

УДК 519.816

**О. В. Андрійчук**

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

## Метод змістової ідентифікації об'єктів баз знань систем підтримки прийняття рішень

*Запропоновано метод змістової ідентифікації об'єктів баз знань систем підтримки прийняття рішень, що дозволяє підвищити адекватність моделей предметних областей. Основна ідея методу полягає в тому, що кожному об'єкту бази знань зіставляється кортеж ключових слів, за якими відбувається змістова ідентифікація. Важливість ключових слів визначається експертним шляхом. Описано авторський підхід до підвищення адекватності моделей предметних областей, та метод змістової ідентифікації, а також можливі варіанти його застосування.*

**Ключові слова:** система підтримки прийняття рішень, модель предметної області, база знань, змістова ідентифікація, інженерія знань, експертне оцінювання, нечіткий топологічний простір, метрика.

### Вступ

Діяльність будь-якого керівника повсякденно пов'язана з необхідністю прийняття рішень. Прийняття рішення є особливим видом діяльності, який полягає у формуванні множини варіантів рішення (альтернатив) і подальшій оцінці їхньої відносної ефективності та розподілі ресурсів між варіантами рішення, відповідно до їхніх оцінок. Більш простими типами рішень є прийняття чи відхилення альтернативи, вибір найкращої альтернативи із заданої множини, ранжування альтернатив.

Під час прийняття комплексних рішень часто виникає потреба у врахуванні численних (десятків чи сотень) взаємопов'язаних факторів, які складним чином взаємодіють між собою. Задля забезпечення високого професійного рівня рішень, необхідна інтеграція знань багатьох спеціалістів-експертів. Утім, людина, через психофізіологічні обмеження, здатна одночасно оперувати лише 7–9 об'єктами [1]. Для подолання цього обмеження застосовують системи підтримки прийняття рішень (СППР) (рис. 1).



Рис. 1. Функціональна схема СППР

При вирішенні задач у слабо структурованих предметних областях, де на теперішній час усе ширше використовуються СППР, набуває актуальності задача підвищення адекватності моделі предметної області для підвищення достовірності рекомендацій, що виробляються за допомогою СППР. Невід'ємною складовою СППР є інформація, що отримується від експертів у вигляді найменувань об'єктів, сформульованих природною мовою, тому дуже важливо для адекватного врахування колективної думки експертів при встановленні зв'язків між такими об'єктами в базах знань (БЗ) СППР однозначно ідентифікувати ці об'єкти (тобто формально представити мовою понять, що притаманні даній предметній області). Для опису складних предметних областей створюються БЗ великих розмірів. Саме в таких ситуаціях питання однозначної ідентифікації об'єктів цих БЗ постає особливо гостро.

На теперішній час при побудові моделей предметних областей у СППР значну увагу дослідники приділяють методам отримання і обробки експертних оцінок. Ці методи досить широко представлені та продовжують висвітлюватися в наукових публікаціях. Таким чином, зазвичай приділяється увага оцінюванню об'єктів із уже побудованих БЗ. У процесі визначення та ідентифікації цих об'єктів експертами, організатори експертиз до теперішнього часу переважно поклалися на чіткість та однозначність їхніх формулювань. При зростанні кількості об'єктів у БЗ СППР, особливо при застосуванні групових методів побудови БЗ, питання безпомилкової ідентифікації стає все більш актуальним.

У ході побудови та супроводу БЗ, найбільш адекватного відображення інтегрованої думки експертів можна досягти шляхом виключення помилкового повторного введення ідентичних за змістом (але різних за формулюванням) назв об'єктів. Для цього доцільно здійснювати пошук споріднених за змістом формулювань. Підвищення ефективності використання знань, отриманих від експертів, стає можливим завдяки повторному використанню раніше побудованих БЗ, але при цьому процес об'єднання БЗ потрібно автоматизувати, оскільки у достатньо великих БЗ,

та, особливо у БЗ, побудованих експертними групами різного профілю, існує досить значна вірогідність помилки.

## Ознаки слабо структурованих предметних областей

СППР використовуються для управління об'єктами в слабо структурованих предметних областях [2]. Ці предметні області мають наступні особливості [3]: унікальність, відсутність цілі функціонування, яку можна формалізувати, відсутність оптимальності, динамічність, неповнота опису, наявність людського фактору, неможливість побудувати аналітичну модель, відсутність еталонів, велика розмірність простору рішення (рис. 2). Прикладами слабо структурованих предметних областей є: управління, сталий розвиток, автоматизація та інформатизація, кадрові рішення, комерційна (ділова) сфера, маркетинг, банківська галузь, промисловість, енергетика, транспортна сфера, медицина та інші.



Рис. 2. Особливості слабо структурованих предметних областей

Розглянемо докладніше перераховані ознаки слабо структурованих предметних областей.

Об'єкти в слабо структурованих предметних областях є унікальними. Фактично, системи управління для цих областей створюються одноразово, з метою вирішення реальних задач; перенесення таких моделей на інші об'єкти потребує великих витрат чи просто є неможливим.

