

Поступила 26.01.2009р.

УДК 683.03

Т.І.Олешко

## **ФОРМУВАННЯ ТА АНАЛІЗ СЕМАНТИЧНИХ ПАРАМЕТРІВ В ІНФОРМАЦІЙНІЙ ТЕХНОЛОГІЇ**

Інформаційні засоби та інформаційна технологія в цілому, яка включає інформаційну систему, в першу чергу орієнтована на роботу з семантичними способами представлення досліджуваних систем чи процесів. У зв'язку з цим необхідно детально розглянути основні семантичні характеристики таких структур. До таких характеристик відносяться: семантична суперечність, семантичні конфлікти, семантична надмірність та семантичні відхилення [1]. Розглянемо більш детально кожен з цих характеристик.

Наведені характеристики є актуальними тільки в тому випадку, коли в рамках інформаційної технології використовуються не тільки формальні моделі, що відображають систему та її функціонування, а й семантичний опис поточного образу системи. Очевидно, що користуватися повним описом всього образу системи не доцільно, оскільки в випадку складної системи такий опис буде занадто громіздким. Тому семантичний образ ( $SO$ ) повинен описувати лише вибрані фрагменти або систему в цілому на різних рівнях узагальнення відображення. Рівень узагальнення опису образу будемо визначати через міру деталізації опису цього образу і позначати  $SO_{d_i}$ . Міру деталізації будемо враховувати наступним чином. Якщо опис образу  $SO$  складається з елементарних синтаксичних елементів, то рівень деталізації опису образу  $SO$  є найвищим, будемо записувати це в формі  $SO(d_i) = \max$ .

До елементарних семантичних елементів будемо відносити ті слова з словника  $C_s$ , які позначають базові компоненти предметної області. Фрагменти опису  $SO$  можуть представляти собою певні підсистеми і такі підсистеми можуть позначатися окремими словами. Тоді відповідні слова будемо відносити до групи слів, що є похідними базових слів стовпця  $C_s$ . Будемо позначати базові слова символом  $S_i(1)$ , тоді похідними словами будуть слова  $S_i(m)$ , де  $m$  – кількість слів, що використовуються для опису підсистеми у вигляді образу  $SO_i$ . В цьому випадку

$SO_i = \langle S_1(1), \dots, S_k(1) \rangle = S_i(k)$ . Опис деякого фрагменту системи  $SO_j$  може використовувати не базові слова, які позначають окремі фрагменти підсистеми, що входять в склад системи, образ якої описується. В цьому випадку такий опис будемо позначати через

$$S_i(m, n) = \langle S_1(1), \dots, S_i(n), \dots, S_m(1) \rangle.$$

Якщо таких фрагментів в описі  $SO$  образу  $S_i$  більше ніж два, то відповідний образ запишеться в вигляді співвідношення:

$$S_i(m, n, \dots, k) = \langle S_1(1), \dots, S_i(n), \dots, S_j(k), \dots, S_m(1) \rangle.$$

Окремі описи  $S_i(m)$  можна характеризувати деякими функціями наступним чином. Кожне слово  $S_i$  має певне інтерпретаційне розширення  $j(S_i)$ , яке описується мовою користувача. На такому інтерпретаційному розширенні визначається тим чи іншим чином оцінка відповідного слова  $S_i(1)$ . Таку оцінку будемо записувати у вигляді  $S_i(\alpha : 1)$ . У випадку похідних слів такий запис буде виглядати наступним чином:  $S_i(\alpha : m, n, \dots, k)$ . Будь-який образ у вигляді семантичного опису можна представити у вигляді похідного слова певного порядку. Таким чином можна використовувати позначення  $SO_i$  і  $S_i(m)$  як еквівалентні. Порядок похідного слова будемо визначати кількістю ієрархічних включень, що використовуються в описі  $S_i$ . Наприклад, нехай похідне слово  $S_i$  має порядок 3. Це означає, що в описі  $S_i$  або  $S_i(m, n)$  використовується трьохрівневе входження похідних слів нижчих порядків. Це можна записати наступним чином:

$$\begin{aligned} S_i^3(m, n) &= \langle S_1(1), \dots, S_i(n, k), \dots, S_m(1) \rangle = \\ &= \langle S_1(1), \dots, S_{ij}(k), \dots, S_{i(j+n)}(k), \dots, S_m(1) \rangle = \\ &= \langle S_1(1), \dots, S_{je}(1), \dots, S_{ij(e+k)}(1), \dots, S_{i(j+n)e}(1), \dots, S_{i(j+n)(e+k)}(1), \dots, S_m(1) \rangle. \end{aligned}$$

