Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут”

Андрійчук Олег Валентинович

УДК 519.816, 681.518.2

МЕТОД ВИЗНАЧЕННЯ ЗМІСТОВОЇ ПОДІБНОСТІ ОБ’ЄКТІВ БАЗ ЗНАНЬ ЕКСПЕРТНИХ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Спеціальність 01.05.04 – Системний аналіз і теорія оптимальних рішень

Автореферат

дисертації на здобуття наукового ступеня

кандидата технічних наук

Київ – 2015

Дисертацією є рукопис.

Робота виконана в Інституті проблем реєстрації інформації НАН України.

Науковий керівник: доктор технічних наук, старший науковий співробітник,
Циганок Віталій Володимирович,
Інститут проблем реєстрації інформації НАН України,
завідувач лабораторії систем підтримки прийняття рішень

Офіційні опоненти: доктор технічних наук, професор,
Данилов Валерій Якович,
Навчально-науковий комплекс «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН України та НАН України, професор кафедри математичних методів системного аналізу

Захист відбудеться ”__” _____ 2016 р. о 15 годині на засіданні спеціалізованої вченої ради Д26.002.03 при Національному технічному університеті України “Київський політехнічний інститут” за адресою: 03056, Київ-56, пр-т. Перемоги, 37,корп. 35, ауд. 1.
З дисертацією можна ознайомитися в бібліотеці Національного технічного університету України “Київський політехнічний інститут” за адресою: 03056, Київ-56, пр-т. Перемоги, 37.
Автореферат розісланий ”__” _____ 2016 р.

Учений секретар спеціалізованої Вченої ради,
д.ф.-м.н., професор Капустян В.О.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. На теперішній час, в умовах жорсткої конкуренції, стрімко зростає “ціна” помилки при прийнятті рішень. Задля забезпечення високого професійного рівня рішень застосовують системи підтримки прийняття рішень (СППР), результатом роботи яких є рекомендації особам, що приймають рішення (ОПР). Якість таких рекомендацій безпосередньо впливає на ефективність прийнятого рішення. Постійне підвищення складності моделей слабко структурованих предметних областей, вимагає адекватного та детального відображення сукупності факторів та їх взаємозв'язків у базі знань (БЗ) СППР. Значний рівень деталізації БЗ призводить до погіршення адекватності моделей слабко структурованих предметних областей, а саме: надлишковість, неоднозначність, наявність протиріч у БЗ.

Для побудови адекватних та детальних моделей слабко структурованих предметних областей, наряду з об'єктивною, необхідно використовувати інформацію отриману від експертів. Властивості слабко структурованих предметних областей унеможливлюють формування якісної навчальної вибірки компактних експертних формулювань “мовою предметної області”, що в свою чергу не дозволяє на основі існуючих методів та моделей визначати змістову подібність для однозначної інтерпретації за змістом кожного об’єкту БЗ експертної СППР.

Тому, розробка методу визначення змістової подібності об’єктів БЗ експертних СППР є актуальною задачею.

Сучасні розробки методів та моделей визначення змістової подібності належать ряду як вітчизняних (А.В. Анісімов, К.С. Ліман, А.А. Марченко, О.Г. Додонов, Д.В. Ланде) так і зарубіжних (Scott Deerwester, Susan T. Dumais, Richard Harshman, Thomas Hofman, David M. Blei, Andrew Y. Ng, Michael I. Jordan, Peter D. Turney, Kai-Wei Chang, Wen-tau Yih, Christopher Meek, Kevin Lund, Curt Burgess, Magnus Sahlgren, Islam Beltagy, Katrin Erk, Raymond Mooney) вчених.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконувалося у відповідності до планів науково-дослідницьких робіт, у рамках виконання держбюджетних та госпрозрахункових тем Інституту проблем реєстрації інформації НАН України:

- "Розробка наукових засад створення систем підтримки прийняття рішень ординального типу" (шифр “РАНГИ”; державний реєстраційний номер 0109U002106);
- "Теоретико-методологічні засади створення корпоративних автоматизованих інформаційно-аналітичних систем підвищеної живучості" (шифр “KORPYS-2007”; державний реєстраційний номер 0107U002354);
- "Розроблення системи кількісних та якісних показників для аналізу та оцінки ефективності виконання Плану заходів з розвитку космічної техніки на 2013 рік" (шифр “Програма-Н (ІПРІ)”; державний реєстраційний номер 0113U002766);

- "Розроблення методики оцінки ефективності виконання Плану заходів з розвитку космічної діяльності та виробництва космічної техніки на 2013 рік" (шифр “Програма-Н (методика)”; державний реєстраційний номер 0113U007093);
- «Дослідження та розробка адаптивних технологій організації обчислень і математичне моделювання вузлів глобальних архітектур для інформаційних систем державного управління» (шифр теми “НОРМА”; державний реєстраційний номер 0112U000667);
- ”Розробка методів та засобів комп’ютерного моделювання систем організаційного управління” (шифр „МОДУС – 2012”; державний реєстраційний номер 0112U002348);
- ”Теоретичні та технологічні засади експертної підтримки прийняття рішень при побудові стратегічних планів” (шифр “СТРАТЕГІЯ”; державний реєстраційний номер 0115U002075).

Мета і завдання дослідження. Метою дослідження є розробка методу визначення змістової подібності об’єктів БЗ експертних СППР, який підвищує адекватність моделей слабко структурованих предметних областей, завдяки чому досягається поліпшення якості рекомендацій для ОПР.

Досягнення поставленої мети потребує виконання наступних завдань:

1) Визначити особливості моделей слабко структурованих предметних областей, поданих у вигляді БЗ експертних СППР та визначити фактори, що впливають на якість рекомендацій для ОПР.
2) Провести аналіз існуючих методів визначення змістової подібності об’єктів предметних областей.
3) Розробити метод, який дозволить в БЗ СППР визначити змістову подібність об’єктів, назви яких подані у вигляді коротких експертних формулювань.
4) Побудувати та реалізувати імітаційні моделі експертних оцінок з метою їх застосування при дослідженні достовірності розробленого методу.
5) Реалізувати розроблений метод визначення змістової подібності в рамках відповідного програмного інструментарію побудови БЗ експертних СППР.

Об’єкт дослідження — процес визначення змістової подібності об'єктів БЗ слабко структурованих предметних областей в експертних СППР.

Предмет дослідження — методи та моделі визначення змістової подібності об'єктів БЗ слабко структурованих предметних областей в експертних СППР.

Методи дослідження. У ході дослідження були використані: системна методологія, теорія нечітких топологічних просторів, методи математичної статистики.

Наукова новизна одержаних результатів. Наукова новизна, дисертаційної роботи полягає у наступному:

1) Вперше розроблено метод визначення змістової подібності об’єктів БЗ слабко структурованих предметних областей в експертних СППР, який, на

відміну від існуючих, не потребує наявності навчальної вибірки формулювань об'єктів та базується на використанні експертної інформації.