У системах, що не створюються людиною (наприклад, екологічних), а також у суспільних та адміністративних системах відсутня ціль функціонування, яку можна формалізувати. Ціллю функціонування таких систем є їхня працездатність у цілому, підтримка деяких параметрів у заданих межах, проте формалізувати таку ціль у вигляді деякого критерію, як правило, неможливо. Наприклад, в екологічній системі всі фактори, які впливають на її функціонування, настільки численні, а зв'язки між ними настільки складні й неочевидні, що задати певну функцію для опису цілі її функціонування, в принципі неможливо.

Через відсутність цілі функціонування, яка піддається формалізації, неможливо побудувати функцію, оптимізація якої забезпечила б найкращий режим функціонування об'єкта. Об'єктивної функції оптимізації не існує, можна вказати тільки окремі фактори, які можливо оптимізувати. Втім, неможливо оптимізувати кожен з цих факторів окремо, оскільки вони тісно пов'язані між собою, і їхні зв'язки будуть порушені в процесі оптимізації. Це може привести до порушення процесу, що регулює або підтримує систему в стійкому стані відносно змінного навколишнього середовища, в якому ця система функціонує, та, можливо, приведе до катастрофічних явищ і незворотних змін у системі.

Оскільки в слабо структурованій предметній області відсутня ціль функціонування, яку можна формалізувати, а також неможливо побудувати функцію, оптимізація якої забезпечить найкращий режим функціонування об'єкта, то неможливо побудувати й аналітичну модель цієї предметної області.

Динамічність пов'язана з тим, що структура та функціонування об'єкта змінюються з часом, тобто об'єкт еволюціонує. Управління такими системами повинно бути адаптивним, здатним змінюватися при зміні об'єкта.

Неповнота опису пов'язана з неточністю, неповнотою, невизначеністю та недостовірністю даних, що описують об'єкт.

Характеристики об'єктів проблематично кількісно описати, тому в слабо структурованих предметних областях недоцільно говорити про існування еталонних значень цих характеристик.

Велика розмірність простору рішень зумовлена великою кількістю та різномірністю критеріїв, які характеризують предметну область.

Об'єктами управління можуть бути люди, які мають свободу волі. Передбачити поведінку людини, як об'єкта управління чи компоненту системи часто буває неможливо. Людина діє в системі, враховуючи свої особисті цілі та інтереси. Тому при моделюванні об'єкта управління поведінку людей важко враховувати.

Описані вище властивості слабо структурованих предметних областей призводять до того, що чи не єдиним джерелом інформації у них являються експерти. Засобами СППР моделюється предметна область, та на основі цього моделювання особа, що приймає рішення, отримує відповідні рекомендації. Тому актуальним питанням є підвищення адекватності моделей предметних областей.

### **Особливості моделей представлення знань у слабо структурованих предметних областях**

Розглянемо особливості, які повинні бути враховані в моделях подання знань у слабо структурованих предметних областях. Сутність методології математично-

го моделювання [4] полягає в переході від досліджуваного об'єкта до його математичної моделі та подальшому дослідженні моделі в рамках обчислювального експерименту за допомогою програмно-реалізованих обчислювально-логічних алгоритмів. Це доцільно у випадках, коли експеримент з реальним об'єктом небезпечний, дорогий, проходить у незручному масштабі простору та часу, неможливий, неповторний, не наочний тощо.

У СППР експертні знання формалізуються за допомогою моделей подачі знань. Це означає, що більша адекватність моделей предметних областей може бути досягнута тією моделлю подачі знань, яка більш повно враховує особливості знань. Знання мають наступні особливості, що відрізняють їх від традиційних даних [5]:

- внутрішню інтерпретованість;
- структурованість;
- зв'язність;
- семантичну метрику;
- активність.

Внутрішня інтерпретованість знання означає, що кожна інформаційна одиниця повинна мати унікальне ім'я, за яким інформаційна система її знаходить, а також відповідає на запити, в яких це ім'я згадується. Структурованість означає, що знання повинні мати гнучку структуру, одні інформаційні одиниці можуть включатися до складу інших (змістові відношення типу «частина – ціле», «елемент – клас», «рід – вид» та інші [3]). Зв'язність знання означає, що в інформаційній системі повинна бути передбачена можливість встановлення різних типів зв'язків між різними інформаційними одиницями (причинно-наслідкові, просторові та ін.). На множині інформаційних одиниць корисно задавати відношення, які характеризують ситуаційну близькість цих одиниць у вигляді семантичної метрики. Активність знання означає, що виконання програм в інтелектуальній системі повинно ініціюватися поточним станом бази знань.

Також часто відокремлюють інші властивості знання, наприклад шкальованість, яка означає, що формально неоднакові поняття насправді відображаються на одній і тій самій шкалі понять, різні поділки якої відповідають інтенсивності прояву одного й того самого фактора. Наприклад, температура може бути високою або низькою, і це породжує такі поняття, як «холодно», «тепло», «гаряче».