Похідне слово вищого порядку будемо записувати в наступному вигляді:  $S_i^3(m, n) = S_i(m, n(k))$ . Очевидно, що така форма запису може поширюватися на похідні слова довільного порядку. Таким чином, слово першого порядку представляє собою образ  $SO_i$ , який описується базовими словами, що мають початкові інтерпретаційні розширення. Приймемо наступні вихідні твердження.

Аксиома 1. Будь-яке словосполучення  $S_i$ , що відображає фрагмент образу  $SO_i$ , має семантичну значущість, меншу ніж сума семантичних значень окремих слів або рівну сумі семантичних значень окремих слів.

В останньому випадку мова може йти тільки про фрагмент тексту, в якому проводиться перелік окремих елементів чи об'єктів. Такі випадки не будуть розглядатися. В даній ситуації виникає задача обчислення семантичної значущості фраз, речень і абзаців, які містять текстовий образ

відповідного фрагмента предметної області інтерпретації. Для цього необхідно розглянути більш детально спосіб визначення семантичної значущості окремих слів. Семантична значущість окремого слова визначається в області власного інтерпретаційного розширення, яке представляє собою оригінальний опис фрагменту відповідної предметної області. Тому необхідно визначити метрику в середовищі інтерпретаційних розширень  $j(x_i)$ . Прикладом найпростішої метрики є кількість слів, що використовуються в даному інтерпретаційному розширенні  $j(x_i) = \sigma(x) = m_i$ .

Більш складним прикладом встановлення метрики на множині слів інтерпретаційного розширення є встановлення частоти використання окремого слова, що використовується в інтерпретаційних розширеннях інших слів семантичного словника. Тоді така частота буде числовою характеристикою окремою слова  $x_i$ , яке описується в списку використовуваних слів  $SW$ . В такому списку по складеному семантичному словнику  $C_s$  визначаються відповідні величини частот  $\lambda_i$  для кожного слова  $\alpha_i$  з інтерпретаційних описів  $I(X)$ . В цьому випадку кожне  $j(x_i)$  може бути представлено в вигляді дискретної послідовності  $j(x_i) = \langle \lambda_1(\alpha_1), \lambda_2(\alpha_2), \dots, \lambda_m(\alpha_m) \rangle$ . Тому залишається визначити функцію  $\varphi(x_i)$  на цій сукупності елементів. Прикладом такої функції може бути залежність наступного вигляду:

$$\sigma(x) = \left( \sum_{i=1}^k \lambda_i \right) / k. \quad (1)$$

Таке визначення семантичної значущості слів, що описують образи  $SO_i$  в предметній області, дозволяє в процесі функціонування системи змінювати міру семантичної значущості окремих слів  $x_i$  за рахунок зміни частоти використання слів  $x_i$  в інтерпретаційних розширеннях  $j(x_i)$ . Така зміна відбувається за рахунок розширень в  $C_s$ , яке приводить до появи нових  $x_i^*$  і, відповідно,  $j(x_i^*)$  та, як наслідок, до зміни величин  $\lambda_i$  окремих  $\alpha_i$  в  $j(x_i^*)$ .

Наведена вище функція  $\sigma(x)$  є найменш чутлива до змін  $x_i$  в  $j(x_i^*)$ , оскільки представляє собою обчислення середнього значення  $x_i$  для всіх  $\alpha_i$  з  $j(x_i^*)$ . Очевидно, що можна побудувати функції для  $\sigma(x_i) = \varphi(\lambda_i, \alpha_i)$  більш чутливі до змін в  $C_s$ . Це приводить до того, що можна всю семантичну частину інформаційної системи зробити більш динамічною при виникненні семантичних відхилень, які автоматично приводять до формування нових  $x_i$  в  $C_s$ . Функція  $\sigma(x_i)$  суттєво ускладнюється навіть у випадку використання її в вигляді (1), якщо у склад  $j(x_i^*)$  входять слова  $x_i'$  з  $C_s$ . В процесі