2) Вперше розроблено імітаційну модель експертних оцінок для перевірки достовірності розробленого методу та на її основі створено моделюючий комплекс, що забезпечує можливість тестувати методи експертної підтримки прийняття рішень без проведення високовартісних експертиз із залученням реальних експертів.
3) Удосконалено програмний інструментарій побудови БЗ експертної СППР, шляхом розробки підсистеми визначення змістової подібності, що дало можливість підвищити якість рекомендацій для ОПР.

Практичне значення отриманих результатів.

- Розроблений у дисертаційній роботі метод реалізовано в вигляді макетів відповідних підсистем для СППР "Солон-3" та "Консенсус" [19].
- Розроблено новий програмний інструментарій експертного оцінювання шляхом парних порівнянь у вигляді комплексу програмних засобів "Рівень" [18].
- Реалізовано технологію побудови БЗ СППР експертними групами у вигляді системи розподіленого збору та обробки експертної інформації "Консенсус".

Система розподіленого збору та обробки експертної інформації для систем підтримки прийняття рішень – "Консенсус", комплекс програмних засобів для експертного оцінювання шляхом парних порівнянь "Рівень" та макет підсистеми визначення змістової подібності знайшли своє практичне застосування в інтеграції з СППР "Солон-3" при проведенні групового експертного оцінювання в рамках побудови системи кількісних та якісних показників для оцінки ефективності виконання Плану заходів з розвитку космічної діяльності та виробництва космічної техніки на 2013 р. в Інституті космічних досліджень НАН України та Державного космічного агентства (ДКА) України.

Окрім того, результати дисертаційного дослідження використовуються ТОВ "Комп'ютерні інформаційні технології" при розробці веб-орієнтованих систем.

Особистий внесок здобувача. Всі основні результати, що виносяться на захист, отримані автором самостійно. У роботах, виконаних у співавторстві, авторові належать: обґрунтування вибору законів розподілу випадкових величин для імітації індивідуальних експертних ранжирувань [1-2] та кардинальних експертних оцінок [2, 4]; конфігурація параметрів законів розподілу випадкових величин для імітації індивідуальних експертних оцінок [1-2, 4], реалізація відповідних експериментальних досліджень [1-4, 7]; технологія експертного оцінювання при застосуванні різних шкал [3]; інтерфейс програмного комплексу для проведення порівняльного дослідження різних технологій експертного оцінювання [3]; методика побудови БЗ СППР для вибору конфігурації сховищ даних [5]; ідея та проведення

експериментальної перевірки методу визначення змістової подібності об’єктів БЗ СППР [7].

Апробація результатів дисертації. Основні положення роботи доповідались та обговорювались на наступних наукових заходах:

- Низка семінарів “Системні дослідження та інформаційні технології” (Учбово-науковий комплекс “Інститут прикладного системного аналізу” при НТУУ “КПІ”, 2014-15 рр.);
- Щорічна підсумкова наукова конференція Інституту проблем реєстрації інформації НАН України (м. Київ, 2010-15 рр.);
- Сьома та восьма конференції з міжнародною участю “Системи підтримки прийняття рішень. Теорія і практика” СППР – 2011-2012 (Інститут проблем математичних машин і систем НАН України, м. Київ, 2011-12 рр.);
- Міжнародний форум “Проблеми розвитку інформаційного суспільства” (м. Львів, 2009 р.);
- XI та XIII Міжнародні симпозіуми з методу аналізу ієрархій, ISAHP 2011 (м. Сорренто, Італія, 2011 р.) та ISAHP 2014 (м. Вашингтон, США, 2014 р.);
- Восьма міжнародна конференція “Якісні та кількісні методи в бібліотеках” QQML-2012 – (м. Лімерик, Ірландія, 2012 р.).

Публікації. Результати дисертації викладені в 19 публікаціях, у тому числі, в 7 статтях у фахових наукових виданнях, тезах 10 доповідей наукових конференцій та 2 свідоцтвах про реєстрацію авторського права на твір.

Структура і обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел з 69 найменувань. Обсяг основної частини – 120 сторінок. Робота містить 10 таблиць і 75 рисунків.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність розробки методу визначення методу визначення змістової подібності; визначені: мета, об’єкт, предмет і методи досліджень; показано зв'язок з науковими програмами, планами, темами; наведено наукову новизну та практичне значення одержаних результатів; висвітлено особистий внесок здобувача.

У **першому розділі** представлена актуальність застосування експертних СППР для підтримки прийняття рішень в слабко структурованих предметних областях; показано роль і місце методу в процесі підтримки прийняття рішень; визначено особливості задач, що вирішуються з застосуванням методу визначення змістової подібності; аналіз стану питання та задачі досліджень (наявні підходи, методи та моделі визначення змістової подібності).

Процес експертної підтримки прийняття рішень у слабко структурованих предметних областях можна розділити на наступні три основні етапи [7]: формування експертних груп, побудова БЗ предметної області та формування рекомендацій для особи, що приймає рішення (ОПР).

При побудові БЗ в експертних СППР однією з найважливіших задач є підвищення адекватності моделей предметних областей, оскільки це безпосередньо впливає на достовірність рекомендацій, які надаються ОПР.

Місце методу визначення змістової подібності розглянемо на Use Case діаграмі експертної СППР (рис. 1). Метод реалізовано у вигляді підсистеми визначення змістової подібності.

Як видно з Use Case діаграми, ОПР формулює головну ціль. В результаті роботи СППР йому будуть надані відповідні рекомендації. Рекомендації розраховуються на основі БЗ, яку будує інженер по знаннях.

При побудові БЗ може використовуватися як об'єктивна, так і експертна інформація.

Відповідно до сформульованої головної цілі, інженер по знаннях формує групу експертів-спеціалістів, компетентних у питаннях, які будуть розглядатись. Експерти приймають участь у груповій декомпозиції цілей.

На кожному етапі декомпозиції експертам пропонується сформулювати множину цілей (серед існуючих або вводяться нові), які безпосередньо впливають на ціль, що розкривається. При формулюванні нових цілей відбувається характеристика їх змісту в рамках розробленого методу. Кожен з експертів дає свою множину індивідуальних формулювань. Потім, за допомогою методу визначення змістової подібності, формуються підмножини формулювань однакових за змістом, в кожній з яких вибирається найкраще формулювання. Далі, аналогічним чином, відбувається декомпозиція інших цілей. Процес декомпозиції завершується, коли в якості цілей виступають проекти, тобто конкретні заходи, на які може вплинути ОПР. Коли декомпозицію завершено, починають групове експертне оцінювання величин відносних ступенів впливу цілей. Таким чином формується БЗ слабко структурованої предметної області, на основі якої проводиться розрахунок рекомендацій для ОПР.

