

УДК 004.451

Б. В. Дурняк, О. Ю.-Ю. Коростіль

Українська академія друкарства

**СПОСОБИ РОЗШИРЕННЯ СЕМАНТИЧНИХ ПАРАМЕТРІВ
ДЛЯ ТЕКСТОВИХ МОДЕЛЕЙ**

Досліджуються способи розширення базових семантичних параметрів, а саме: семантична значимість, семантична суперечність, семантичний конфлікт для побудови текстових моделей.

Ключові слова: *текстові моделі, семантичні параметри.*

Текстові моделі (ТМ) призначені не тільки для опису об'єктів соціальних (SO_i) або інших, які важко описати формальними засобами, для їх статичного відображення, а й для опису їх взаємодії з зовнішніми факторами та взаємодії між собою. Це обумовлює необхідність розширення асортименту параметрів, що їх описують. Оскільки йдеться про текстові форми опису, розширенню підлягають семантичні параметри. До способів розширення семантичних параметрів можна віднести:

- додавання фрагментів тексту до існуючого текстового опису деякого об'єкта;
- елімінацію окремих фрагментів з тексту;
- перестановку фрагментів тексту в межах текстового опису всієї моделі.

Для реалізації відповідних перетворень необхідно розв'язати певні задачі:

- виявлення фрагмента тексту, найменший розмір якого може представляти одну фразу ϕ_i , з точки зору його семантичної значимості чи семантичних характеристик;
- визначення місця входження відповідного фрагмента тексту та куди саме відповідний фрагмент передбачається перенести у випадку елімінації відповідного фрагмента, координати місця його переміщення приймаються рівними нулю;
- встановлення другого фрагмента тексту, який переноситиметься на місце фрагмента, що вибраний першим для перестановки.

Такі перетворення є замкнутими, оскільки вони реалізуються в межах деякого текстового опису ТМ_i. Відкритий тип перетворень — це процес виве-

дення нових фрагментів тексту, якими буде окрема фраза чи речення. До відкритих перетворень можна віднести: процеси виводу фраз чи речень; процеси генерації фрагментів тексту; обумовлені перетворення, що можуть здійснюватися в межах текстової моделі.

Для реалізації процесів виводу, крім семантичних параметрів, використовуються граматичні структурні перетворення, які ґрунтуються на використанні правил граматики та визначенні класів слів і словосполучень. Відповідні граматичні структури визначають допустимі схеми взаємного розміщення різних класів у схемах фраз ϕ_i чи речень ψ_i . Такі правила є нормалізованими.

Розглянемо визначення, що описують обмеження, у рамках яких окреслюються перетворення та особливості відповідних розширень семантичних параметрів.

Визначення 1. Взаємодія TM_i із зовнішнім оточенням полягає в доповненні TM_i текстовими елементами $IP_i \in Q(tm_i, \dots, tm_n)$, що адресовані відповідним TM_i , та в аналізі й перетвореннях відповідних розширень текстових описів, або:

$$TM_i^* = F[TM_i, Q(tm_i, \dots, tm_n)],$$

де F — функція, що описує семантичний аналіз тексту TM_i , доповненого розширенням $Q(tm_i, \dots, tm_n)$.

Визначення 2. Взаємодія TM_i з $Q(tm_i, \dots, tm_n)$ є допустимою, якщо виконується умова

$$A(TM_i) \geq k_i Q(tm_i, \dots, tm_n),$$

де k_i — коефіцієнт, що визначає міру допустимої взаємодії, TM_i з $Q(tm_i, \dots, tm_n)$, A — функція визначення розміру TM_i .

Визначення 3. Величина ефективності дії $Q(tm_i, \dots, tm_n)$ на TM_i визначається мірою семантичної узгодженості $Q(tm_i, \dots, tm_n)$ з TM_i , що формально описується співвідношенням:

$$\sigma(TM_i) \Rightarrow \mu \cdot \sigma[Q(tm_i, \dots, tm_n)],$$

де $\sigma(TM_i)$ та $\sigma[Q(tm_i, \dots, tm_n)]$ — інтегральні семантичні параметри, що характеризують TM_i і $Q(tm_i, \dots, tm_n)$; μ — міра їх семантичної узгодженості.