### **Врахування особливостей знань у моделі подачі знань «граф ієрархії цілей»**

Розглянемо як враховуються вищеописані особливості знань у моделі подачі знань «граф ієрархії цілей». У рамках цієї моделі шляхом експертного оцінювання будується ієрархія цілей або БЗ, що задана орієнтованим графом типу мережа (рис. 3). Вершини графа є цілями або об'єктами БЗ. Дуги відображують вплив досягнення одних цілей на досягнення інших: дуги (ребра), що виходять з цілей, входять у їхні безпосередні надцілі. Цілі можуть бути кількісними та якісними.

При побудові ієрархії цілей використовується метод ієрархічного цільового оцінювання альтернатив [2], у рамках якого будують ієрархію цілей, визначають відповідні часткові коефіцієнти впливу та оцінюють відносну ефективність проек-

тів. Спочатку формулюється головна ціль проблеми і можливі варіанти її досягнення (проекти), що, зрештою, будуть оцінюватися. Потім у два етапи здійснюється побудова графа ієрархії цілей: «згори-донизу» та «знизу-догори» [2]. Етап «згори-донизу» полягає у послідовній декомпозиції кожної цілі на підцілі чи проекти, досягнення яких впливає на досягнення даної цілі чи проекту. Головна ціль підлягає декомпозиції на більш прості складові — цілі, які впливають на неї. Далі ці сформульовані цілі також підлягають декомпозиції на більш прості складові — підцілі, які потім також будуть підлягати декомпозиції. При декомпозиції цілі до переліку підцілей, що впливають на її досягнення, окрім щойно сформульованих підцілей, можуть бути включені вже наявні у ієрархії цілі, які були сформульовані раніше, при декомпозиції інших цілей. Процес декомпозиції зупиняється, коли множини підцілей, що впливають на цілі, які розкриваються, будуть складатися з уже розкритих цілей та варіантів рішень, що оцінюються. Таким чином, після завершення процесу декомпозиції не залишається нерозкритих цілей. Етап «знизу-догори» полягає у визначенні для кожного проекту та кожної цілі всіх надцілей, на досягнення яких впливає досягнення даної цілі чи проекту.

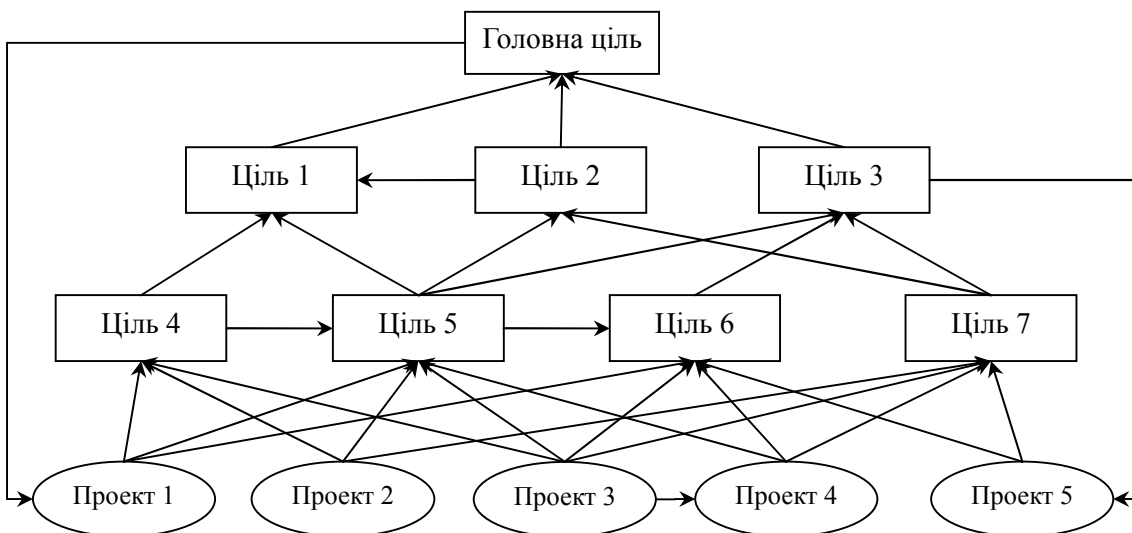


Рис. 3. Ієрархія цілей

Як було сказано вище, експерти будують ієрархію цілей, що задається орієнтованим графом типу мережа (рис. 3). Його вершини позначені формулюваннями цілей. Наявність дуги, що йде з однієї вершини в іншу (цілями), означає наявність впливу досягнення однієї цілі на досягнення іншої. У результаті описаного процесу побудови ієрархії цілей, відповідний граф, що її задає, є одностороннім зв'язним, тому що з будь-якої вершини графа існує шлях до вершини, яка позначає головну ціль. Кожній цілі поставлено у відповідність показник ступеня досягнення від 0 до 1. Цей показник дорівнює 0 за відсутності будь-якого прогресу в досягненні цілі, а при повному її досягненні — дорівнює 1. Кожний вплив однієї цілі на іншу може бути як позитивним, так і негативним, а його ступінь виражається від-

повідним показником — частковим коефіцієнтом впливу (ЧКВ). У методі цільового динамічного оцінювання альтернатив (МЦДОА) також враховується часова затримка впливу [2]. Для проектів враховується тривалість виконання. Існує також підхід (оснований на МЦДОА), який дозволяє відобразити складну динамічну поведінку об'єкта, що еволюціонує з часом [3]. ЧКВ визначаються експертним шляхом, а для підвищення достовірності експертного оцінювання застосовують метод парного порівняння.

