

ФОРМАЛЬНА ОЦІНКА ПРОЦЕСІВ ОБРАЗНОГО АНАЛІЗУ ТЕКСТУ

Вінницький національний технічний університет

Анотація

Запропоновано формальний критерій процесів образного аналізу тексту на основі моделі одиниці сенсу природно-мовної інформації. Визначено, що модель одиниці сенсу є технологічно досяжною засобами сучасних лінгвістичних пакетів.

Ключові слова: текст, образний аналіз, критерій, одиниця сенсу, нечітке відношення, функція належності.

Abstract

The formal criterion of the processes of figurative analysis of the text based on the model of the sense unit of the natural language information is proposed. It is determined that the sense unit is a technologically achievable by modern linguistic packages.

Keywords: text, figurative analysis, criterion, unit of meaning, fuzzy relation, membership function.

Вступ

Зазвичай поняття образного аналізу літературного твору означає глибоке вивчення сутності його головних героїв та їх відносин як результат літературознавчого дослідження [1]. Такого роду дослідження тісно пов'язані з розкриттям змісту твору та його сюжетних ліній, визначенням мотивів та сенсу дій (вчинків) провідних персонажів.

Зрозуміло, що якісне проведення образного аналізу вимагає залучення висококваліфікованих експертів, а тому застосовується, в основному, для дійсно помітних літературних творів. Проте моделювання процесу образного аналізу має важливе значення для кожного тексту, оскільки дозволяє отримати формальні оцінки суттєвих семантичних ознак цієї текстової інформації [2]. Поняття образного аналізу лінгвістичної інформації введено у [3], а у [4] показано місце та роль образного рівня у традиційній тріаді морфологія–синтаксис–семантика засобів комп'ютерної лінгвістики.

Метою роботи є формальне визначення критерію процесів образного аналізу тексту на основі моделювання одиниці сенсу природно-мовної інформації.

Результати дослідження

Формалізація процесів образного аналізу тексту має визначити головні поняття, що фігурують у тексті та зв'язки між ними. Будемо вважати, що всі слова тексту є вербальними ознаками деякої множини образів I . Використаємо бінарне нечітке відношення, що задане на одній базисній множині (універсумі) образів I , яке визначимо як нечітке відношення

$$Q = \{ \langle i_l, i_j \rangle, \mu_Q(\langle i_l, i_j \rangle) \}, \quad (1)$$

де $\mu_Q(\langle i_l, i_j \rangle)$ – функція належності нечіткого бінарного відношення, що задається як відображення $\mu_Q : I \times I \rightarrow [0, 1]$, а $\langle i_l, i_j \rangle$ кортеж з двох елементів, причому $i_l \in I$, $i_j \in I$. Якщо носій Q_S нечіткого відношення Q є скінченним, то потужність цього нечіткого відношення дорівнює кількості кортежів його носія [5] і позначається як $card(Q_S)$.

Розглянемо можливі способи формального визначення бінарних нечітких відношень (1) на універсумі I , зокрема:

- Список з безпосереднім перерахуванням всіх кортежів нечіткого відношення та відповідних цим кортежам значень функцій належності – у випадку скінченного та невеликого числа таких корте-

жів.

- Аналітично у формі деякого математичного виразу, що забезпечує можливість обчислення значення функцій належності для кожного з кортежів. У цьому випадку нечітке відношення записують у вигляді (1), де

$$\mu_Q(<i_l, i_j >) = f(i_l, i_j) - \quad (2)$$

певна задана функція двох змінних з універсуму I , що задовольняє стандартні вимоги до функції належності. Функцію (2) називають також [5] функцією, що породжує нечітке відношення (1).

- Графічно у формі певної поверхні або сукупності окремих точок у тривимірному просторі, в якому дві координати (незалежні змінні) відповідають значенням елементів i_l та i_j з універсуму I , а третя координата – функції належності зі значенням з інтервалу $[0, 1]$.
- У формі квадратної матриці A_Q нечіткого відношення Q , рядкам якої відповідають перші елементи кортежів $i_l \in I$, а стовпцям – другі елементи кортежів $i_j \in I$. Елементами (l, j) матриці A_Q є відповідні значення функції належності $\mu_Q(<i_l, i_j >)$ цього відношення.
- У формі орієнтованого нечіткого графа $G = (V, E, \mu_G)$, де $V = \{v_1, v_1, \dots, v_n\}$ – множина вершин нечіткого графа, а $E = \{e_1, e_1, \dots, e_m\}$ – множина дужок нечіткого графа, кожній з яких приписано значення функції належності. Натуральні числа n та m визначають кількість вершин та дужок нечіткого графа ($card(Q) = m$), а кожному елементу універсуму $i_l \in I$ відповідає окрема вершина $v_l \in V$ ($l = \overline{1, n}$) орієнтованого нечіткого графа. Довільному k -му кортежу нечіткого відношення $<i_l, i_j > \in Q$ відповідає дужка графа $e_k = <v_l, v_j >$ з початком у вершині v_l , закінченням у вершині v_j та значенням функції належності $\mu_Q(<i_l, i_j >)$.

