

АВТОМАТИЗАЦІЯ ПРОЦЕСУ ПОБУДОВИ ЛОГІКО-ЛІНГВІСТИЧНОЇ МОДЕЛІ ПРОСТОГО РЕЧЕННЯ

Національний авіаційний університет

Сформульовано основні підходи для вирішення проблеми формалізації текстової інформації та методів її представлення з метою виявлення логічних протиріч, ознайомлено з основними принципами формалізації інформації, запропоновано алгоритм автоматизованого формування логіко-лінгвістичних моделей простих речень, продемонстровано роботу системи на прикладі

Вступ

Принципово нові можливості комп'ютерів та мереж зумовлюють необхідність якісно нових засобів опрацювання інформації, які розвиватимуться, насамперед, у напрямку інтелектуалізації.

Суспільна потреба у розробці ефективних лінгвістичних технологій, на яких базуватимуться технології оперування знаннями, вимагає створення універсальної системи підтримки лінгвістичних досліджень та розробок [1]. Зокрема, з метою представлення текстової інформації у формалізованій формі для можливості подальшого порівняння текстів, необхідне створення системи відшукування протиріч.

Один із підходів створення такої системи базується на побудові логіколінгвістичної моделі тексту, який підлягає перевірці на логічне протиріччя відносно других текстів. Основною проблемою на цьому шляху є автоматизація процесу побудови такої моделі, що дозволить надалі застосовувати до логіколінгвістичних моделей алгоритми доведення логічних протиріч як для формальних моделей, представлених предикатами першого порядку.

Постановка задачі

Логіко-лінгвістична модель представляє собою вираз або вирази у вигляді термінів природної мови та лінгвістичних змінних (які можуть бути не лише числами, але й словами і словосполученнями природної або штучної мови). Основою для логіколінгвістичних моделей служить логіка предикатів, де предикат – це функція P , яка приймає значення 0 або 1, і аргументи якої

пробігають значення із довільної множини $M: P(x_1, x_2, \dots, x_n)$ [2].

Логіко-лінгвістична модель здатна відображати зміст речень тексту і є засобом представлення його у формалізованій формі. Формалізація – це представлення предметної області (розмірковувань, доведень, процедур класифікації, пошуку інформації) у вигляді формальної системи або обчислень. Формальні системи – це системи операцій над об'єктами, які розуміються як послідовності символів, самі операції також представляються як операції над символами [3]. Формалізація передбачає посилення ролі формальної логіки, оскільки у випадку формальних теорій вже не можна задовольнятися інтуїтивними твердженнями, так як та чи інша аргументація узгоджена з логічними правилами, набутими завдяки здатності до правильного мислення. Існує твердження, що повністю можуть бути формалізовані лише елементарні теорії з простою логічною структурою і невеликим запасом понять.

Проте запропонована методика перетворення речення в логіколінгвістичну модель спростовує це твердження і показує, що формалізувати можна будь-яку текстову інформацію, представлену у вигляді різних типів речень.

Згідно з методикою перетворення речення в логіколінгвістичну модель (для неї суб'єктом управління буде підмет речення, об'єктом управління – додаток, над яким виконується дія; сама дія, що виконується суб'єктом управління, – присудок речення) для будь-якого типу речення з однією граматичною основою, логіко-лінгвістична модель має вигляд:

$$P(x, \{U_i(x)\}, \{Y_j(\{D_j(x)\}, \{Q_k(D_j(x)))\}), \{O_n(x)\}), \quad (1)$$

де i, j, k, n – натуральні числа;

P – предикат, що позначає присудок;

x – підмет;

$\{U_i(x)\}$ – множина означень, що характеризують підмет;

$\{D_j(x)\}$ – множина додатків;

$\{Q_k(D_j(x))\}$ – множина означень, що характеризують додаток речення;

$\{Y_j(\{D_j(x)\}, \{Q_k(D_j(x)))\})\}$ – множина лексично пов'язаних додатків і означень;

$\{O_n(x)\}$ – множина обставин.

