

А.І. ВАВІЛЕНКОВА

ЛОГІКО-ЛІНГВІСТИЧНА МОДЕЛЬ ЯК ЗАСІБ ВІДОБРАЖЕННЯ СИНТАКСИЧНИХ ОСОБЛИВОСТЕЙ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Abstract. It is suggested using a logico-linguistic model as a method of display the syntactical features of text information. There are listed and justified the rules for their forming and also are shown an algorithm of creation of all components of such models in this paper.

Key words: logico-linguistic model, syntactical analysis, inference rule, rules, text information.

Анотація. Запропоновано використовувати логіко-лінгвістичні моделі як засіб відображення синтаксичних особливостей текстової інформації, перелічено та обґрунтовано правила їх формування, наведено алгоритм створення всіх складових таких моделей.

Ключові слова: логіко-лінгвістична модель, синтаксичний аналіз, правило виводу, правила, текстова інформація.

Аннотация. Предложено использовать логико-лингвистические модели как способ отображения синтаксических особенностей текстовой информации, перечислены и обоснованы правила их формирования, приведен алгоритм создания всех составных таких моделей.

Ключевые слова: логико-лингвистическая модель, синтаксический анализ, правило вывода, правила, текстовая информация.

1. Вступ

Однією із сфер застосування універсальної системи підтримки лінгвістичних досліджень є розвиток такої галузі комп'ютерних технологій, як інтелектуальна обробка текстової інформації, що потребує представлення текстових документів у формалізованій формі для можливості подальшого порівняння текстів та відшукування у них протиріч. З точки зору конструктивної семантики, формалізація передбачає перехід до оперування символами, при якому не потрібно додаткового аналізу речей об'єктивного світу і вся теорія розвивається у знаковій області [1].

Знання, які можуть бути представлені за допомогою логіки предикатів, являються фактами, що відображаються логічними формулами. Автоматичне перетворення речень, написаних на природній мові, в мову формальних систем типу логіки предикатів, називається розумінням природної мови. Логіці предикатів властивий високий рівень модульності знань, і разом з цим вона дозволяє отримати єдину систему представлення, в якій логічно роз'яснюються властивості знань як одного цілого. Першочерговим завданням логіки предикатів являється пояснення логічних основ природної мови. Оскільки будь-яка флективна мова надзвичайно складна, то об'єкти, якими оперує логіка предикатів, обмежені тими компонентами, елементи яких легко піддаються формалізації. Тому введемо базові поняття для систематизації цієї логіки, що дасть змогу вилучати знання з яких завгодно різних структур текстової інформації.

2. Постановка задачі

Логіко-лінгвістична модель (ЛЛМ) подає зв'язок між елементами пропозиційної логіки та синтаксичною структурою речень природної мови. ЛЛМ в загальному вигляді задається такими формулами:

$$(A \& B) \vee (A \rightarrow B) \vee (A \vee B) \vee (A \sim B) \vee A, \quad (1)$$

$$\forall_{S_i} S_i = M_{S_j} (S_k), k / i \in K_j, j = \overline{1, m}. \quad (2)$$

Кількість складових ЛЛМ дорівнює $(m + 1)$: формула (1) з урахуванням значень складних виразів A і B та аксіоми (2), кількість яких дорівнює кількості складних елементів формальної системи S_j^i .

Послідовно розберемо кожну з аксіом.

В аксіомі (1) A є складним логічним висловлюванням, яке описує просте речення, що входить до складу складного або описує будь-яке односкладове речення;

B – складне логічне висловлювання, яке описує просте речення, що входить до складу складного і є залежним.

3. Правила формування логіко-лінгвістичних моделей

Оскільки досі не існує єдиних правил формування ЛЛМ, не виведена загальна формула для ЛЛМ і не визначено чіткої схеми її створення, запропоновано для всіх типів речень флективних мов впровадити формулу (1) ЛЛМ, підтвердженням чому є ряд обґрунтувань.

1) Речення поділяються на прості та складні. Нехай просте речення в ЛЛМ характеризується висловлюванням A . Це ж висловлювання може бути частиною складного речення; тоді в реченні з декількома граматичними основами головна частина речення позначається висловлюванням A , а залежна – висловлюванням B .

