

УДК 621.391

А.О. Феклістов, В.А. Лупандін, Г.В. Мегельбей

Харківський університет Повітряних Сил ім. І. Кожедуба, Харків

МЕТОД ФОРМУВАННЯ ІНТЕГРОВАНІХ ПОНЯТІЙНИХ КОМПОНЕНТІВ В ІНФОРМАЦІЙНИХ СИСТЕМАХ ОРГАНІВ ВІЙСЬКОВОГО УПРАВЛІННЯ

Розглядається метод формування інтегрованих понятійних компонентів в базах даних інформаційних систем органів військового управління на основі матриць семантичної сумісності.

Ключові слова: інформаційна система органу військового управління, база даних, матриця семантичної сумісності, класифікатор понять (термінів).

Вступ

Постановка проблеми у загальному вигляді.

В умовах реформування Збройних Сил (ЗС) України однією з важливих складових забезпечення воєнної безпеки держави як складової національної безпеки є розвиток воєнної науки, формування науково-технічної і технологічної бази для створення високоефективних засобів збройної боротьби [1 – 3].

Серед науково-технічних задач інформаційного та інформаційно-аналітичного забезпечення органів військового управління (ОВУ) існує задача інтеграції класифікаторів, які відображають існуючі в даній предметній галузі (ПрГ) понятійні складові в базах даних (БД) інформаційних систем. Наявність різноманітних класифікацій понять (термінів) обумовлена об'єктивною складністю розробки єдиної термінологічної бази та існуванням прикладних задач ОВУ, які використовують тільки частину потрібних їм понять з урахуванням галузевих особливостей. Необхідність в розробці інтегрованих понятійних компонентів (ІПК) БД обумовлена наступними вимогами користувачів інформаційних систем ОВУ [4 – 6]:

– можливість вирішення на основі ІПК БД прикладних задач, в яких використовуються поняття з різних класифікаторів ПрГ таким чином, що вирішення задачі на основі одного класифікатора може бути інтерпретовано в термінах іншого класифікатора;

– можливість підключення к ІПК БД нових класифікаторів і пов'язаних з ним прикладних задач, алгоритми вирішення яких автоматично налагоджуються на використання існуючих в системі класифікаторів;

– можливість розробки на основі ІПК БД інформаційно-пошукових підсистем, які використовують в якості тезауруса поняття із інтегрованого компонента.

Аналіз відомих методів інтеграції інформаційних складових в БД свідчить про те, що на даний

час існує необхідність в методах інтеграції понятійних складових класифікаторів понять (термінів).

Таким чином, актуальним є розробка метода формування ІПК БД інформаційних систем ОВУ на основі матриць семантичної сумісності.

Мета статті – розробка метода формування ІПК БД інформаційних систем ОВУ на основі матриць семантичної сумісності.

Викладення матеріалів досліджень

Класифікатори, які використовуються в ПрГ, можуть бути умовно поділені на базові та додаткові.

Базові класифікатори – це офіційно затверджені класифікації понять (термінів), які систематично використовуються ОВУ в повсякденній діяльності, наприклад, в якості довідкової та нормативно-методичної документації.

Додаткові класифікатори – це різноманітні класифікації понять (термінів) та комбінації понять з базових класифікаторів, які не затверджені офіційно, але практично використовуються під час вирішення конкретних прикладних задач.

Інтеграція класифікаторів полягає в формуванні ІПК БД ОВУ (рис. 1).

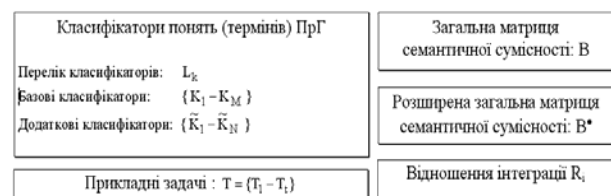


Рис. 1. Загальний вид ІПК БД ОВУ

Структурно ІПК БД складається з наступних компонентів: перелік класифікаторів (L_k); базові класифікатори ($\{K_1 - K_M\}$); додаткові класифікатори ($\{\tilde{K}_1 - \tilde{K}_N\}$); загальна матриця семантичної сумісності (B); розширена загальна матриця семанти-

чної сумісності (B^\bullet); відношення інтеграції (R_i); прикладні задачі ($\{T_1 - T_t\}$).

В загальному вигляді технологічна схема формування ІПК БД складається з трьох послідовних етапів ($E_1 - E_3$) наступної структури:

$$E_i = \langle I_i^{BX}, M_i, I_i^{VIX} \rangle,$$

де $i \in [1, 3]$ – номер етапу; I_i^{BX} – вхідна інформація; M_i – метод обробки вхідної інформації; I_i^{VIX} – вихідна інформація.

Метод обробки інформації визначається наступними складовими:

$$M_i = \langle p_i, pa_i(\Delta I), lim_i(\Delta I) \rangle,$$

де p_i – алгоритм роботи метода; $pa_i(\Delta I)$ – параметри оцінки змін якостей інформації (ΔI); $lim_i(\Delta I)$ – цільові установки, яким повинні задовольняти результати роботи правил метода, визначені через параметри зміни якостей інформації на вході та виході етапу.