функціонування системи такі випадки є не тільки можливими, але й неминучими, оскільки в більшості випадків семантичні розширення відбуваються на основі використання понять і термінів, що сформовані при початковому складанні словника  $C_s$  і, відповідно, словників  $SPO_i$ . Тоді функція (1) запишеться наступним чином:

$$\sigma'(x_i) = \varphi(x_i) = \left[ \left( \sum_{i=1}^r \lambda_i \right) / r \right] + \left[ \left( \sum_{j=1}^m \lambda_j \right) / m_1 \right] + \dots + \left[ \left( \sum_{j=1}^{m_k} \lambda_j \right) / m_k \right],$$

що представляє собою суму середніх значень семантичних значущостей оригінальних складових  $\lambda_j$  з  $j(x_i)$  та середніх значень складових, що запозичені з семантичного словника  $C_s$ . Наведене співвідношення можна записати в більш інтегральній формі, якій відповідає наступне співвідношення:

$$\sigma(x_i) = \left\{ \left( \left( \sum_{i=1}^r \lambda_i \right) / r \right)_1 + \left( \left( \sum_{j=1}^m \lambda_j \right) / m_1 \right)_2 + \dots + \left( \left( \sum_{j=1}^{m_k} \lambda_j \right) / m_k \right)_n \right\} / n. \quad (2)$$

Наступною семантично значимою компонентою для текстового опису є фраза  $\psi_i(x_1, \dots, x_k)$ . В якості проміжної граматичної конструкції приймемо речення  $R$ , яке може складатися з однієї або кількох фраз  $R = \varphi_r(\psi_1, \dots, \psi_m)$ . Опис всього образу будемо представляти у вигляді одного абзацу тексту або  $SO_i = V_i(R_1, \dots, R_m)$ . Оскільки інформаційна технологія формується як деяка формалізована система [2], то можна ввести систему правил  $\gamma$ , які описують допустимі способи формування фраз, що можна записати в наступному вигляді:

$$[(x_1, \dots, x_n) \& (\gamma_1, \dots, \gamma_k)] \Rightarrow \psi_1(x_{11}, \dots, x_{1k}), \dots, \psi_n(x_{n1}, \dots, x_{nk}).$$

Створення такої системи правил у вигляді детермінованих співвідношень є допустимим завдяки тому, що предметна область, для якої вони створюються, є обмеженою і достатньо точно визначеною своїми функціональними можливостями. Аналогічно можна створити систему правил побудови речень з окремих фраз:

$$\{[\psi_1(x_{11}, \dots, x_{1k}), \dots, \psi_k(x_{k1}, \dots, x_{km})] \& (\Omega_1, \dots, \Omega_m)\} \Rightarrow R_i[\psi_1, \dots, \psi_m].$$

Виходячи з аналізу предметної області, відповідні умови можна сформулювати для формування цілих образів текстового опису, які вище називались образами  $OS_i$  фрагментів предметної області. При дослідженні семантичних закономірностей обмежимося правилами формування фраз  $\gamma = \{\gamma_1, \dots, \gamma_n\}$ . Прийmemo, що речення можуть складатися з сукупності фраз, сполучення яких в межах речення будемо проводити виходячи з умови семантичного узгодження.

Розглянемо основні семантичні закономірності в межах однієї фрази.

Оскільки опис предметної області є нормалізованим і представляє собою текстовий опис у вигляді технічної документації, то кількість правил формування фраз зведено до мінімуму. Це означає, що кількість слів різних граматичних типів є обмеженою у порівнянні з текстами інших типів, наприклад, з технічною літературою і тим більше в порівнянні з текстами художньої літератури.