Як уже зазначалося, людина має певні психофізіологічні обмеження на кількість блоків інформації, що вона може одночасно тримати в короткочасній пам'яті [1]. Тобто більшість людей з нормальними середніми здібностями, зазвичай, можуть одночасно тримати у пам'яті 5 об'єктів, з високими здібностями — 7, а при тренуванні людини на певний тип інформації, чи при особливо високих здібностях — 9. При цьому, у складних ієрархіях цілей у процедурі «згори-донизу» у кожній цілі може бути більш ніж 9 підцілей та проектів. Аналогічно в процедурі «знизу-догори» для кожного проекту та кожної цілі може бути більш ніж 9 надцілей. Таким чином, при досить значній кількості цілей в ієрархії, для людини виникають певні складнощі пошуку цілей/проектів, схожих за змістом, а також можливість помилкового пропуску впливу експертом.

З'ясуємо, як врахована внутрішня інтерпретованість знань у моделі подачі знань «граф ієрархії цілей». У процесі визначення та ідентифікації цілей ієрархії експерти, в основному, покладаються на чіткість та однозначність їхніх формулювань. Але при досить великих розмірах БЗ існує можливість помилкового введення однакових за змістом цілей. Зокрема, при об'єднанні декількох ієрархій, що сформульовані експертними групами різної спеціалізації, можлива ситуація, коли різними формулюваннями описується одна й та сама ціль. Отже, внутрішня інтерпретованість знань у моделі граф ієрархії цілей не завжди може бути врахована.

Процес декомпозиції цілей на підцілі, який було описано вище, дозволяє досягнути структурованості та зв'язності знань. Активність знань досягається шляхом використання в ієрархії кількісних цілей [2]. Шкальованість знань досягається шляхом використання експертами при оцінюванні вербальних шкал [2, 6], які задіяні, зокрема, у відповідній адаптивній технології експертного оцінювання [7, 8].

Отже в моделі подачі знань «граф ієрархії цілей» відсутня семантична метрика та не завжди може бути врахована внутрішня інтерпретованість знань. Для врахування цих особливостей знань пропонується використовувати метод змістової ідентифікації об'єктів БЗ СППР [5, 9].

Слід зазначити, що граф ієрархії цілей подібний до прогнозного графа, для якого також характерна описана особливість — відсутність семантичної метрики. У доповіді про використання прогнозного графа для прогнозування і керування науковими дослідженнями [10] В.М. Глушков, серед наукових проблем математичної обробки інформації, відзначав важливе питання ототожнення близьких за значенням проблем, яке не було вирішене. Питання полягало у визначенні семантично подібних науково-технічних проблем, адже завжди потрібно ототожнювати близькі за змістом проблеми, які вирішуються більш-менш схожими засобами. З іншого боку, існують «тонкі» приклади, коли проблеми зовсім мало відрізняються в постановці і, начебто, їх слід ототожнити, а в дійсності, за методами розв'язання вони цілком різні, і тому їх треба, навпаки, розділити. Розв'язання цього питання

було актуальним для ефективного використання методики прогнозного графа, зокрема, для запобігання розростанню графа до нескінченності. Проблеми мали відрізнятися якісно, а не тільки за деякими параметрами. Передбачалося використання спеціального органу експертів, на зразок інституту наукової інформації, з кваліфікованими спеціалістами, який міг би розв'язувати проблеми такого роду. Якщо ж відповідний орган був не в змозі чітко класифікувати проблеми за змістом, передбачалося звернення до спеціалістів більш високої кваліфікації для отримання остаточного судження.

### Сутність підходу

Змістова ідентифікація полягає у пошуку однозначної відповіді на питання: чи співпадають об'єкти БЗ між собою за змістом у межах заданого відносного відхилення? В основі методу лежить підхід, який, у загальних рисах, певною мірою перекликається з підходом, що застосовується при оцінці компетентності експертів [2]. Сутність підходу полягає в наступному: для відображення змісту, кожному об'єкту БЗ ставиться у відповідність кортеж ключових слів (КС) з відповідними коефіцієнтами важливості. Цей кортеж має мінімальну потужність, **необхідну** для збереження унікальності змісту цілі. Коефіцієнти важливості КС для кожної цілі ієрархії визначаються експертним шляхом і представляють собою відповідні дійсні числа. Змістова ідентифікація відбувається за подібністю коефіцієнтів важливості КС об'єктів БЗ (див. нижче).