Також при побудові БЗ можуть використовуватися декілька експертних груп, які працюють паралельно, або для економії часу та коштів можуть використовуватися фрагменти раніше побудованих БЗ. Для цього інженер по знаннях здійснює об'єднання знань наданих різними експертними групами. При цьому мають бути встановлені відсутні впливи між цілями. Для цього використовується пошук об'єктів БЗ за заданим змістом в рамках розробленого методу. Також повинні бути виключені з БЗ протиріччя, надлишковість та неоднозначність. Для цього використовують рейтинг змістової відмінності об'єктів БЗ отриманий за допомогою методу визначення змістової подібності.

Особливості задач, що вирішуються з застосуванням методу визначення змістової подібності:

- Об'єкти, до яких застосовується метод, – це досить компактні текстові формулювання назв об'єктів БЗ “мовою предметної області” дані експертами.
- Само по собі формулювання назви об'єкту БЗ не може повністю відображати його зміст (повний збіг тексту формулювання не означає однозначного збігу за змістом).

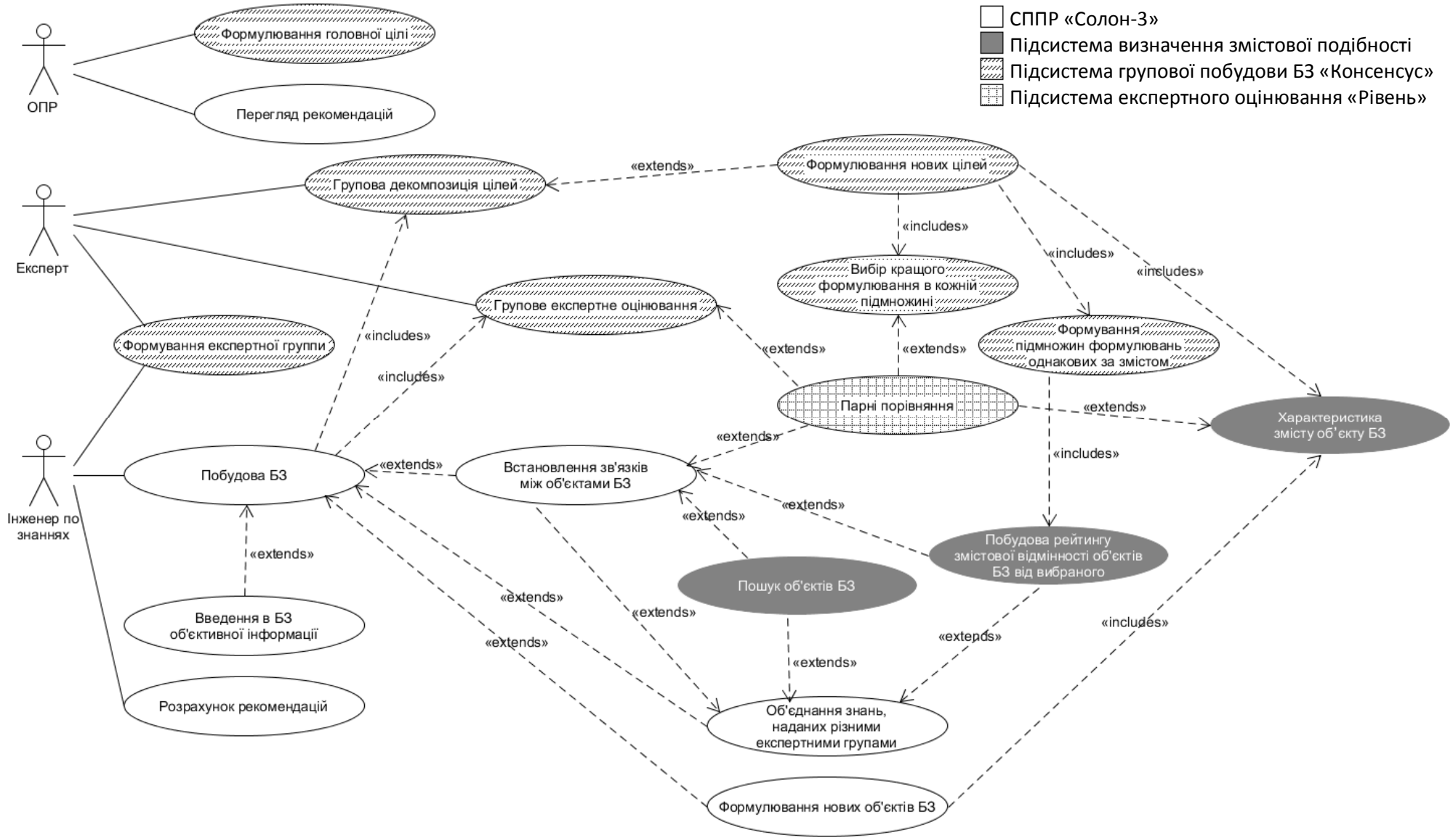

Рис. 1. Use Case діаграма експертної СППР.

- Властивості слабко структурованих предметних областей, такі як: унікальність об'єктів, відсутність еталонів, а також неточність, помилковість, неповнота, неоднозначність, суперечливість їх опису, унеможливлюють застосування онтологій, а також формування якісної, репрезентативної, несуперечливої, адекватної навчальної вибірки формулювань.
- Найбільш повною інформацією про сутність формулювання володіє тільки експерт, який його дав.
- У більшості випадків є можливість звернутися до експертів для отримання додаткової інформації про сутність формулювання.
- Зазвичай, БЗ СППР містять значну кількість об'єктів (мають високий рівень деталізації).

У відповідності до вище приведених особливості задач, що вирішуються з застосуванням методу визначення змістової подібності, до методу визначення змістової подібності об’єктів БЗ СППР висуваються наступні вимоги:

- Кожний об’єкт БЗ повинен однозначно інтерпретуватися за змістом в СППР, тобто система повинна мати змогу його знаходити, відрізняти від інших об’єктів БЗ за змістом, а також відповідати на запити, в яких це ім‘я згадується.
- Змістова подібність має визначатися без навчальної вибірки формулювань.
- Змістова сутність формулювання має визначатися експертом, який його дав.

Наявні підходи до визначення змістової подібності:

- статистичний: модель векторного простору з представленням кожного речення за допомогою моделі “мішок слів” (“bag of words”) та косинус між векторами, як міра змістової подібності (Charles T. Meadow, Mehran Sahami, Timothy D. Heliman);
- співставлення речень (sentence alignment): співставлення слів / фраз / частин з одного речення словам / фразам / частинам з іншого (Rada Mihalcea, Cuortney Corley, Carlo Strapparava);
- екстрагування характеристик (multiple features extraction): витягування якомога більше характеристик (лексичних, семантичних, синтаксичних) з речень та шляхом застосування машинного навчання порівнювати їх за змістом (Mladen Karan, Goran Glavaš, Frane Šarić, Jan Šnajder, Bojana Dalbelo Bašić). До цієї групи методів також належать методи, що використовують глибоке машинне навчання для визначення змістової подібності (Richard Socher, Christopher Manning);
- онтологічний: найкоротший шлях між концептами в онтологічній семантичній мережі (Philip Resnik, Claudia Leacock, Zhaohui Wu), перетин словникових описів концептів (Michael Lesk, Siddharth Patwardhan).