Нормалізовані текстові описи (TO) характеризуються цілим рядом семантичних параметрів, значення яких змінюється в певному діапазоні, який обмежується максимальним і мінімальним значеннями $\sigma^i := [\alpha^i, \beta^i]$. Згідно з такою інтерпретацією семантичних параметрів можна прийняти, що в межах відповідних діапазонів $[\alpha^i, \beta^i]$ вони змінюються лінійно. Щоб можна було їх суміщати з точки зору масштабів їх вимірювання, необхідно перейти до відносного способу визначення величини семантичних параметрів, прийнявши $\alpha^i = 0$ і $\beta^i = 1$. В такому випадку інтегральний семантичний параметр можна визначати як середнє значення відносних величин усіх параметрів, що описується співвідношенням:

$$\sigma^I = (\sum_{i=1}^m \sigma^i) / m. \quad (1)$$

Щоб $Q(tm_1, \dots, tm_n)$ міг вплинути на TM_i , міра семантичної узгодженості повинна бути більшою від одиниці.

Формальна семантика ґрунтується на використанні семантичних параметрів, значення яких визначаються числами. Для таких параметрів вводяться визначення, які дозволяють обчислювати їх величини [1]. До таких семантичних параметрів відносяться:

- семантична суперечність σ^S ,
- семантичний конфлікт σ^K ,
- семантична надмірність σ^N ,
- семантична значимість σ^Z тощо.

Інтерпретаційна семантика визначається мірою зв'язку між окремими компонентами, що описують елементи предметної області, або між фразами, які для інтерпретаційної семантики є одиничними елементами. Як одиничні елементи можуть використовуватися окремі речення, якщо останні нормалізовані й у рамках тексту, який досліджується, таких речень є не менше заданої кількості. Критичною величиною такої кількості виберемо число сім. Очевидно, що інтерпретаційна семантика, яку будемо називати i -семантикою ($i\sigma$), пов'язана з формальною семантикою σ , і такий зв'язок полягає в наступному:

- формальна семантика σ найефективніше використовується в рамках окремих фраз і оцінює величину відповідних зв'язків між словами фрази на основі тим чи іншим способом прийнятої методики оцінки інтерпретаційних описів відповідних слів;
- інтерпретаційні семантичні параметри $i\sigma$ використовуються для визначення семантичних зв'язків між окремими фразами чи окремими реченнями.

Можна аналізувати формальні семантичні параметри між фразами з використанням узагальненого семантичного параметра фрази як деякої середньої величини параметра фрази. При цьому на рівні фрази частково нівелюється значимість інтерпретаційного опису окремого слова, що визначається за семантичним словником S_C , оскільки в рамках фрази з'являється власна її інтерпретаційна значимість. Це зумовлюється тим, що окремі слова в рамках цілої фрази втрачають свою первинну значимість на рівні міжфразових з'єднань. Така часткова втрата на рівні міжфразових з'єднань окремими словами своєї семантичної значимості обумовлюється тим, що ціллю формування фрази чи речення є створення нового інтерпретаційного значення, яке щонайменше виступає узагальненням інтерпретаційних значень слів, з яких складається фраза. Без цього в рамках всієї фрази потрібно було б використовувати не окремі слова, а цілі описи їх інтерпретації. Підсумовуючи сказане, введемо наступні визначення.

Визначення 4. Інтерпретаційний семантичний параметр $i\sigma$, який формується на основі аналізу семантичного віртуального словника S_V , що вміщує ідентифікатори всіх основних фраз, які знаходяться в описі текстової моделі TM_i та у всіх інформаційних потоках, які активізовані й направлені на відповідну TM_i , визначається з інтерпретаційного опису фрази у відповідному словнику, що використовує розширене текстове відображення відповідної фрази.

Формально елемент віртуального семантичного словника описується співвідношенням:

$$(\varphi_i \equiv x_i^\varphi) := x_i^\varphi = \langle a_1^\varphi, \dots, a_n^\varphi \rangle I \langle P_1^\varphi, \dots, P_1^\varphi \rangle,$$

де $[a_i^\varphi \equiv (x_i \in S_C)] = \langle a_{i1}, \dots, a_{im} \rangle I \langle P_{i1}, \dots, P_{ik} \rangle$, P_{ij}^φ — значення деякого атрибуту фрази φ_i , якщо в останню включено числові або додаткові величини.