### Постановка задачі змістової ідентифікації

Задачу змістової ідентифікації об'єктів БЗ СППР можна звести до наступної постановки.

*Дано:*

1) множина всіх об'єктів БЗ  $G = \| G_i \|_{i=\overline{1,m}}$ , де  $m$  — кількість об'єктів у БЗ; кожний об'єкт БЗ  $G_i = \| g_{ij} \|_{i=\overline{1,m}; j=\overline{1,n}}$ , де  $g_{ij}$  — ваговий коефіцієнт КС з номером  $j$  у об'єкті БЗ з номером  $i$ ;  $n$  — загальна кількість КС, причому  $0 \leq g_{ij} \leq 1$ ,  $i = \overline{1,m}$ ,  $j = \overline{1,n}$ ,  $\sum_{j=1}^n g_{ij} = 1$ ;

2) симетрична матриця відмінностей КС між собою  $D$ , де  $D = \| d_{ij} \|_{i=\overline{1,n}; j=\overline{1,n}}$ ,  $0 \leq d_{ij} \leq 1$  &  $d_{ij} = d_{ji}$ ,  $i = \overline{1,n}$ ,  $j = \overline{1,n}$ ;

3) відносне відхилення  $\varepsilon$ , де  $0 < \varepsilon \leq 1$ .

*Потрібно визначити:* які об'єкти БЗ є подібними до об'єкта  $G_i$  в межах заданого відносного відхилення  $\varepsilon$ ?

### Сутність методу змістової ідентифікації

Розглянемо сутність методу. При формулюванні цілі експертом, для характеристики її змісту та забезпечення однозначної відповідності змісту уявленню експерта, задається кортеж КС з відповідними коефіцієнтами важливості. Цей кортеж



КС із відповідними вагами може бути: а) заданий експертним шляхом; б) сформульований на основі нормативних документів; в) оснований на раніше побудованих ієрархіях. Кортеж КС об'єкта БЗ має бути мінімальної потужності і при цьому містити всю необхідну інформацію для його ідентифікації. Коефіцієнти важливості КС для об'єкта БЗ визначаються експертним шляхом. Вагові коефіцієнти КС можуть визначатися за матрицею парних порівнянь, наприклад, методами власного вектора [6], «трикутник», «квадрат» [2], комбінаторним [11, 12], чи іншими методами обробки експертної інформації. Будується загальна для предметної області матриця відмінностей КС, основана на порівняннях усіх КС між собою за семантичною подібністю.

Нехай  $K$  буде множина всіх КС БЗ. Тоді побудований вищеописаним чином об'єкт  $G_i$  є нечіткою підмножиною [13] множини  $K$  ( $G_i \subseteq K$ ). Значення вагових коефіцієнтів  $g_{ij}$  об'єкта  $G_i$  задають таблицю значень відповідної функції приналежності  $\mu_{G_i} \equiv g_{ij}$ . Таким чином, об'єкту  $G_i$  відповідає наступна сукупність пар:

$$\{(K_j, \mu_{G_i}(K_j)) \mid K_j \in K\} = \{(K_j, g_{ij}) \mid K_j \in K\}, \quad j = \overline{1, n},$$

де  $K_j$  —  $j$ -те КС.

$K_j$  теж можна розглядати як нечітку множину з наступною функцією приналежності  $\mu_{K_j}$ :

$$\mu_{K_j}(K_i) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}, \quad i = \overline{1, n}.$$

Візьмемо множину добутків нечітких множин  $K_j$  на відповідні коефіцієнти важливості КС  $\{g_{ij} \times K_j, j = \overline{1, n}, i = \overline{1, m}\}$  як бази [14] нечіткої топології [15]  $\tau$  на множині  $K$ . Згідно з [14], базою топології простору є така сукупність його відкритих підмножин, що будь-яка відкрита множина в просторі може бути представлена як сума деякого числа цих підмножин. Топологію простору можна задати, вказавши в цьому просторі деяку її базу; ця топологія співпадає із сукупністю множин, що можуть бути представлені як сума множин з цієї бази.

Пара  $(K, \tau)$  є скінченим нечітким топологічним простором за Чангом [16]:  $K$  — множина;  $\tau$  — нечітка топологія на ній, тобто деяке сімейство її нечітких підмножин, що задовольняє трьом аксіомам:

- 1)  $0, 1 \in \tau$ ;
- 2) якщо  $U, V \in \tau$ , то  $U \wedge V \in \tau$ ;
- 3) якщо  $U_i \in \tau$  для всіх  $i \in I$ , то  $\bigvee_i U_i \in \tau$ .

Тоді, оскільки всі об'єкти БЗ представляються сумою деякої скінченої кількості елементів бази топології  $\tau$ ,  $G \subseteq \tau$ , тобто якщо кожний об'єкт БЗ належить топології, то і вся множина об'єктів БЗ належить їй. Всі об'єкти БЗ є відкритими множинами, оскільки вони належать топології  $\tau$  [14]. Як сказано вище, для пред-

ставлення змісту кожного об'єкта БЗ використовується кортеж КС з відповідними нормованими ваговими коефіцієнтами, тому всі об'єкти БЗ лежать на симплексі в просторі  $(K, \tau)$ .