Наявні методи та моделі визначення змістової подібності: Latent Semantic Indexing (Scott Deerwester, Susan T. Dumais, Richard Harshman), Probabilistic Latent Semantic Indexing (Thomas Hofman), Latent Dirichlet Allocation (David M.

Blei, Andrew Y. Ng, Michael I. Jordan), Latent Relational Analysis (Peter D. Turney), Multi-Relational Latent Semantic Analysis (Kai-Wei Chang, Wen-tau Yih, Christopher Meek), Hyperspace Analogue to Language (Kevin Lund, Curt Burgess), Random Indexing (Magnus Sahlgren), Probabilistic Soft Logic (Islam Beltagy, Katrin Erk, Raymond Mooney), модифікована шляхоорієнтовна міра семантичної схожості (А.В. Анісімов, К.С. Ліман, А.А. Марченко), імовірнісна модель виявлення латентних зв'язків у мережах понять (О.Г. Додонов, Д.В. Ланде) та інші через властивості слабко структурованих предметних областей *не придатні* для визначення змістової подібності об’єктів БЗ експертних СППР.

Отже, *актуальною задачею* є розробка методу визначення змістової подібності об’єктів БЗ СППР, що задовольняє всім відповідним вимогам.

Другий розділ присвячено методу визначення змістової подібності об’єктів БЗ експертних СППР [6].

Ідея визначення змістової подібності, певною мірою, перекликається з підходом до визначення компетентності експертів відносно питання, що обговорюється (В.Г. Тоценко).

Сутність підходу визначення змістової подібності:

1. Для відображення змісту, кожному об’єкту БЗ ставиться у відповідність кортеж ключових слів (КС) з відповідними коефіцієнтами важливості. Кортеж КС повинен мати мінімальну потужність, необхідну для збереження унікальності змісту формулювання назви об’єкту БЗ.
2. Коефіцієнти важливості КС для кожного об’єкту БЗ визначаються експертним шляхом та представляють собою відповідні дійсні числа.
3. Визначення змістової подібності об’єктів БЗ відбувається за близькістю коефіцієнтів важливості їх КС.

Задача визначення змістової подібності об’єктів БЗ СППР має наступну постановку:

- *Дано:*

множина всіх об’єктів БЗ $G = \| G_i \|_{i=\overline{1,m}}$, де m – кількість об’єктів в БЗ;

- *Потрібно визначити:*

відносний показник змістової подібності s_t об’єктів БЗ до заданого об’єкту $G_t \in G$.

Основна ідея методу полягає в тому, що кожному об’єкту БЗ ставиться у відповідність кортеж ключових слів (КС), за якими визначається змістова подібність. Важливість КС визначається експертним шляхом. Експериментальне дослідження [7] підтвердило достовірність та практичну цінність методу визначення змістової подібності.

Схема методу визначення змістової подібності представлена на рис.2.

Для характеристики змісту об’єкту БЗ G_i, задається кортеж КС з відповідним коефіцієнтами важливості $G_i = \| g_{ij} \|_{i=\overline{1,m} j=\overline{1,n}}$. Кортеж КС має бути мінімальної потужності і при цьому містити всю необхідну інформацію для визначення відповідного об’єкту БЗ.

Цей кортеж КС із відповідними вагами може бути отриманий наступними шляхами:

1. заданий експертним шляхом (вагові коефіцієнти КС можуть визначатись методами власного вектору, «трикутник», «квадрат», комбінаторним, чи іншими);
2. сформульований на основі нормативних документів;
3. оснований на раніше побудованих ієрархіях.

Загальна для предметної області матриця відмінностей КС $D = \| d_{ij} \|_{i=\overline{1,n};\, j=\overline{1,n}}$, основана на порівняннях всіх КС між собою за змістовою відмінністю. Матриця *D* може бути отримана наступними шляхами:

1. експертним (попарні порівняння КС за змістовою відмінністю);
2. представлення множини КС як семантичної мережі, де відмінність КС між собою, відповідає відстаням між відповідними її вузлами;
3. використання імовірнісної моделі виявлення зв'язків між поняттями (в якості понять беруться КС) за допомогою інформаційно-аналітичних систем на основі обробки інформаційного потоку, сформованого в Інтернет.

Рис. 2. Метод визначення змістової подібності об'єктів БЗ експертних СППР.

Третій розділ присвячено перевірці достовірності методу визначення змістової подібності об'єктів БЗ СППР [7].

Через наступні властивості слабко структурованих предметних областей: унікальність об'єктів, відсутність еталонів, а також неточність, помилковість, неповнота, неоднозначність, суперечливість їх опису відсутність еталонів, перевіряти достовірність розробленого методу потрібно з залученням експертів:

1. експерт оцінює змістову відмінність об'єктів відповідно до своїх переваг шляхом парних порівнянь їх за змістом,
2. експерт вказує КС об'єктів та їх відносну важливість,

3. змістова відмінність об’єктів розраховується розробленим методом.

Для вищеописаного процесу потрібно залучати групу експертів, робота яких коштує досить дорого та потребує значного часу. Оскільки вихідними даними методу є експертні оцінки важливостей КС об’єктів, то доцільним є застосування імітаційного моделювання експертних оцінок для його перевірки.

Експериментальна перевірка достовірності методу за допомогою імітаційного моделювання експертних оцінок:

- в якості об'єктів БЗ візьмемо публікації (наукові статті, доповіді) одного автора;
- запропонуємо автору оцінити змістову відмінність своїх публікацій шляхом парних порівнянь;
- проведемо імітаційне моделювання експертних оцінок важливості КС об'єктів БЗ, де в якості еталонну беруться ранжирування КС, взятих з відповідних публікацій автора.

Оскільки кортежи КС в такому випадку будуть ранжуваннями, то потрібно перейти від їх порядку до їх ваг. Для цього використовувалися наступні шляхи:

- закон Ньюкомба-Бенфорда (закон першої цифри):
  перша цифра d (d {1, …, b − 1}) з основою b трапляється з ймовірністю:

$$P(d) = \log_b(d+1) - \log_b(d) = \log_b(1+\frac{1}{d})$$

  ця ймовірність й береться в якості ваги КС;
- вага КС w_i пропорційна його номеру n_i у ранжируванні:

$$w_i = \frac{N - n_i + 1}{\sum_{j=1}^{N} n_j}$$

Імітаційне моделювання кардинальних експертних оцінок проводиться наступним чином:

1. генеруємо девіацію Δ, як випадкову величину, розподілену за експоненційним або за півнормальним законом: шкала відношень $\Rightarrow$ $\Delta = (a_{\Delta+}/a_e) = (a_e/a_{\Delta-})$, де a_e – значення елемента ідеально узгодженої МПП, що піддається збуренню; $a_{\Delta+}$ та $a_{\Delta-}$ – значення елементів при відхиленні у бік зростання та зменшення.
2. відхилення вибирається випадковим чином: у бік зростання ($a_{\Delta+}=a_e \cdot \Delta$) або зменшення ($a_{\Delta-}=a_e/\Delta$).
3. округлення до найближчої поділки шкали (можливі значення: {1/9, 1/8, …, 1/2, 1, 2, …, 8, 9}).