Розглянемо метод оцінки семантичної значущості фрази, яка складається з певної кількості слів. Кожне слово  $\lambda_i$ , що використовується в  $j(x_i)$ , має певним чином визначену семантичну значущість  $\alpha_i$  або  $\lambda_i(\alpha_i)$ . Два слова  $x_i$  і  $x_j$ , що відносяться до одного граматичного типу, якими можуть служити іменники, дієслова, прикметники і т.п., описуються інтерпретаційними розширеннями  $j(x_i)$  і  $j(x_j)$ . Такі розширення можуть в ідеальному випадку використовувати абсолютно різні слова мови інтерпретації або має місце:

$$\forall(\lambda_i \in j(x_i))\forall(\lambda_j \in j(x_j))[\lambda_i \neq \lambda_j].$$

Практично ідеальна ситуація або взагалі не існує в області  $j(x_i)$  або її існування є окремим випадком. Найчастіше має місце ситуація, коли області  $j(x_i)$  і  $j(x_j)$  частково перекриваються. Прийmemo наступну аксіому.

Аксіома 2. Послідовність використання слів одного граматичного типу визначається величиною їх семантичної значущості.

Прикладом правила, що ґрунтується на цій аксіомі, може бути наступна граматична функція, що запишеться в вигляді:

$$\gamma_i(x_i, x_j) = [\sigma(x_i) > \sigma(x_j)] \Rightarrow \psi_i[x_i, x_j].$$

Такі правила формуються наступними способами:

- формується початковий текст технічного опису предметної області для кожного фрагменту, по сформульованому тексту будуються правила формування фраз  $\psi_i$ ;
- формуються правила  $\gamma_i$  формування фраз, що описують технічний текст предметної області і по цих правилах складається опис фрагментів опису у вигляді окремих образів  $SO_i$ ,
- формуються аксіоми, які приймаються як вихідні положення або принципи, на базі яких формуються правила побудови фраз, на основі таким чином сформованих правил будується текстовий опис фрагментів предметної області інтерпретації.

Формально наведені правила записуються у вигляді наступних співвідношень:

$$\begin{aligned} & [\psi_i(x_1, \dots, x_i), \dots, \psi_k(x_i, \dots, x_j)] \rightarrow (\gamma_1, \gamma_2, \dots, \gamma_k) \\ & [\gamma_1, \dots, \gamma_z] \rightarrow [\psi_i(x_1, \dots, x_k), \dots, \psi_m(x_i, \dots, x_k)] \\ & [A_1, \dots, A_z] \rightarrow [\gamma_i(x_i, x_j), \dots, \gamma_k(x_1, x_e)] \rightarrow \psi_i(x_1, \dots, x_n), \dots, \psi_k(x_i, \dots, x_k). \end{aligned}$$

Очевидно, що способи формування правил можуть представляти собою синтез наведених вище способів. Ці способи використовуються в першу чергу для початкового формування текстових образів  $SO_i$  предметної області. В процесі функціонування телекомунікаційних систем можуть виводитись нові правила формування окремих фраз текстового опису.

В даному випадку такі правила формуються на основі семантичних параметрів, які є спільними для всіх слів, що визначаються словником  $C_s$ . До цього часу в якості такого параметра розглядалась семантична значущість або семантична вага окремого слова. При збільшенні кількості таких семантичних параметрів існує можливість збільшити кількість правил побудови фраз, кількість способів їх формування і, відповідно, виникає необхідність формулювати нові аксіоми, що відображають роль нового семантичного параметра.

Введемо такий семантичний параметр як міра семантичної оригінальності слова, яке визначається в словнику  $C_s$ . Розглянемо наступне визначення цього параметра.

Визначення. Міра семантичної оригінальності слова  $x_i$ , яку будемо позначати  $\mu(x_i)$ , визначається як кількість слів в описі інтерпретаційного розширення  $j(x_i)$ , які вже визначені в словнику  $C_s$ .

Міра семантичної оригінальності слова  $x_i$ , є максимальною, якщо в  $j(x_i)$  не існує ні одного  $x_j$  з  $C_s$ . Міра оригінальності зменшується прямо пропорційно до збільшення кількості слів з  $C_s$  при формування опису  $j(x_i)$ . Міра семантичної оригінальності визначається в діапазоні [0, 1] відповідно з наступним співвідношенням:

$$\mu(x_i) = 1 - \frac{m}{n},$$

де  $m$  – кількість слів з  $C_s$  в  $j(x_i)$ ,  $n$  – загальна кількість слів в  $j(x_i)$ .