Для визначення змістової відстані потрібно ввести метрику в описаному вище просторі об'єктів БЗ  $(K, \tau)$ . Використаємо підхід, аналогічний знаходженню відстані за таблицею вимірів з ієрархічного кластерного аналізу [17]. Приклад таблиці вимірів представлений нижче.

Приклад таблиці вимірів

№ об'єкта	Характеристика $a$	Характеристика $b$	Характеристика $c$
1	0,5	0,3	0,2
2	0	0,6	0,4
3	0,7	0,15	0,15
4	0,1	0,9	0
Ваги характеристик	1,2	1,7	2,1

Головним моментом у кластерному аналізі вважається вибір метрики, від якого залежить кінцевий варіант розбиття об'єктів на групи при заданому алгоритмі розбиття. Вибір метрики впливає на форму кластерів [18]. Для кожної конкретної задачі метрика обирається з урахуванням головних цілей дослідження, фізичної та статистичної природи інформації, що використовується.

Для кожного об'єкта БЗ як виміри кожної з характеристик візьмемо відповідні значення вагових коефіцієнтів КС  $g_{ij}$ ,  $i = \overline{1, n}$ , як ваги кожної із характеристик — відповідно, сумарну відмінність кожного КС від решти інших КС БЗ  $\sum_{i=1}^n d_{ij}$ . Змістову відстань  $\rho$  між кожними двома об'єктами БЗ  $G_k$  та  $G_l$  знайдемо, наприклад, як зважену відстань Хемінга:

$$\rho(G_k, G_l) = \sum_{i=1}^n (|g_{ki} - g_{li}| \sum_{j=1}^n d_{ij}).$$

Доведемо виконання аксіом метрики [15] для введеної змістової відстані:

- 1)  $\rho(G_k, G_l) = 0 \Leftrightarrow G_k = G_l$  (аксіома тотожності);
- 2)  $\rho(G_k, G_l) = \rho(G_l, G_k)$  (аксіома симетрії);
- 3)  $\rho(G_k, G_l) \leq \rho(G_l, G_p) + \rho(G_p, G_l)$  (аксіома трикутника або нерівність трикутника).

Доведення виконання перших двох аксіом (тотожності та симетрії) очевидне, тому приведемо доведення виконання аксіоми трикутника:

$$\begin{aligned} \downarrow \rho(G_k, G_l) &= \sum_{i=1}^n (|g_{ki} - g_{li}| \sum_{j=1}^n d_{ij}) = \sum_{i=1}^n (|g_{ki} - g_{pi} + g_{pi} - g_{li}| \sum_{j=1}^n d_{ij}) \leq \\ &\leq \sum_{i=1}^n (|g_{ki} - g_{pi}| + |g_{pi} - g_{li}|) \sum_{j=1}^n d_{ij} = \sum_{i=1}^n (|g_{ki} - g_{pi}| \sum_{j=1}^n d_{ij}) + \sum_{i=1}^n (|g_{pi} - g_{li}| \sum_{j=1}^n d_{ij}) = \\ &= \rho(G_l, G_p) + \rho(G_p, G_l) \uparrow \end{aligned}$$

Таким чином, побудовано скінчений чангівський нечіткий топологічний простір об'єктів БЗ та на ньому введено метрику.

Тепер обчислимо вектор відстаней від об'єкта  $G_t$  до інших об'єктів БЗ  $v_t$ :

$$v_t = \| \rho(G_t, G_i) \|_{i=1, \overline{m}}.$$

Для нормування поділимо знайдений вектор  $v_t$  на його максимальний елемент  $\max_{i=1, m} \{v_i\} \neq 0$  та отримаємо вектор  $v_t^{norm}$ :

$$v_t^{norm} = \frac{v_t}{\max_{i=1, m} \{v_i\}}.$$

Нехай множина  $X_\varepsilon$  містить шукані об'єкти БЗ подібні до об'єкта  $G_t$  в межах заданого відносного відхилення  $\varepsilon$ :

$$X_\varepsilon = \{G_i \mid v_i^{norm} \leq \varepsilon, i = \overline{1, m}\}.$$

Розглянемо *приклад*. Нехай є 4 об'єкти БЗ, зміст яких заданий за допомогою 3-х КС. Візьмемо з наведеного вище прикладу таблиці вимірів (див. табл.) значення коефіцієнтів важливості КС і сумарних відмінностей кожного КС від решти інших КС БЗ. Визначимо, які об'єкти БЗ, є подібними до об'єкта  $G_1$  в межах заданого відносного відхилення  $\varepsilon = 0,33$ .