Функції щільності ймовірностей обраних законів розподілу для імітації експертних оцінок:

- експонеційний закон розподілу:

$$f(x) = \lambda \cdot e^{-\lambda x},\ x \geq 0 \qquad (MExp = 1/\lambda)$$

  де $\lambda > 0$ – параметр розподілу;
- нормальний (Гаусівський) закон розподілу:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty; \infty)$$

де параметр μ – середнє/очікуване значення (пік) і σ^2 – середнє квадратичне відхилення (міра ширини розподілу)

- півнормальний закон розподілу (при $\mu = 0$):

$$f(x) = \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, x \in [0;\infty) \qquad (MExp = \sqrt{2/\pi} \cdot \sigma)$$

Підбір параметрів обох розподілів для імітації експертних оцінок:

- прирівнюємо відповідні математичні сподівання:

$$(MExp = 1/\lambda = \sqrt{2/\pi} \cdot \sigma)$$

- з того, що інтегральна функція експоненційного розподілу

$$1 - e^{-\lambda x} = 0.95 \Rightarrow 1/\lambda = -x/\ln(0.05)$$

(в нашому випадку *x=1*) і це відношення ≈ *0.3338 (33.38%)*.

- таким чином, можемо обчислити параметри для обох розподілів виходячи з наступного відношення:

$$1/\lambda = \sqrt{2/\pi} \cdot \sigma = 0.3338$$

Обрані закони розподілу відхилень індивідуальних експертних оцінок від початкових “еталонних” значень показані на рис. 3.

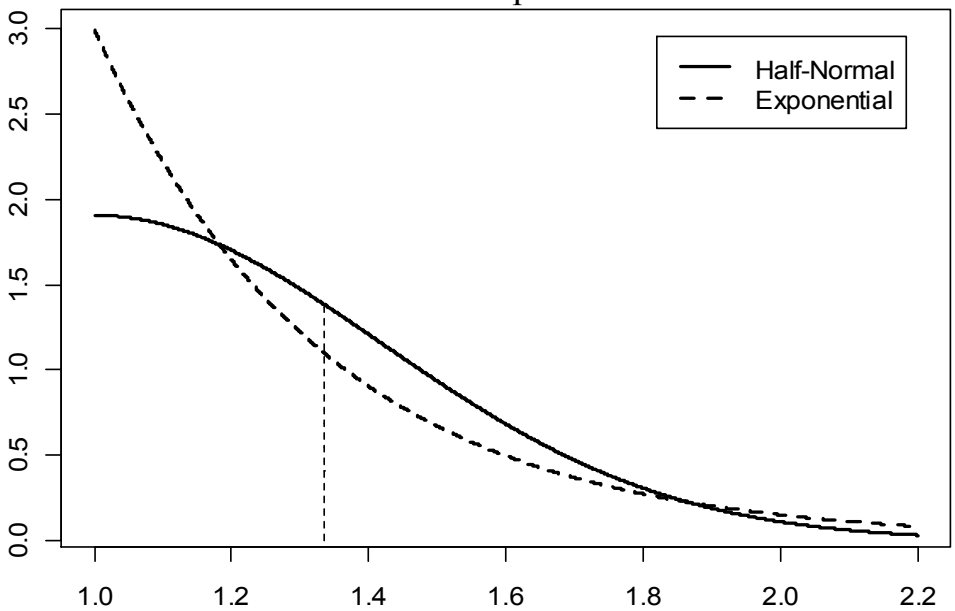

Рис. 3. Закони розподілу відхилень індивідуальних експертних оцінок від початкових “еталонних” значень.

Результати експериментального дослідження методу змістової подібності наведені у таблиці 1.

Отримані результати імітації 66-ти повторів для кожного з наявних еталонних значень зведені в таблицю. Таким чином, із 660 повторів експерименту для 10 різних еталонів, розрахований показник подібності об’єктів відрізняється від еталонного значення менш ніж на 10% в 598 виходах експерименту, що складає 90.6% усіх випробувань.

Таблиця 1. Результати експериментального дослідження методу змістової подібності

	P_1P_2	P_1P_3	P_1P_4	P_1P_5	P_2P_3	P_2P_4	P_2P_5	P_3P_4	P_3P_5	P_4P_5
p'	57	62	53	65	62	64	49	64	59	63
$p = p' / 66$	0.864	0.939	0.803	0.985	0.939	0.970	0.742	0.970	0.894	0.955

Статистична достовірність результатів дослідження:

- Пробна серія експерименту із 30 повторів для кожного з 10-ти еталонів наведена у таблиці 2:

Таблиця 2. Пробна серія експерименту із 30 повторів для кожного з 10-ти еталонів

	P_1P_2	P_1P_3	P_1P_4	P_1P_5	P_2P_3	P_2P_4	P_2P_5	P_3P_4	P_3P_5	P_4P_5
p'	25	28	18	29	28	26	21	29	27	26
$p = p' / 30$	0.833	0.933	0.6	0.967	0.933	0.867	0.7	0.967	0.9	0.867

- Оцінка статистичної достовірності проводилась на основі центральної граничної теореми:

$$n \geq \frac{p \cdot (1-p)}{\beta^2} \left(F^{-1}\left(P_\beta\right)\right)^2$$

$$F^{-1}(0.9) \approx 1.65$$

$$\left(F^{-1}(0.9)\right)^2 \approx 2.72$$

$$n \geq \frac{0.6 \cdot (1-0.6)}{0.1^2} 2.72 = 65.28$$

У **четвертому розділі** представлено практичну реалізацію методу визначення змістової подібності об'єктів БЗ СППР у вигляді макетів відповідних підсистем для СППР “Солон-3” та системи “Консенсус” [19].

Система розподіленого збору та обробки експертної інформації для систем підтримки прийняття рішень – „Консенсус” [19] призначена для проведення оцінювання розподіленими групами експертів з подальшим використанням зібраної та обробленої інформації в СППР. Система реалізує технологію побудови бази знань для слабко структурованих предметних областей. Система “Консенсус” надає експертам можливість формулювання цілей та встановлення між ними зв'язків (рис. 4).

Для кожного формулювання експерт задає КС об'єкту БЗ та їх ваги у підсистемі змістової ідентифікації (рис. 5).

Важливість КС експерт вказує шляхом відповідних парних порівнянь з використанням розробленої технологія експертного оцінювання, яка реалізована у вигляді комплексу програмних засобів „Рівень” [18].