Бінарне нечітке відношення (1) застосуємо як характеристику множини пар образів, що описуються значимими поняттями тексту – тоді функцію належності (2) можна вважати природною чисельною мірою сенсу. Згідно з [5], значення $\mu_Q(<i_l, i_j >) = 1$ називатимемо одиницею сенсу розміром один *Sav* (Синтагматичної асоціації вага, рос. – Синтагматической ассоциации вес) або *Saw* (Syntagmatic association weight). Окрім цього, англійське слово *Saw* (друга форма неправильного дієслова *to See*) перекладається як «побачене» і, тим самим, вказує на образний шлях появи сенсу в лінгвістичній системі. При такому підході значення елемента (l, j) матриці A_Q або k -ї дужки $e_k = <v_l, v_j >$ графа G залежить від статистики появи зв'язку для кортежу $<i_l, i_j >$ в процесі образного аналізу тексту.

Отже, в загальному вигляді функцію належності (базовий рівень) нечіткого відношення сенсу для пар образів задамо як

$$\mu_Q(<i_l, i_j >) = f(k_{lj}, t_L), \quad (3)$$

де k_{lj} – кількість зафіксованих лінгвістичною системою зв'язків між l -м та j -м образами на момент часу t_L . Значення k_{lj} неважко отримати автоматизовано, через парсер вільно доступних лінгвістичних пакетів шляхом підрахунку кількості синтаксичних зв'язків між словоформами речень певного тексту. Мету дослідження досягнуто – на основі технологічно досяжної моделі одиниці сенсу природно-мовної інформації (3) отримано формальний критерій процесів образного аналізу тексту.

Висновки

Запропоновано формальний критерій процесів образного аналізу тексту на основі моделі одиниці сенсу природно-мовної інформації. Критерій характеризує силу зв'язку між вербальними ознаками образів тексту через функцію належності нечіткого бінарного відношення. Визначено, що модель одиниці сенсу є технологічно досяжною засобами сучасних лінгвістичних пакетів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Губа Л. В. Художній текст як об'єкт лінгвістичних досліджень / Л. В. Губа // Наукові записки Національного університету "Острозька академія". Серія : Філологічна. - 2015. - Вип. 58. - С. 93-96. - Режим доступу: http://nbuv.gov.ua/UJRN/Nznuoaf_2015_58_37.
2. Бісікало О.В. Образний аналіз текстової інформації з Wikipedia / О.В. Бісікало, І.О. Назаров // Вісник інженерної академії України. — 2012. — № 2. — С. 52–56.
3. Бісікало О.В. Образний аналіз та синтез лінгвістичної інформації / О. В. Бісікало, І. В. Богач // Інтелектуальні технології лінгвістичного аналізу : тези доповідей міжнар. наук.-техн. конф. (Київ, 22-23 жовтня 2013 р.). — Київ : НАУ. — 2013. — С. 11.
4. Бісікало О.В. Формальне введення образного рівня до традиційної лінгвістичної тріади морфологія–синтаксис–семантика / О.В. Бісікало, І.В. Богач // Бионика интеллекта. — 2013. — № 2 (81). — С. 27–30.
5. Бісікало О.В. Формалізація понять мовного образу та образного сенсу природно-мовних конструкцій / О.В. Бісікало // Математичні машини і системи. – 2012. – № 2. – С. 70–73.

Бісікало Олег Володимирович — д-р техн. наук, професор, декан факультету КСА, Вінницький національний технічний університет, м. Вінниця, e-mail: obisikalo@gmail.com

Ляховченко Наталія Володимирівна — к.п.н., доцент кафедри інтеграції навчання з виробництвом, Вінницький національний технічний університет

Bisikalo Oleh V. — Dr.Sc. (Eng.), Professor, Dean of the Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia, email: obisikalo@gmail.com

Liahovchenko Nataliya V — the candidate of pedagogical Sciences, Associate Professor at the Department of integration of education with production, Vinnytsia National Technical University, Vinnytsia