2) В залежності від типу зв'язків між головною та залежною частинами речень розрізняють безсполучникові, складносурядні та складнопідрядні речення, для яких характерні певні знаки пунктуації та сполучні слова. Пунктуація базується на трьох основних принципах: семантичному, синтаксичному та інтонаційному. Вони пов'язані між собою і діють одночасно: зміст оформляється певною синтаксичною структурою, а інтонація передає і зміст, і змістовну структуру. Оскільки інтонація на письмі формально не відображається, знаки пунктуації обумовлені синтаксичною структурою речень і є обов'язковими, а зміст оформляється цією ж синтаксичною структурою, то в основу формування ЛЛМ покладено механізм здійснення синтаксичного аналізу.

3) Проаналізувавши граматичне оформлення речень, тобто способи вираження членів речення, підрядного та сурядного зв'язку, засоби вираження синтаксичних відношень між частинами складного речення, можна зрозуміти, що в реченні може бути чотири типи зв'язку: одночасність виконання ($A \& B$), наслідок ($A \rightarrow B$), альтернативність ($A \vee B$) та тотожність, еквівалентність ($A \sim B$). Якщо речення складне, то якийсь із типів зв'язку обов'язково буде характерним для вхідного тексту.

4) Якщо в реченні зустрічаються знаки пунктуації, сполучники або сполучні слова, то в залежності від знаку («», «;», «-», «:») в ЛЛМ буде присутня відповідна операція булевої алгебри і можна сформулювати правила створення складного висловлювання (1).

5) Знакам пунктуації «», «;» та єднальним сполучникам відповідає логічна операція кон'юнкції «&» та еквівалентності «~» відповідно. Це означає, що компоненти складного речення A і B рівноправні за змістом і можуть передавати одночасність, приєднання, сумісність дій або станів. За таких умов аксіома (1) ЛЛМ набуває вигляду ($A \& B$) або ($A \sim B$).

б) Для протиставних та розділових сполучників в ЛЛМ характерна логічна операція диз'юнкції « \vee ». Тобто складні висловлювання A і B являються частинами складносурядного речення, в якому протиставляються або зіставляються, взаємно виключаються або чергуються певні явища, йдеться про невпевненість, сумнів припущення. В такому разі ЛЛМ набуває вигляду $(A \vee B)$.

7) Знакам пунктуації «:», «-», сполучним словам та сполучникам, характерним для підрядного зв'язку, відповідає логічна операція імплікації « \rightarrow ». Це означає, що складні висловлювання A і B являються частинами складнопідрядного речення, вони не рівноправні за змістом, залежна частина речення може уточнювати час, місце, причину, спосіб, про який йдеться в головній частині складнопідрядного речення. За таких умов ЛЛМ матиме вигляд $(A \rightarrow B)$.

Таким чином, ЛЛМ (1) являється відображенням синтаксичної структури як завгодно складного речення природної мови, що дозволяє вилучити з текстової інформації знання, структуровані за певними правилами [2].

Для повного розуміння змісту речення необхідно прослідкувати зв'язки між простими елементами формальної системи (словами) та принципом їх об'єднання у складні елементи (словосполучення) [3]. Для цього в ЛЛМ передбачені аксіоми (3):

$$\forall_{S_i} S_i = M_{S_j} (S_k), k / i \in K_j, j = \overline{1, m} , \quad (3)$$

де S_i – головний елемент словосполучення;

K_j – множина елементів (слів), що входять до складного S_j^i , тобто до кожного j -го словосполучення;

$M_{S_j^i}$ – множина значень, які може приймати властивість залежного слова у словосполученні (допоміжних частин складного елемента S_j^i), наприклад, головне слово може позначати об'єкт, кількість, якість, що залежить від типу словосполучення ;

k – частина складного елемента S_j^i .

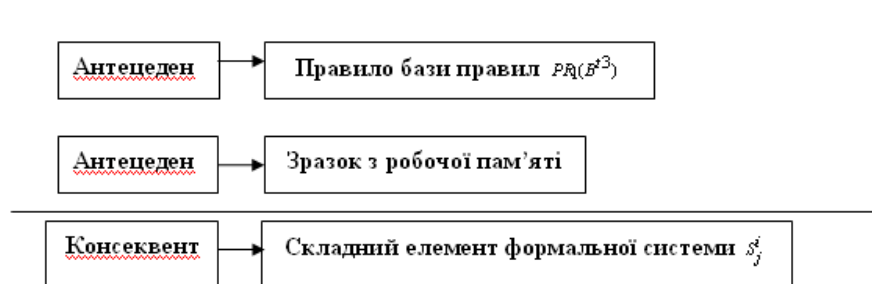


Рис. 1. Загальна схема дії механізму логічного виводу бази знань B^{t3} для визначення структури складних елементів ФС

Формування аксіом (3) відбувається шляхом звернення до бази знань B^{t3} . В робочу пам'ять послідовно заносяться прості елементи формальної системи $S_i, \overline{i = 1, n}$.