Автоматизація процесу побудови логіко-лінгвістичної моделі простого речення

З урахуванням граматичної та лексичної семантики, граматична підсистема має бути реалізована у такий спосіб, що забезпечував би виокремлення у спостережуваних компонентів граматичних станів української лексичної системи, тобто виділенням певних лексико-граматичних класів слів або підмножин слів за окремими

граматичними значеннями, а також за іншими граматичними параметрами та їх комплексами. Особлива увага в процесі автоматизації побудови логіколінгвістичної моделі приділяється таким граматичним параметрам, як відмінок, рід, число, а також синтаксичній структурі речення, тобто тому, яким членом речення виступає те чи інше слово, що розглядається. Кінцевою метою процесу автоматизації є створення моделі (1).

Алгоритм процесу автоматизації побудови логіко-лінгвістичної моделі простого речення складається з декількох етапів:

В програму водиться просте речення, слова в якому розділені пробілами, а кінцем речення вважається крапка. Будь-які інші знаки пунктуації в реченні відсутні. Програма розбиває речення на масив, кожним елементом якого є слово речення (рис. 1)

```

Введіть будь-яке просте речення:
вони авансували нам десять абажурів.
mas_slov[0]= вони
mas_slov[1]= авансували
mas_slov[2]= нам
mas_slov[3]= десять
mas_slov[4]= абажурів

```

Рис. 1. Перетворення речення в масив

2. Відбувається пошук кожного із елементів масиву в таблицях бази даних, якщо слово знайдене, із таблиць беруться дані про

граматичні параметри даного слова, згідно з яким визначається синтаксична його роль у реченні (рис. 2 – 6)

```

Стрічка, повернена з бази даних:
Вони, займенник, Н.в., -, -, особовий, суб'єкт
суб'єкт вони,
(вони, )

```

Рис. 2. Виведення граматичних параметрів першого елементу масиву

Стрічка, повернена з бази даних: авансували, дія
дія
авансували_
авансували_(вони,)

Рис. 3. Виведення граматичних параметрів другого елементу масиву

Стрічка, повернена з бази даних:
нам, займенник, Д.в.,-,-,-,особовий, об'єкт
об'єкт нам,
авансували_(вони, нам)

Рис. 4. Виведення граматичних параметрів третього елементу масиву

Стрічка, повернена з бази даних:
десять, числівник, Н.в.,-,-,-,кількісний, умова
умова десять,
авансували_(вони, нам, десять)

Рис. 5. Виведення граматичних параметрів четвертого елементу масиву

$$P(x, \{U_i(x)\}, \{Y_j(\{D_j(x)\}, \{Q_k(D_j(x))\})\}, \{O_n(x)\})$$

де P – предикат, що позначає присудок – «авансували»;

x – підмет – «вони»;

$\{U_i(x)\}$ – множина означень, що характеризують підмет – «»;

$\{D_j(x)\}$ – множина додатків – «нам», «абажурів»;

$\{Q_k(D_j(x))\}$ – множина означень, що характеризують додаток речення – «»;

$\{Y_j(\{D_j(x)\}, \{Q_k(D_j(x))\})\}$ – множина лексично пов'язаних додатків і означень – «»;

$\{O_n(x)\}$ – множина обставин – «декілька».

База даних, що використовується в процесі автоматизації, представляє собою інструментальний комплекс, що поєднує довідкову функцію, побудовану на основі формалізації лінгвістичних критеріїв, згідно з якими вся множина слів мови розбивається на підмножини [1]. Такими підмножинами є

Стрічка, повернена з бази даних:
абажурів, іменник, Р.в.,-,-,-, об'єкт
об'єкт абажурів,
авансували_(вони, нам, абажурів, десять)

Рис. 6. Виведення граматичних параметрів п'ятого елементу масиву

3. В результаті отримано логіколінгвістичну модель даного простого речення (рис. 7), що побудована згідно прототипу (1).

авансували_(вони, нам, абажурів, десять)

Рис. 7. Отримана логіко-лінгвістична модель

В представленій методиці первинний текст перетворюється в структуру даних, яка відображає синтаксичну структуру вхідної послідовності. Для побудови логіколінгвістичної моделі згідно з запропонованою методикою необхідно правильно визначити всі члени речення і зв'язки між ними, частини мови, їх граматичні форми та морфологічні ознаки.