Завдяки параметру  $\mu(x_i)$  існує можливість характеризувати окремі фрази мірою деталізації опису образу  $SO_i$  або  $S_i(m)$ , який відповідна фаза описує. Цей параметр використовується при формуванні правил побудови речень, які складаються з ряду фраз.

В рамках даного підходу існує можливість зв'язати з семантикою такий параметр як час. Суть цього параметра полягає в наступному. В процесі функціонування системи можуть формуватися нові слова, які представляють собою наступне по відношенню до системи:

- позначення того чи іншого узагальнення, яке доцільно використовувати при описі поточного функціонування системи,
- позначення нових процесів, які виникають в системі як результат розширення її функціональних можливостей чи процесів, що виникають в

системі в процесі її еволюції,

- слова, що позначають новий базовий елемент системи, який вводитьься в систему при її модифікації.

Введемо наступне визначення, що стосується часових аспектів, що пов'язуються з семантикою.

Визначення. Семантичною координатою часу інформаційної системи будемо вважати умовно проведену вісь вздовж списку слів в словнику  $C_s$ .

Початком відліку семантичного часу  $\tau$  будемо вважати точку на координаті часу, яка відповідає останньому по порядку слову в стовпці  $C_s$ , яке впроваджене як базове.

Всі слова, які появляються в системі в процесі її функціонування, послідовно дописуються в список слів словника  $C_s$ , продовжуючи умовну вісь часу новою точкою відліку. В цьому випадку семантичний інтервал часу, який проіснувала система протягом деякого фізичного інтервалу часу  $\Delta t_i$ , вимірюється кількістю нових слів, які появилися в системі, а точніше в текстових образах системи чи просто в семантичних та інформаційних компонентах системи. Будемо говорити, що  $\tau(\delta S)$  є семантичним часом існування системи. Отже, якщо протягом певного часу функціонування системи остання обходиться тільки набором базових слів, а нові слова не з'являються, то будемо говорити, що така система в семантичному сенсі або часі не функціонує. Очевидно, що з кожним новим словом, що появляється в словнику  $C_s$ , появляється новий інтервал часу семантичного функціонування системи. Величину семантичних інтервалів функціонування системи або інформаційної компоненти системи будемо визначати як різницю між семантичною оригінальністю поточного слова і попереднього слова, які з'являються в словнику  $C_s$ .

Величини семантичних інтервалів  $\Delta\tau(\delta S)$  будемо формально описувати співвідношенням:

$$\tau(\delta S_i) = \mu(x_i) - \mu(x_{i-1}),$$

де  $\mu(x_{i-1})$  – міра семантичної оригінальності, що визначалась при попередньому розширенні словника  $C_s$ ,  $\mu(x_i)$  – міра оригінальності поточного семантичного розширення. Завдяки параметру  $\Delta\tau_i(\delta S_i)$  можна встановити зв'язок між параметром  $\mu(x_i)$  і семантичним часом. А через параметр  $\mu(x_i)$  можна встановити зв'язок з параметром  $\sigma(x_i)$ .

З метою більшого відображення семантичного часу розглянемо поняття швидкості зміни семантичного часу. Введемо уявлення про ознаки еволюційного розвитку системи. В кожний поточний момент система в цілому володіє певною сумарною семантичною значимістю, яка визначається по ефективному словнику  $C_s$ , яку можна обчислити по співвідношенню:

$$\sigma_c = (C_s) = \sum_{i=1}^n \sigma(x_i).$$

Крім того, в рамках інформаційних засобів телекомунікаційних систем існують описи образів  $SO_i$  фрагментів предметної області, до якої відноситься система, що описуються співвідношеннями, побудованими на основі використання співвідношення (2), яке можна записати

$$\sigma(S) = \sum_{i=1}^m \sigma(S_i); \quad \sigma(S_i) = \sum_{i=1}^n \sigma(\psi_i),$$

де  $\sigma(\psi_i)$  обчислюється за допомогою співвідношення:

$$\sigma(\psi_i) = \left( \sum_{i=1}^n \sigma(x_i) \right) \& (x_i \in \psi_i) - \sum_{i=1}^n (\sigma(x_i)) \cap \sigma(x_{i+1}).$$