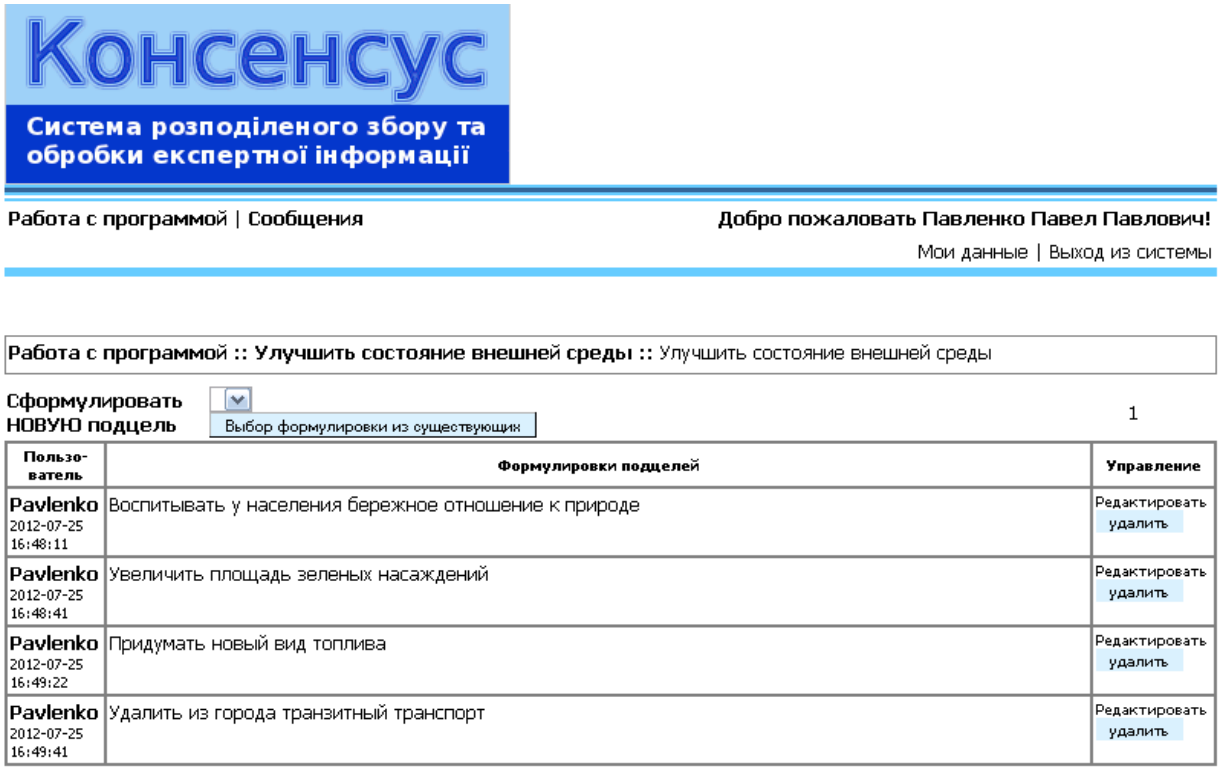

Рис. 4. Формулювання експертами цілей та встановлення між ними зв'язків.

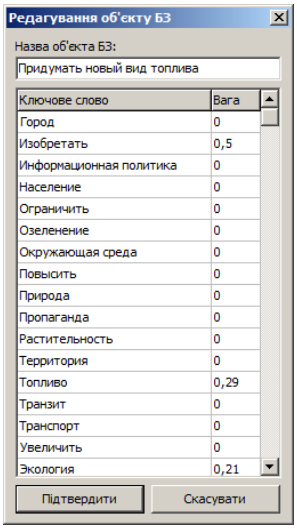

Рис. 5. Задавання КС об'єкту БЗ та їх ваг.

Комплекс програмних засобів для експертного оцінювання шляхом парних порівнянь „Рівень” [18] призначений для проведення експертизи в системах підтримки прийняття рішень і дозволяє отримувати знання від експертів шляхом надання їм можливості попарно порівнювати об’єкти між собою. Він дає змогу експертові вказувати наявність переваги між об’єктами з можливістю подальшого поступового уточнення ступеня цієї переваги до досягнення рівня, відповідного дійсним знанням експерта про об’єкт (рис. 6).

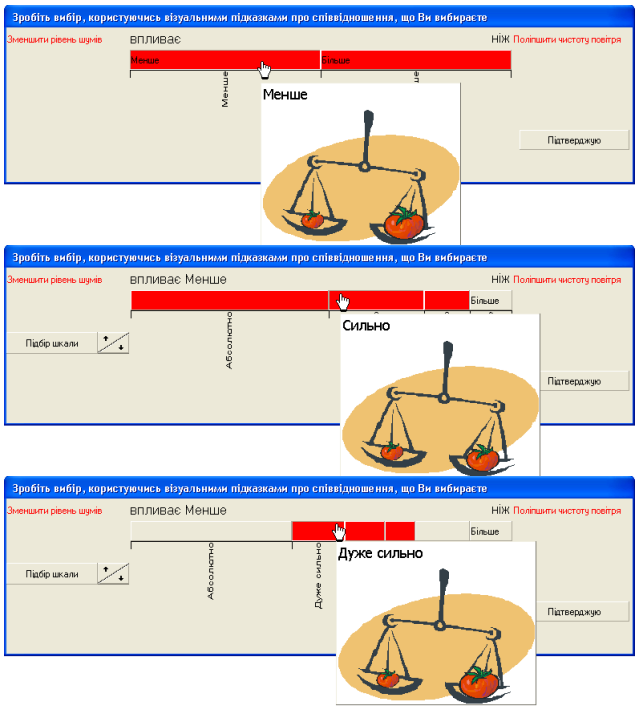

Рис. 6. Поступове уточнення ступеня переваги в комплексі „Рівень” (визначення наявності переваги впливу, визначення ступеня переваги впливу у вербальній шкалі з 3-ма поділками, визначення ступеня переваги впливу у більш докладній вербальній шкалі).

Підсистема визначення змістової подібності надає інженеру по знаннях можливість перегляду рейтингу змістової відмінності об'єктів БЗ від вибраного (рис. 7). За допомогою цього експерт може виключити з БЗ об'єкти, які співпадають за змістом, але різні за формулюванням, для запобігання надлишковості. Також є можливість довстановлювати впливи між об'єктами БЗ, яких не вистачає.

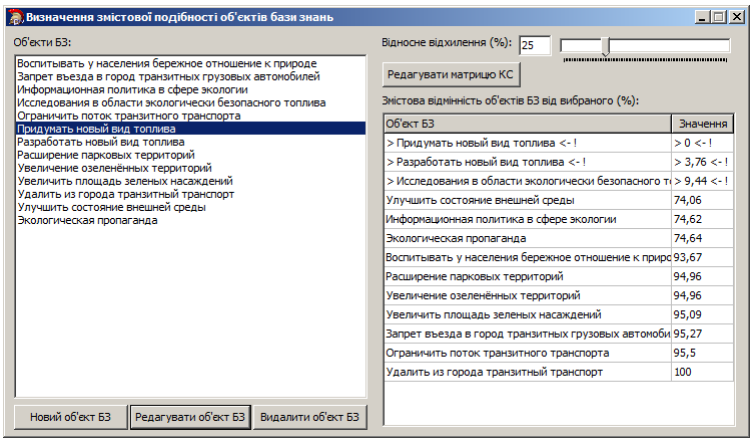

Рис. 7. Рейтинг змістової відмінності об'єктів БЗ від вибраного.