Виходячи зі складу S_V , можна записати, що $S_V = \{TM \cup IP(T)\}$. Словник S_V в момент t_0 формується на основі множини текстових моделей, що описують певну предметну область $\{TM_1, \dots, TM_n\}$. В міру активізації інформаційних потоків $IP_i(\Delta T_{i1}, \dots, \Delta T_{in})$ по відношенню до окремих TM_i словник S_V розширюється за рахунок фраз з кожного потоку IP_i . Величина $i\sigma^Z$ обчислюється за аналогією з обчисленням σ^Z на основі використання S_C , і може змінюватися за появи нових IP_i . Прикладом одного з можливих способів визначення величини $i\sigma^Z$ може служити визначення кількості випадків використання окремої фрази ϕ_i у текстах, що описують TM_i . У цьому випадку інтегральний інтерпретаційний семантичний параметр можна визначити як середнє значення відносних величин усіх інтерпретаційних семантичних параметрів відповідно до виразу (1). Слід зауважити, що різні типи $i\sigma$ відповідають типам σ^i , які приведені вище, наприклад, семантичної суперечності, семантичного конфлікту і т.д.

Структурна семантика ($s\sigma$) відображає зв'язки між окремими реченнями ψ_i і ψ_j текстових описів TM_i . Оскільки семантична залежність між реченнями в TM відрізняється від семантичної залежності між фразами, це приводить до можливості просторового рознесення семантичних зв'язків речень у структурі всього текстового опису TM_i . Така структурна розподіленість тексту відображає типові випадки розриву одних сюжетних послідовностей іншими. Для визначення структурної семантики, або $s\sigma$, необхідно ввести уявлення про сюжетну послідовність речень у фрагментах тексту, прикладом якого може служити текстовий опис окремої текстової моделі TM_i .

Визначення 5. Величина структурного семантичного параметра $s\sigma$ між двома вибраними реченнями визначається різницею середніх значень інтерпретаційних семантичних параметрів відповідних речень.

Формально це можна описати співвідношенням

$$s\sigma(\psi_i, \psi_j) = |[\sum_{k=1}^g i\sigma(\varphi_{ki})]/g] - [\sum_{e=1}^r i\sigma(\varphi_{rj})/r]|.$$

Визначення 6. Сюжетною лінією фрагмента $\Phi(\psi_1, \dots, \psi_n)$, де ψ_i — речення, є така їх послідовність, для якої величина $s\sigma(\psi_i, \psi_j)$ більша заданого порогу $\delta(s\sigma)$ для довільної пари речень $\psi_i, \psi_j \in \Phi$, що описується співвідношенням:

$$[s\sigma(\psi_i, \psi_j) \geq \delta(s\sigma)] \rightarrow [h_i(\Phi(\psi_1, \dots, \psi_n))],$$

де h_i — елемент сюжетної лінії в тексті Φ .

Очевидно, що сюжетна лінія $h_i(\Phi)$ складає фрагмент сюжету. Сюжет загалом деякого текстового опису TM_i може описуватися довільною сукупністю сюжетних ліній, які можуть між собою перетинатися. В такому випадку уявлення про сюжет тексту дещо відрізняється від уявлення про сюжет, прийняте у художніх творах і, відповідно, в семантиці таких текстів [2]. В останніх сюжет означає опис деякої послідовності подій або дій, реалізація яких описується відповідним текстом. У рамках запропонованої роботи тексти, які є описами деякого об'єкта, для прикладу TM_i , також мають сюжет. Це означає, що описи окремого об'єкта та його фрагмента являють собою сюжети різних рівнів. Окремий сюжет $H_i(\Phi_i)$ буде відповідати опису окремого об'єкта чи опису окремих фрагментів об'єкта. Отже, введемо уявлення про ранг сюжету.