В цьому випадку ознака еволюційного розвитку системи запишеться у вигляді співвідношення:

$$[\eta(S)]_i = \{[\sigma(S)]_{i+1} - [\sigma(S)]_i\} / [\Delta \tau_i (\delta S)]_i. \quad (3)$$

Для кожної окремої реалізації телекомунікаційної системи на основі ознаки  $\eta(S)$  можна встановити критерій здійснення еволюційного кроку, який задається у вигляді певного граничного значення для  $\eta(S)$ . Таким чином можна записати наступне:

$$K_e(\eta) = (\eta_i(S) - \eta_{i-1}(S)) - A_c, \quad (4)$$

де  $A_c$  – деяка гранична константа, що визначає величину зміни критерія еволюційного розвитку не менше, ніж на задану константу  $A_c$ .

На змістовному рівні  $\eta(S)$  означає, що завдяки збільшенню семантичної значущості всієї системи, яка обумовлена введенням нових слів, що можуть позначати нові поняття, та яка не включає семантичну значущість самого словника  $C_s$  по відношенню до величини зміни оригінальності слів, які вводяться, або до величини семантичного часу системи, здійснюється прогресивний розвиток системи  $S$  в цілому. Якщо такий прогресивний розвиток, що обчислюється у відповідності з співвідношенням (3), більший деякої заданої константи  $A_c$ , то має місце один крок еволюційного процесу в системі  $S$ . Будемо позначати процес еволюційного розвитку системи у вигляді послідовності:

$$\eta_1(K_e, S) \rightarrow \eta_2(K_e, S) \rightarrow \dots \rightarrow \eta_m(K_e, S).$$

Швидкість зміни семантичного часу в системі будемо визначати співвідношенням:

$$V(\Delta \tau) = [\eta_i(K_e, S) - \eta_{i+1}(K_e, S)] / \Delta t_i,$$

де  $\Delta t_i$  – інтервал фізичного часу функціонування системи, протягом якого відбулася зміна рівня розвитку системи, в даному випадку еволюційного



розвитку системи рівна  $\eta_i(K_e, S) = \eta_i(K_e, S) - \eta_{i+1}(K_e, S)$ . Якщо  $\Delta\tau$  не міняється, то це означає, що в рамках системи не відбувається розвитку.

1.Емельянов В.В.,Курейчик В.В.,Курейчик В.М. Теория и практика эволюционного моделирования.- М.:Физматгиз, 2003.

2.Олешко Т.І.Дослідження взаємозв'язку між інформаційною моделлю та базами даних, що входять в склад інформаційної технології. Захист інформації – 2005. Спецвипуск, с.30-35.

*Поступила 28.01.2009р.*

УДК 683.03

Б.В.Дурняк, Ю.М.Коростиль, Г.А.Максименко, В.И.Сабат

## **ОСОБЕННОСТИ ФОРМИРОВАНИЯ КОМПОНЕНТ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ СИСТЕМ МОНИТОРИНГА**

Одной из главных компонент систем обобщённого мониторинга или расширенной системы мониторинга (ОСМ) является информационная модель социальных объектов (СО), а также и других компонент, которые используются в системе ОСМ. Особенностью такой информационной модели является способ её представления. Очевидно, что информационная модель, как и любая другая модель описывает определённые объекты. В соответствии с классическими представлениями о модели [1], последняя должна не только описывать моделируемый объект, но и моделировать процессы, которые происходят в объекте в процессе его функционирования. Одной из важных особенностей рассматриваемой информационной модели является способ описания объекта моделирования, который представляет собой составление текста описания на естественном языке. Поэтому, в дальнейшем будем под информационной моделью подразумевать текстовую информационную модель (ТИМ). Как принято, при составлении описания модели, будем её текстовое описание формировать в виде некоторой семантической иерархической структуры. Это означает, что в первую очередь будет формироваться самое общее описание моделируемого объекта. После этого описания будет размещаться описание более детальное с точки зрения их семантических значений. Этот процесс соблюдения иерархических уровней описания будет продолжаться до тех пор, пока в текстовом виде не окажутся описанными все детали и элементы моделируемого объекта, которые определяют точность или адекватность его моделирования. В этом описании отдельные уровни иерархии будут разделяться условными разделами, которые могут иметь своё отдельное текстовое описание, а элементы одного