Знайдемо змістові відстані від об'єкта  $G_1$  до всіх об'єктів БЗ:

$$\begin{aligned} \rho(G_1, G_1) &= 0, \\ \rho(G_1, G_2) &= 1,53, \\ \rho(G_1, G_3) &= 0,6, \\ \rho(G_1, G_4) &= 1,92. \end{aligned}$$

Маємо наступний вектор відстаней  $v_1$ :

$$v_1 = (0 \quad 1,53 \quad 0,6 \quad 1,92).$$

Знайдемо нормований вектор відстаней  $v_1^{norm}$  :

$$v_1^{norm} \approx (0 \quad 0,797 \quad 0,313 \quad 1).$$

Тоді маємо наступну множину  $X_\varepsilon$ , яка містить шукані об'єкти БЗ подібні до об'єкта  $G_1$  в межах заданого відносного відхилення  $\varepsilon$  :

$$X_\varepsilon = \{G_1, G_3\}.$$

Слід зазначити, що інформацію для матриці відмінностей КС між собою  $D$  пропонується отримати одним із 3-х альтернативних способів:

1) за допомогою підходу, використаного В.Г. Тоценком для оцінювання компетентності експертів у групі [2]. Матриця  $D$  у цьому випадку отримується експертним шляхом. Цей підхід не є найзручнішим, адже, внаслідок великої кількості КС і психофізіологічних обмежень людини [1] доведеться розбивати множину КС на групи для порівняння; відповідно, знадобиться більший обсяг роботи експертів і кошти на її оплату;

2) шляхом представлення множини КС як семантичної мережі. Відмінність КС між собою у цьому випадку відповідає відстаням між її відповідними вузлами;

3) шляхом заповнення матриці  $D$  із використанням імовірнісної моделі виявлення зв'язків між поняттями (як поняття беруться КС) за допомогою інформаційно-аналітичних систем на основі обробки інформаційного потоку, сформованого в Інтернеті [19, 20].

### Шляхи використання результатів роботи методу

У результаті обчислення змістових відстаней для кожної пари об'єктів БЗ отримаємо симетричну матрицю змістових відстаней між усіма об'єктами БЗ. Далі за цією матрицею можна проводити кластерізацію об'єктів за змістом, що, зокрема, дасть можливість побачити та встановити відсутні зв'язки між подібними за змістом об'єктами, з якими, наприклад, працювали різні за спеціалізацією експерти.

При встановленні впливів у ієрархії цілей, коли експертові потрібно вибрати всі підцілі, що впливають на деяку ціль, метод може використовуватися для пошуку формулювань цілей з відомою експертові семантикою, але з невідомим формулюванням. У раніше побудованих ієрархіях можна знаходити помилково введені семантично однакові цілі, а також формулювати відсутні, але необхідні за змістом цілі ієрархії. Також можливе використання методу як додаткового інструменту для аналізу адекватності БЗ предметній області.

Накопичені в процесі роботи методу дані (а саме: КС і матриця відмінностей КС) можна використовувати при врахуванні компетентності експертів (оскільки врахування компетентності експертів у малих експертних групах є необхідною умовою забезпечення достовірності колективних рішень [21–23]), а також для формування експертних груп у відповідній предметній області.

## Висновки

Розглянуто особливості слабо структурованих предметних областей та моделей подачі знань. Обґрунтовано необхідність удосконалення моделі подачі знань «граф ієрархії цілей».

Запропоновано підхід і метод змістової ідентифікації об'єктів БЗ СППР. Суть методу полягає в тому, що кожному об'єкту БЗ ставиться у відповідність кортеж КС, важливість яких визначається експертним шляхом. Таким чином, кожен об'єкт БЗ ставиться у відповідність елементу скінченного нечіткого топологічного простору об'єктів БЗ. Змістова ідентифікація проходить за відстанню між об'єктами БЗ, що є метрикою у цьому просторі.

Використання підходу та методу змістової ідентифікації дозволяє підвищити адекватність моделей предметних областей, запобігти помилковому введенню у БЗ однакових за змістом цілей, зокрема при об'єднанні ієрархій цілей, сформованих різними експертними групами. Запропонований метод також може використовуватися для пошуку цілей ієрархії, точні формулювання яких невідомі, за ключовими словами.

1. *Миллер Дж.* Магическое число сем плюс или минус два. О некоторых пределах нашей способности перерабатывать информацию / Дж. Миллер. — Инженерная психология. — М.: Прогресс, 1964. — 696 с.

2. *Тоценко В.Г.* Методы и системы поддержки принятия решений. Алгоритмический аспект / В.Г. Тоценко. — К.: Наук. думка, 2002. — 382 с.

3. *Андрійчук О.В.* Розширення можливостей систем підтримки прийняття рішень шляхом врахування динаміки впливів / О.В. Андрійчук // Реєстрація, зберігання і оброб. даних: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 01-02 березня 2012 року / НАН України. Інститут проблем реєстрації інформації / відп. ред. В.В. Петров. — К.: ІПІ НАН України, 2012. — 158 с.

4. *Самарский А.А.* Математическое моделирование: Идеи. Методы. Примеры / А.А. Самарский, А.П. Михайлов. — М.: Физматлит, 2001. — 320 с.