таблиці основних та службових частин мови, а також таблиці закінчень та суфіксів, що дозволяють визначити, до якої частини мови належить те чи інше слово.

В основі побудованої логіколінгвістичної моделі лежить логіка висловлювань – в якості основних понять взятий клас об'єктів, їх синтаксичні відношення та операції над цими об'єктами. Ці основні поняття розглядають як вихідні [4]. З іншої сторони, вони вибрані не довільним чином, а так, що відповідати змісту, який повинна описувати математична теорія.

В основі даної методики перетворення речення в логіколінгвістичну модель лежить синтаксичний зв'язок між словами речення (виявити його дозволяє синтаксичний парсинг речення).

Подальша робота в області автоматизації процесу побудови логіколінгвістичної моделі речення буде направлена на дослідження структури складних речень та застосуванні знаків пунктуації для розбору речення, а також розбиття отриманих логіколінгвістичних моделей на більш прості лексеми.

Висновки

З точки зору конструктивної семантики (науки, що займається вивченням і формуванням законів синтезу знакових систем, які забезпечують процес відображення об'єктивного світу в його моделях) [5] формалізація передбачає перехід до оперування символами, при якому не потрібно додаткового аналізу речей об'єктивного світу і вся теорія розвивається в знаковій області, без звернення до досвіду, емпірії.

Формалізованою формою представлення текстових документів можуть бути логіколінгвістичні моделі, які здатні відображати зміст речень тексту[6].

Методика перетворення речення в логіколінгвістичну модель базується на синтаксичному парсингу речення. Синтаксичний парсинг – це процес аналізу вхідної послідовності символів з метою розробки граматичної структури, зазвичай у відповідності із заданою формальною граматикою. При парсингу первинний текст перетворюється в структуру даних (наприклад, дерево), яке відображає синтаксичну структуру вхідної послідовності і підходить для подальшої обробки. В представленій методиці первинний текст перетворюється в структуру даних, яка відображає синтаксичну структуру вхідної послідовності. Для побудови логіколінгвістичної моделі згідно з запропонованою методикою необхідно правильно визначити всі члени речення і зв'язки між ними, частини мови, їх граматичні форми та морфологічні ознаки.

Алгоритм автоматизації процесу побудови логіколінгвістичної моделі простого речення, що розглядається у даній статті, є першим кроком до автоматизованого формування логіколінгвістичних моделей різних типів речень. Покладена в основу процесу автоматизації ідея використання синтаксичного парсингу речення, дає можливість стандартизувати логіколінгвістичну модель під всі типи речень.

Автоматизація процесу перетворення речення в логіколінгвістичну модель, яка здатна відображати зміст речень дає змогу піддавати тексти логічній перевірці, використовуючи при цьому алгоритми доведення логічних протиріч, що застосовуються в логіці предикатів. Основу для побудови такої моделі зв'язку пізнання та структури (в даному випадку структури речення) дає теоретична когнітологія, яка базується на логіці та аналізі побудови й функціонування природної мови.

Список літератури

1. Широков В.А., Бугаков О.В., Грязнухіна Т.О. та ін. Корпусна лінгвістика. – К.: «Довіра», 2005. – 471 с.
2. Поспелов Д.А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 232 с.
3. Кузнецов О.П., Адельсон Вельский Г.М. Дискретная математика для инженера. – М.: Энергия, 1980. – 344 с.
4. Джордж Ф., Люгер Искусственный интеллект: стратегии и методы решения сложных проблем, 4-тое издание. – М.: ООО «Вильямс», 2005. – 864 с.
5. Попов Э.В. Общение с ЭВМ на естественном языке. – М.: Наука. Главная редакция физико-математической литературы, 1982. – 360 с.
6. Джарратано Д. Экспертные системы: принципы разработки и программирование, 4-е издание.: Пер. с англ. – М.: ООО «Вильямс», 2007. – 1152 с.