Визначення 7. Якщо деякий текст tm_i описує структурно або функціонально окрему компоненту об'єкта SO_i , який складається з окремих чи виділених компонент, текстові описи tm_s , що описують окремі компоненти, мають ранг відповідного сюжету $R[h_i(tm_i)]$, який є нижчим у відношенні до рангу сюжету $R[H(TM_i)]$, що описує SO_i в цілому.

Враховуючи сформульоване визначення, можна стверджувати, що чим дрібнішу складову деякого об'єкта SO_i описує фрагмент тексту, що має власний сюжет, тим нижчий ранг відповідного сюжету. Очевидно, що має місце:

$$[h(tm_i) = \min] \rightarrow h(\psi_i, \psi_j).$$

Це означає, що мінімальний сюжет повинен мати хоча б два речення. Якщо деякий сюжет описує виділену компоненту об'єкта SO_i , це означає, що $\min K(SO_i) \rightarrow \min R(h_i)$.

Уявлення про сюжетні лінії опису об'єкта SO_i дозволяє реалізовувати переходи від текстових описів TM_i об'єктів SO_i до структурних описів відповідних TM_i . Структурне представлення TM_i дозволяє проводити аналіз відповідних SO_i на доволі загальному рівні, що особливо важливо для розв'язку задач формування інформаційних потоків, які забезпечували б ефективну дію на вибраній SO_i [3]. Окрім того, досить важливою є задача моделювання взаємодії різних типів SO_i між собою в рамках деякого середовища. Особливо актуальна задача прогнозування моментів і причин активізації окремих SO_i до взаємодії з іншими SO_j з відповідного оточення SO_i [4].

1. Афанасьєва О. Ю. Методи семантичних перетворень в стеганосистемах / О. Ю. Афанасьєва // Моделювання та інформаційні технології. — К. : ІПМЕ НАН України, 2010. — Вип. 56. — С. 188–196. 2. Андрусенко Т. Б. Лингвистические структуры в компьютерных учебных средах / Т. Б. Андрусенко. — К. : Наук. думка, 1994. — 160 с. 3. Чернявская В. Е. Лингвистика

текста. Поликодовость. Интертекстуальность / В. Е. Чернявская. — М. : ЛИБРОКОМ, 2009. — 248 с. 4. Шурыгин А. М. Математические методы прогнозирования / А. М. Шурыгин. — М. : Горячая линия – Телеком, 2009. — 180 с.

СПОСОБЫ РАСШИРЕНИЯ СЕМАНТИЧЕСКИХ ПАРАМЕТРОВ ДЛЯ ТЕКСТОВЫХ МОДЕЛЕЙ

Статья посвящена анализу способов расширения базовых семантических параметров, к которым относятся: семантическая значимость, семантическое противоречие, семантический конфликт для описания текстовых моделей.

METHODS OF EXPANSION OF SEMANTIC PARAMETERS FOR TEXT MODELS

The methods of expansion of base semantic parameters are investigated, namely: semantic meaningfulness, semantic contradiction, semantic conflict for the construction of text models.

УДК 004.451

Б. В. Дурняк

Українська академія друкарства

М. М. Кляп

*Державний вищий навчальний заклад
«Ужгородський національний університет»*

ОСНОВНІ МЕТОДИ ПРОГНОЗУВАННЯ ПОДІЙ РІЗНИХ ТИПІВ

У статті проаналізовано методи прийняття керівних рішень при реалізації технологічних процесів у друкарському виробництві та здійснено прогнозування подій різних типів.

Ключові слова: *прогнозування подій, регресійні моделі, авторегресія.*

Для прийняття адекватних керівних рішень при реалізації технологічних процесів у друкарському виробництві необхідно розв'язувати певні задачі:

- оцінити ряд параметрів, які характеризують ситуацію, що не передбачуваним способом впливає на процес;
- здійснити аналіз зв'язку конкретної ситуації з подіями, наявними в технологічному процесі до моменту настання непередбачуваної події;
- прийняти рішення щодо реалізації адекватної реакції системою управління (особливо при управлінні технологічним процесом у напівавтоматизованому режимі, коли обслуговуючий персонал приймає рішення та реалізує відповідну керівну дію).