5. *Андрійчук О.В.* Підвищення адекватності моделей предметних областей шляхом змістової ідентифікації / О.В. Андрійчук // Реєстрація, зберігання і оброб. даних: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 27-28 лютого 2013 року / НАН України. Інститут проблем реєстрації інформації; відп. ред. В.В. Петров. — К.: ІПІ НАН України, 2013. — 258 с.

6. *Саати Т.* Принятие решений. Метод анализа иерархий / Т. Саати. — М.: Радио и связь, 1993. — 278 с.

7. *Експериментальний аналіз технології експертного оцінювання* / В.В. Циганок, П.Т. Качанов, С.В. Каденко [та ін.] // Реєстрація, зберігання і оброб. даних. — 2012. — Т. 14, № 1. — С. 91–100.

8. *Свідомство* про реєстрацію авторського права на твір № 44521 Державної служби інтелектуальної власності України від 03/07/2012 / Циганок В.В., Андрійчук О.В., Качанов П.Т., Каденко С.В. // Комп'ютерна програма «Комплекс програмних засобів для експертного оцінювання шляхом парних порівнянь «Рівень».

9. *Андрійчук О.В.* Метод змістової ідентифікації об'єктів баз знань систем підтримки прийняття рішень / О.В. Андрійчук // Системи підтримки прийняття рішень. Теорія і практика: зб. до-

повідей науково-практичної конференції з міжнародною участю. — К.: ІППММС НАНУ, 2011. — 204 с.

10. Глушков В.М. Прогнозування і керування науковими дослідженнями [Електронний ресурс] / В.М. Глушков // Наукові доповіді на засіданнях Президії НАН України 1970–2001 рр.; ІПРІ НАН України. — Київ, 2008. — 1 DVD.

11. Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом / В.В. Циганок // Реєстрація, зберігання і оброб. даних. — 2000. — Т. 2, № 2. — С. 92–102.

12. Tsyganok V.V. Investigation of the Aggregation Effectiveness of Expert Estimates Obtained by the Pairwise Comparison Method / V.V. Tsyganok // Mathematical and Computer Modelling. — 2010, Aug. Vol. 52, № 3-4. — P. 538–544.

13. Кофман А. Введение в теорию нечётких множеств: пер. с франц. / А. Кофман. — М.: Радио и связь, 1982. — 432 с.

14. Колмогоров А.Н. Элементы теорий функций и функционального анализа / А.Н. Колмогоров, С.В. Фомин. — М.: Главная редакция физико-математической литературы изд-ва «Наука», 1976. — 544 с.

15. Шостак А.П. Два десятилетия нечеткой топологии: основные идеи, понятия и результаты / А.П. Шостак // Успехи математических наук. — 1989. — Т. 44. — Вып. 6(270). — С. 99–147.

16. Chang C. Fuzzy Topological Spaces / C Chang // J. Math. Anal. Appl. — 1968. — Vol. 24. — P. 182–190.

17. Жамбю М. Иерархический кластер-анализ и соответствия / М. Жамбю. — М.: Финансы и статистика, 1988. — 342 с.

18. Мандель И.Д. Кластерный анализ / И.Д. Мандель. — М.: Финансы и статистика, 1988. — 176 с.

19. Додонов А.Г. Выявление понятий и их взаимосвязей в рамках технологии контент-мониторинга / А.Г. Додонов, Д.В. Ланде // Реєстрація, зберігання і оброб. даних. — 2006. — Т. 8, № 4. — С. 45–52.

20. Додонов О.Г. Імовірнісна модель виявлення латентних зв'язків у мережах понять / О.Г. Додонов, Д.В. Ланде // Реєстрація, зберігання і оброб. даних. — 2011. — Т. 13, № 2. — С. 38–46.

21. Циганок В.В. Урахування компетентності експертів при визначенні групового ранжування / В.В. Циганок, О.В. Андрійчук // Реєстрація, зберігання і оброб. даних. — 2011. — Т. 13, № 1. — С. 94–105.

22. Цыганок В.В. Имитационное моделирование экспертных оценок для тестирования методов обработки информации в системах поддержки принятия решений [Текст] / В.В. Цыганок, С.В. Каденко, О.В. Андрейчук // Международный научно-технический журнал «Проблемы управления и информатики». — 2011. — № 6. — С. 83–93 / Tsyganok V.V. Simulation of Expert Judgements for Testing the Methods of Information Processing in Decision-Making Support Systems / V.V. Tsyganok, S.V. Kadenko, O.V. Andriichuk // Journal of Automation and Information Sciences. — 2011. — Vol. 43. — Issue 12. — P. 21–32.

23. Tsyganok V.V. Significance of Expert Competence Consideration in Group Decision-Making Using AHP / V.V. Tsyganok, S.V. Kadenko, O.V. Andriichuk // International Journal of Production Research. — 2012, Sept. — Vol. 50, N 17. — P. 4785–4792.

Надійшла до редакції 18.11.2013