modeling, information security analysis and decision-making support.

Keywords: scenario analysis, analytical activity, information operations, network of terms, scenario modeling.

Дмитро Володимирович Ланде, доктор технічних наук, старший науковий співробітник, завідувач відділом спеціалізованих засобів моделювання, Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна.

E-mail: dwlande@gmail.com.

Андрій Васильович Бойченко, науковий співробітник, Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна.

E-mail: boychenko.a@gmail.com.

Дмитрий Владимирович Ландэ, доктор технических наук, старший научный сотрудник, заведующий отделом специализированных средств моделирования, Институт проблем регистрации информации Национальной академии наук Украины, Киев, Украина.

Андрей Васильевич Бойченко, научный сотрудник, Институт проблем регистрации информации Национальной академии наук Украины, Киев, Украина.

Dmytro Lande, doctor of technical science, senior researcher, head of the specialized modeling tools department, Institute for information recording of National academy of science of Ukraine, Kyiv, Ukraine.

Andrii Boichenko, senior researcher, Institute for information recording of National academy of science of Ukraine, Kyiv, Ukraine.