Основними параметрами оцінки змін якостей інформації на i -му етапі є шум (φ_i), повнота (β_i) та час (τ_i): $pa_i(\Delta I) = \langle \varphi_i, \beta_i, \tau_i \rangle$.

Цільові установки полягають в максимальному зменшенні шумів та збільшенні повноти вихідної інформації при мінімальному часі реалізації кожного етапу:

$$lim_i(\Delta I) = \langle \varphi^\Sigma, \beta^\Sigma, \tau^\Sigma \rangle, \varphi^\Sigma = F_\varphi(\varphi_1, \varphi_2, \varphi_3) \rightarrow \min;$$

$$\beta^\Sigma = F_\beta(\beta_1, \beta_2, \beta_3) \rightarrow \max, \tau^\Sigma = F_\tau(\tau_1, \tau_2, \tau_3) \rightarrow \min.$$

Функції F_φ , F_β та F_τ оцінюють зміни якостей інформації на всіх етапах і визначаються особливостями методів обробки інформації конкретної інформаційної системи.

В цілому ефективність роботи технологічної схеми (Q) може бути представлена функцією F_Q , яка оцінює відповідність змін вхідної та вихідної інформації ($I_1^{BX} \Rightarrow I_3^{VIX}$) вимогам прикладних задач T , а також якість обробки інформації:

$$E = F_E(I_1^{BX}, I_3^{VIX}; T; \varphi^\Sigma, \beta^\Sigma, \tau^\Sigma) \rightarrow \max.$$

Етап 1. "Нормалізація класифікаторів" полягає в приведенні позиційних кодів класифікаторів до єдиного формату, додаванні не визначених та виключених зайвих інформаційних елементів та додаванні понятійного образу інформаційного елемента в структуру класифікатора.

Вхідною інформацією для першого етапу є базові та додаткові класифікатори, вихідною – нормалізовані класифікатори (K^H).

Позначимо d_i^p – інформаційний елемент, який має позиційний код p та знаходиться на i -му рівні ієрархії класифікатора, "text" – поняття, яке надано у вигляді природно-мовного виразу. Наприклад, інформаційний елемент $d_3^{6.2.1}$ відповідає поняттю "ра-

діоелектронна боротьба", яке має позиційний код "6.2.1" та знаходиться на третьому рівні ієрархії в класифікаторі "Військовий стандарт 02.104.001. Боротьба радіоелектронна. Терміни та визначення".

Формування понятійного образу для даного інформаційного елемента полягає в визначенні основ понять на основі методу квазізакінчень (наприклад, для вислову "радіоелектронна боротьба" це "радіоелектронн"+"боротьб").

Коефіцієнт шуму φ_1 визначається "випадково" невизначеними та зайвими інформаційними елементами; елементами, які мають неповні понятійні образи, та елементами, позиційні коди яких необхідно привести до загального формату.

Етап 2. "Інтеграція нормалізованих класифікаторів" K_m^H та K_n^H ($m, n \in [1, M+N]$) полягає в формуванні матриць семантичної сумісності $B_{m,n}$, елемент $b_{i,j}$ якої має значення "1" – якщо між елементом $d_{i,i}^p$ класифікатора K_m^H та елементом $d_{j,j}^p$ класифікатора K_n^H існує відношення інтеграції R_i , "0" – в іншому випадку, "X" – якщо невідомо про наявність або відсутність між даними елементами відношення інтеграції.

Таким чином, вхідною інформацією другого етапу є нормалізовані класифікатори. Визначення відношення інтеграції між елементами може здійснюватися в "ручному" та автоматизованому режимі експертом в ПрГ по формальним правилам. В загальному випадку для отримання коректних відношень інтеграції потрібно залучення додаткових (екстралінгвістичних) знань, які повинні зберігатися в тезаурусі інформаційної системи.

Сукупність всіх можливих матриць семантичної сумісності між двома класифікаторами складає загальну матрицю семантичної сумісності B , яка є вихідною інформацією другого етапу. В якості прикладу наведена узагальнена матриця B для чотирьох нормалізованих класифікаторів (табл. 1).

Таблиця 1

Узагальнена матриця семантичної сумісності для чотирьох нормалізованих класифікаторів

	K_1^H	K_2^H	K_3^H	K_4^H
K_1^H	B_{11}	B_{12}	B_{13}	B_{14}
K_2^H	B_{21}	B_{22}	B_{23}	B_{24}
K_3^H	B_{31}	B_{32}	B_{33}	B_{34}
K_4^H	B_{41}	B_{42}	B_{43}	B_{44}

В загальному випадку для повного заповнення матриці B достатньо знати показник β_2^1 , який визначає повноту інтеграції класифікаторів, що входять до складу ІПК БД, наступним чином:

$$\beta_2^1 = (n_k^2 - n_k) / 2,$$

де $n_k \in [1, M + N]$ – загальна кількість нормалізованих класифікаторів.