Згідно з правилом modus ponens буде здійснено логічний вивід. Схема реалізації дії механізму логічного виводу зображена на рис. 1.

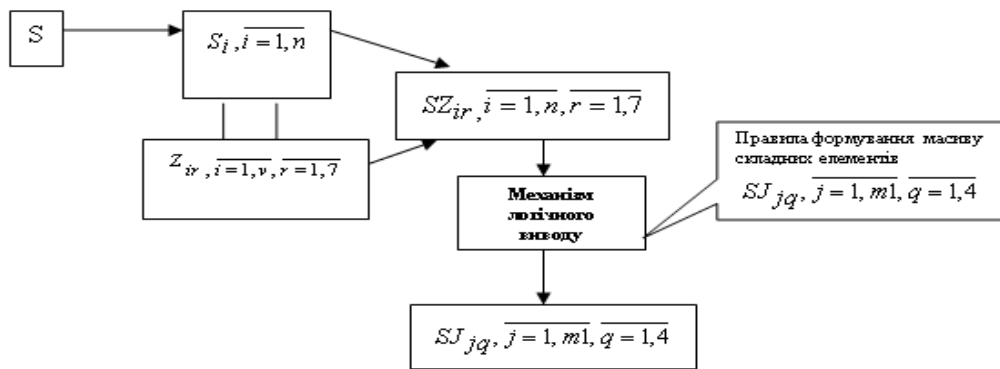


Рис. 2. Загальна схема формування аксіом

Результатом роботи механізму логічного виводу буде двовимірний масив SJ , кількість рядочків якого рівна кількості словосполучень у реченні, тобто $\overline{j=1, m1}$, а значення елементів масиву представляють собою словосполучення, його головний елемент, властивість залежного слова у словосполученні та безпосередньо залежну частину складного елемента формальної системи:

$$S \Rightarrow S = \begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_i \\ \dots \\ S_n \end{pmatrix} \Rightarrow SZ = \begin{pmatrix} S_1 & cm_{11} & g_{12} & n_{13} & k2_{14} & t_{15} & h_{16} & l_{17} \\ S_2 & cm_{21} & g_{22} & n_{23} & k2_{24} & t_{25} & h_{26} & l_{27} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_i & cm_{i1} & g_{i2} & n_{i3} & k2_{i4} & t_{i5} & h_{i6} & l_{i7} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_n & cm_{n1} & g_{n2} & n_{n3} & k2_{n4} & t_{n5} & h_{n6} & l_{n7} \end{pmatrix} \Rightarrow SJ = \begin{pmatrix} S_1^i & S_i & M_{S_1^i} & S_k \\ S_2^i & S_i & M_{S_2^i} & S_k \\ \dots & \dots & \dots & \dots \\ S_j^i & S_i & M_{S_j^i} & S_k \\ \dots & \dots & \dots & \dots \\ S_{m1}^i & S_i & M_{S_{m1}^i} & S_k \end{pmatrix}.$$

Кількість рядочків двовимірного масиву SJ дорівнює кількості аксіом (3) в ЛЛМ. Загальний алгоритм побудови логіко-лінгвістичної моделі речення, тобто схема формування аксіом (1) – (3), показана на рис. 2.

4. Висновки

Аналіз синтаксичних особливостей текстової інформації дозволив прослідкувати зміну форми ЛЛМ в залежності від знаків пунктуації та сполучних слів, які зустрічаються в реченні. Ця залежність відображена в аксіомі (1). Модель (2) являється своєрідним токеном, призначеним для вилучення з тексту якомога більшої кількості інформації. Подальше дослідження і аналіз загальної форми ЛЛМ, а також синтаксичних особливостей різних мов дасть змогу повністю формалізувати процес вилучення знань з текстової інформації.

СПИСОК ЛІТЕРАТУРИ

1. Корпусна лінгвістика / [В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна та ін.]. – К.: Довіра, 2005. – 471 с.
2. Фостер Дж. Автоматический синтаксический анализ / Фостер Дж.; пер. с англ. В.В. Мартынюка; под ред. Э.З. Любимского. – М.: Мир, 1975. – 70 с.
3. Мельников Г.П. Системология и языковой аспект кибернетики / Мельников Г.П.; под ред. Ю.Г. Косарева. – М.: Советское радио, 1978. – 368 с.

Стаття надійшла до редакції 12.08.2009