В результаті роботи системи “Консенсус” формуються відповідні таблиці реляційної бази даних. У подальшому СППР інтерпретує ці проміжні дані в БЗ відповідної предметної області.

ВИСНОВКИ

У рамках виконання завдань дисертаційного дослідження отримано наступні нові наукові результати:

1. Визначено особливості моделей слабко структурованих предметних областей, поданих у вигляді БЗ експертних СППР та визначено найважливіші фактори, що впливають на якість рекомендацій для ОПР, а саме: надлишковість, неоднозначність та наявність протиріч у БЗ СППР.
2. Проведено аналіз наявних підходів та методів визначення змістової подібності, в результаті якого вибрані напрями досліджень для вирішення задачі визначення змістової подібності об’єктів БЗ експертних СППР.
3. Розроблено метод визначення змістової подібності об’єктів БЗ слабко структурованих предметних областей в експертних СППР, який, на відміну від існуючих, не потребує наявності навчальної вибірки (корпусу текстів). Метод дозволяє підвищити адекватність моделей предметних областей, запобігти помилковому введенню у БЗ однакових за змістом об'єктів, зокрема, при об’єднанні БЗ, сформованих різними експертними групами, а також може використовуватись для пошуку за КС у БЗ об’єктів, точні формулювання яких невідомі.
4. Запропоновано імітаційну модель ЕО та створено відповідний моделюючий комплекс для перевірки достовірності методу визначення змістової подібності. Отримані результати тестування (більш ніж 90% усіх випробувань відрізняються менш ніж на 10% від заданих експертним шляхом еталонних значень) дають змогу зробити висновок про достовірність та практичну цінність методу.
5. Метод визначення змістової подібності об’єктів БЗ експертних СППР реалізовано у вигляді підсистеми визначення змістової подібності в рамках системи групової побудови БЗ “Консенсус”, яка дозволяє проводити територіально розподілену експертизу. Також розроблено програмний інструментарій експертного оцінювання (система “Рівень”), який, зокрема в рамках методу визначення змістової подібності, дозволяє уникати спотворень при отриманні експертної інформації.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Циганок В.В. Урахування компетентності експертів при визначенні групового ранжирування / В.В. Циганок, О.В. Андрійчук // Реєстрація, зберігання і обробка даних. – 2011. – т.13. – №1. – С. 94-105.
2. Цыганок В.В. Имитационное моделирование экспертных оценок для тестирования методов обработки информации в системах поддержки принятия решений / В.В. Цыганок, С.В. Каденко, О.В. Андрійчук // Международный научно-технический журнал «Проблемы управления и информатики». – 2011. – №6. – С. 84-94. (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. – Volume 43 / Issue 12. p.21-32.).

3. Циганок В.В. Експериментальний аналіз технології експертного оцінювання / В.В. Цыганок, П.Т. Качанов, С.В. Каденко, О.В. Андрійчук, Г.А. Гоменюк // Реєстрація, зберігання і обробка даних. – 2012. – т.14. – №1. – С. 91-100.
4. 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. – v.50, issue 17. – P.4785-4792.
5. Andriichuk O.V. An Approach to Decision Support System Usage for Data Storage Configuration Variant Selection / O.V. Andriichuk, S.V. Kadenko, V.V. Tsyganok // Qualitative and Quantitative Methods in Libraries (QQML) issue 3 – 2012, pp. 327 –333.
6. Андрійчук О.В. Метод змістової ідентифікації об’єктів баз знань систем підтримки прийняття рішень / О.В. Андрійчук // Реєстрація, зберігання і обробка даних. – 2014, - Т.16, №1 – С.65-78.
7. Циганок В.В. Експериментальне дослідження методу визначення змістової подібності об’єктів баз знань систем підтримки прийняття рішень / В.В. Циганок, О.В. Андрійчук // Реєстрація, зберігання і обробка даних. – 2014, - Т.16, №4 – С.65-76.
8. Березін Б.О. Підтримка прийняття рішень при побудові систем довготермінового зберігання інформації / Б.О. Березін, П.Т. Качанов, В.В. Цыганок, О.В. Андрійчук // Матеріали міжнародного форуму “Проблеми розвитку інформаційного суспільства” (Львів, 7-9 жовтня 2009 р.). – К. УкрІРТІ, 2009. – С. 145-153.
9. Tsyganok V.V. Significance of Expert Competence Consideration while Group Decision-Making using AHP / V.V. Tsyganok, S.V. Kadenko, O.V. Andriichuk // Proceedings of the Eleventh International Symposium for the Analytic Hierarchy/Network Process. (Sorrento, Naples, Italy. June 15-18, 2011) – Access mode: http://204.202.238.22/isahp2011/dati/pdf/154_078_Tsyganok.pdf
10. Андрійчук О.В. Метод змістової ідентифікації об’єктів баз знань систем підтримки прийняття рішень / О.В. Андрійчук // Системи підтримки прийняття рішень. Теорія і практика: Збірник доповідей науково-практичної конференції з міжнародною участю. – Київ: ІПММС НАНУ, 2011. – С. 161-163.
11. Андрійчук О.В. Розширення можливостей систем підтримки прийняття рішень шляхом врахування динаміки впливів / О.В. Андрійчук // Реєстрація, зберігання і обробка даних: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 01-02 березня 2012 року / НАН України. Інститут проблем реєстрації інформації. – К.: ІПРІ НАН України, 2012. – С 154-157.
12. Андрійчук О.В. Застосування системи підтримки прийняття рішень для вибору конфігурації системи електронного документообігу / О.В. Андрійчук, С.В. Каденко, В.В. Цыганок // Системи підтримки