Коефіцієнт повноти інтеграції β_2^2 оцінює кількість елементів матриці B , які мають невизначені значення "X":

$$\beta_2^2 = \text{card}(b_{ij} | b_{ij} = "X") / \text{dim}(B),$$

де $\text{card}(M)$ – кількість елементів множини M ; $\text{dim}(B)$ – кількість елементів в матриці B .

Коефіцієнт шуму φ_2 визначається як величина зворотна коефіцієнту повноти і характеризує якість відношень інтеграції, що формуються при автоматичному заповненні матриці B .

Етап 3. "Прогнозування можливих відношень інтеграції" полягає в формуванні гіпотез, які описують усі можливі відношення інтеграції, на основі загальної матриці семантичної сумісності B (вхідна інформація). Вихідною інформацією є розширена загальна матриця семантичної сумісності B^* з новими доданими та перевіреними відношеннями інтеграції.

Фрагмент можливої матриці B^* , який графічно демонструє одне з правил прогнозування, згідно якого "якщо між елементами d_{11}^p та d_{12}^p існує відношення інтеграції, то можливо воно існує між елементами d_{12-1}^p та d_{12+1}^p , які знаходяться вище та нижче по рівню ієрархії", наведений в табл. 2.

Таблиця 2

Графічна демонстрація процесу генерації гіпотез відносно наявності відношень інтеграції між елементами матриць семантичної сумісності

	d_{12-1}^p	d_{12}^p	d_{12+1}^p
d_{11-1}^p		1	
d_{11}^p	1	↑ ← 1 ⇒ ↓	1
d_{11+1}^p		1	

МЕТОД ФОРМИРОВАНИЯ ИНТЕГРИРОВАННЫХ ПОНЯТИЙНЫХ КОМПОНЕНТОВ В ИНФОРМАЦИОННЫХ СИСТЕМАХ ОРГАНОВ ВОЕННОГО УПРАВЛЕНИЯ

А.А. Феклистов, А.В. Лупандин, А.В. Мегельбей

Рассматривается метод формирования интегрированных понятийных компонентов в информационных системах органов военного управления на основе матриц семантической смежности.

Ключевые слова: информационная система органа военного управления, база данных, матрица семантической смежности, классификатор понятий (терминов).

METHOD OF FORMATION INTERGRATIVE TERM COMPONENTS IN INFORMATION SYSTEMS OF MILITARY MANAGEMENT DEPARTMENTS

A.O. Feklistov, A.V. Lupandin, A.V. Megelbey

Consider method of formation intergrative term components in information systems of military management departments based on matrix of semantic relations.

Keywords: information system of military management department, database, matrix of semantic relations, vocabulary of words (terms).

Коефіцієнти шуму (φ_3) та повноти (β_3) визначаються відповідно кількістю коректних та некоректних гіпотез, перевірка яких здійснюється експертом в "ручному" режимі або автоматизовано за визначеними формальними правилами.

Висновки

Наведений метод формування ІПК БД інформаційних систем ОВУ на основі матриць семантичної сумісності показав практичну цінність при проектуванні інформаційної системи підтримки прийняття рішень під час підготовки та проведення тендерів з закупівлі продукції оборонного призначення для потреб ЗС України.

Список літератури

1. Замаруєва І.В. Державне та військове управління як об'єкт інформаційної боротьби / І.В. Замаруєва, С.В. Ленков, А.О. Рось // Наука і оборона. – 2007. – № 3. – С. 28-34.
2. Руснак І.С. Розвиток форм і способів ведення інформаційної боротьби на сучасному етапі / І.С. Руснак, В.М. Телелим // Наука і оборона. – 2000. – № 2. – С. 18-23.
3. Стратегічне управління військово-технічним співробітництвом в інтересах забезпечення воєнної безпеки України: монографія / В.М. Бегма, О.М. Загорка, В.О. Косеєвцов, В.М. Шемаєв; під заг. ред. І.С. Руснака. – К.: ІПНБ, НАОУ, 2005. – 228 с.
4. Оленович І.Ф. Методологія дослідження складних систем воєнного призначення / І.Ф. Оленович, А.І. Сбітнев. – К.: НАОУ, 2002. – 400 с.
5. Барабаш Ю.Л. Основи теорії оцінювання ефективності складних систем (методологія військово-наукових досліджень) / Ю.Л. Барабаш. – К.: НАОУ, 1999. – 40 с.
6. Воронин А.Н. Сложные технические и эргатические системы / А.Н. Воронин, Ю.К. Зиатдинов, А.В. Харченко. – Х.: Факт, 1997. – 240 с.
7. КНД 50-054-95. Правила побудови та написання назв об'єктів у класифікаторах техніко-економічної та соціальної інформації. – К.: Держстандарт України, 1996. – 22 с.

Надійшла до редколегії 9.09.2009

Рецензент: д-р техн. наук, проф. О.М. Сотніков, Харківський університет Повітряних Сил ім. І. Кожедуба, Харків.