прийняття рішень. Теорія і практика: Збірник доповідей науково-практичної конференції з міжнародною участю. – Київ: ІПММС НАНУ, 2012. – С.22-26.
13. Andriichuk O.V. An Approach to Decision Support System Usage for Data Storage Configuration Variant Selection / O.V. Andriichuk, S.V. Kadenko // Qualitative and Quantitative Methods in Libraries (QQML-2012) – Limeric (Ireland), May 22-25, 2012, p. 29.
14. Андрійчук О.В. Підвищення адекватності моделей предметних областей шляхом змістової ідентифікації / О.В. Андрійчук // Реєстрація, зберігання і обробка даних: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 27-28 лютого 2013 року / НАН України. Інститут проблем реєстрації інформації. – К.: ІПРІ НАН України, 2013. – С 246-250.
15. Андрійчук О.В. Особливості застосування методу змістової ідентифікації при експертній підтримці прийняття рішень / О.В. Андрійчук // Реєстрація, зберігання і обробка даних: зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 14-15 травня 2014 року / НАН України. Інститут проблем реєстрації інформації. – К.: ІПРІ НАН України, 2014. – 146 с.
16. Tsyganok V.V. Aggregating pair-wise comparisons given in scales of different detail degree / V.V. Tsyganok, O.V. Andriichuk // Proceedings of the Thirteenth International Symposium for the Analytic Hierarchy/Network Process. – (Washington, USA, June 29 – July 2, 2014) – Access mode: http://www.isahp.org/uploads/p729679.pdf
17. Андрійчук О.В. Визначення змістової подібності при експертній підтримці прийняття рішень / О.В. Андрійчук // Реєстрація, зберігання і обробка даних: зб. Наук. Праць за матеріалами Щорічної підсумкової наукової конференції 26-27 травня 2015 року / НАН України. Інститут проблем реєстрації інформації. – К.: ІПРІ НАН України, 2015. – 160 с.
18. Свідоцтво про реєстрацію авторського права на твір № 44521 Державної служби інтелектуальної власності України. Комп'ютерна програма “Комплекс програмних засобів для експертного оцінювання шляхом парних порівнянь «Рівень»” / В.В. Цыганок, О.В. Андрійчук, П.Т. Качанов, С.В. Каденко // від 03.07.2012.
19. Свідоцтво про реєстрацію авторського права на твір № 45894 Державної служби інтелектуальної власності України. Комп'ютерна програма “Система розподіленого збору та обробки експертної інформації для систем підтримки прийняття рішень - «Консенсус»” // В.В. Цыганок, П.Т. Качанов, О.В. Андрійчук, С.В. Каденко // від 03.10.2012.

АНОТАЦІЯ

Андрійчук О.В. Метод визначення змістової подібності об'єктів баз знань експертних систем підтримки прийняття рішень. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.04 – системний аналіз і теорія оптимальних рішень. – Національний технічний університет України "Київський політехнічний інститут", Київ, 2015 р.

Дисертаційна робота присвячена розробці методу визначення змістової подібності об'єктів баз знань експертних систем підтримки прийняття рішень, який дозволяє поліпшити якість рекомендацій для осіб, що приймають рішення, завдяки підвищенню адекватності моделей слабко структурованих предметних областей. У роботі запропоновано метод, який, на відміну від існуючих, не потребує наявності навчальної вибірки формулювань об'єктів та базується на використанні експертної інформації.

Для перевірки достовірності розробленого методу запропоновано імітаційну модель експертних оцінок та на її основі створено моделюючий комплекс, що забезпечує можливість тестувати методи експертної підтримки прийняття рішень без проведення високовартісних експертиз із залученням реальних експертів.

Для експертного оцінювання в рамках методу визначення змістової подібності розроблено та протестовано новий програмний інструментарій, який, на відміну від існуючих, дозволяє експерту при кожному порівнянні використовувати різні шкали та поступово уточнювати свої оцінки, що дозволяє уникнути спотворень при отриманні експертної інформації.

Ключові слова: система підтримки прийняття рішень, слабко структурована предметна область, база знань, змістова подібність, інженерія знань, експертне оцінювання.

АНОТАЦИЯ

Андрейчук О.В. Метод определения смыслового подобия объектов баз знаний экспертных систем поддержки принятия решений. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.04 – системный анализ и теория оптимальных решений. – Национальный технический университет Украины "Киевский политехнический институт", Киев, 2015 г.

Диссертационная работа посвящена разработке метода определения смыслового подобия объектов баз знаний экспертных систем поддержки принятия решений, который позволяет улучшить качество рекомендаций для лиц, принимающих решения, благодаря повышению адекватности моделей слабо структурированных предметных областей. В работе предложен метод определения смыслового подобия объектов баз знаний слабо структурированных предметных областей в экспертных системах поддержки принятия решений, который, в отличие от существующих, не требует наличия

обучающей выборки формулировок объектов и базируется на использовании экспертной информации.

Основная идея метода заключается в том, что каждому объекту базы знаний сопоставляется кортеж ключевых слов, по относительным коэффициентам важности которых и определяется смысловое подобие. Важность ключевых слов определяется экспертным путем. Метод даёт возможность устранить избыточность, неоднозначность и наличие противоречий в базе знаний, что и приводит повышению адекватности соответствующих моделей предметных областей.

Для проверки достоверности разработанного метода предложена имитационная модель экспертных оценок и на ее основе создан моделирующий комплекс, обеспечивающий возможность тестировать методы экспертной поддержки принятия решений без проведения дорогостоящих экспертиз с привлечением настоящих профессионалов.

Для экспертного оценивания в рамках метода определения смыслового подобия разработано и протестировано новый программный инструментарий, который, в отличие от существующих, позволяет эксперту при каждом сравнении использовать различные шкалы и постепенно уточнять свои оценки, что позволяет избежать искажений при получении экспертной информации. Даная технология реализована в виде комплекса программных средств для экспертного оценивания путем парных сравнений «Уровень».

Разработанный метод реализован в системе распределенного сбора и обработки экспертной информации "Консенсус", которая позволяет проводить территориально распределенную экспертизу при работе экспертов в сети Интернет.

Метод определения смыслового подобия объектов баз знаний экспертных систем поддержки принятия решений нашел практическое применение при построении базы знаний системы количественных и качественных показателей для анализа и оценки эффективности выполнения Плана мероприятий по развитию космической деятельности и производства космической техники, а также при построении базы знаний для выбора варианта конфигурации системы электронного документооборота.

Ключевые слова: система поддержки принятия решений, слабо структурированная предметная область, база знаний, смысловое подобие, инженерия знаний, экспертное оценивание.

ANNOTATION

Andriichuk O.V. A method for defining semantic similarity of objects in knowledge bases of expert decision support systems. – Manuscript.

The dissertation for acquiring of the PhD scientific degree of engineering sciences, specialty 01.05.04 – System analysis and optimal decisions theory. National Technical University of Ukraine “Kyiv Polytechnic Institute”, Kyiv, 2015.

The dissertation is dedicated to development of a method for defining semantic similarity of objects in knowledge bases of expert decision support systems. The method provides an opportunity to improve the quality of recommendations for

decision-makers through increasing the adequacy of models of weakly structured domains. The author suggests a method, which, in contrast, to other existing methods, does not require a training sample of object formulations, and is based on expert data utilization.

In order to verify the relevance of the developed method, a simulation model of expert estimation has been developed, and a modeling complex has been created on its basis. The modeling complex provides an opportunity to test expert decision support methods without conducting costly expert examinations involving real experts.

To facilitate expert estimation within semantic similarity definition method, a special software toolkit has been developed and tested. In contrast to other existing software toolkits, it allows the expert to use different scales for every pair comparison, and gradually increase his/her estimates, thus, preventing information distortion during expert data input.

Keywords: decision support system, weakly structured domain, knowledge base, semantic similarity, knowledge engineering, expert estimation.