

ЗАСТОСУВАННЯ БАЗ ЗНАНЬ ДО НЕСТРУКТУРОВАНОЇ ТЕКСТОВОЇ ІНФОРМАЦІЇ

А. І. Катаєва

На сьогоднішній день, велика кількість неструктурованої текстової інформації зосереджено в мережі Інтернет. Більшість можна структурувати, застосовуючи засоби побудови реляційних баз даних (БД). Та що робити, з інформацією, яку структурувати неможливо або доволі складно? На допомогу приходять бази знань (БЗ) – сучасний інструмент для утворення сукупності знань з певної предметної області, формально представлених так, щоб на їх основі можна було здійснювати міркування. Зазвичай база знань є сукупністю правил виводу [1]. Тому важливим моментом постає питання утворення та роботи бази знань із неструктурованою текстовою інформацією.

Розглянемо існуючі моделі побудови БЗ, які можуть вирішити поставлену задачу:

- продукційна модель подання знань – модель, заснована на правилах, дозволяє представити знання у вигляді пропозицій типу – «Якщо (умова), то (дія)»;
- семантична модель знань – це модель, основою якої є орієнтований граф, який відображає відношення (зв'язки типу: «це» («АКО — A-kind-of» «is-a»), «має частиною» («has part»), «належить», «любить») між поняттями (абстрактними або конкретними об'єктами) [2];
- фреймова модель даних – її розуміють як мережну моделі уявлення знань, коли фрагмент мережі представляється фреймом з відповідними слотами і значеннями [3];
- логічна модель знань – це модель, у якій вся інформація, необхідна для вирішення прикладних завдань, розглядається як сукупність фактів і тверджень, що представляються як формули в деякій логіці [3].

Отже, перед нами стоїть задача проаналізувати вхідний текст та на його основі утворити БЗ. Процеси аналізу та синтезу текстової інформації традиційно базуються на апараті формальних граматик. Фактично, йдеться про набір правил, що дозволяють описувати деякі граматичні закономірності природних та штучних мов [4]. Для того щоб проаналізувати цілий текст необхідно розділити його на частини.

Проаналізувавши існуючі моделі формування БЗ, зупинимось на семантичній. Так як в основі семантичної структури природної мови лежить модель оточуючого нас світу, то можливість точного опису природної мови прямо залежить від можливості побудови релевантних моделей реальної дійсності. Для формалізації семантики природною мовою достатньо мовної моделі світу, а ця модель міститься в словниковому запасі природної мови [5]. Тому в основі БЗ буде певний словник. Одним з ключових понять для визначення структури словникових системи є поняття лексикографічного параметру – деякого «кванту» лінгвістичної інформації, що може мати самостійний інтерес для користувача, але, як правило, виступає в комбінації з іншими параметрами – «квантами» – і знаходить своє специфічне вираження у словниках; іншими словами – це особливе словникове відображення окремих структурних рис мови [6].

Для формування словника, який буде основою БЗ, наведемо формальну модель онтології – впорядковану трійку виду:

$$O = \langle X, \mathfrak{R}, \Phi \rangle, \quad (1)$$

де X – скінченна множина концептів (понять, термінів) предметної області, яку представляє онтологія O ; \mathfrak{R} – скінченна множина відношень між концептами (поняттями, термінами) предметної області; Φ – скінченна множина функцій інтерпретації, заданих на концептах та/чи відношеннях онтології O .

Введемо обмеження у загальне визначення онтології. Як наслідок, отримаємо таксономію:

$$\begin{aligned} O &= T^0 = \langle X, \mathfrak{R}, \{\} \rangle, \\ X &= \bigcup_{i=1}^N \alpha_i, \\ \mathfrak{R} &= \{is_a\}. \end{aligned} \quad (2)$$

Під таксономічною структурою розуміється ієрархічна система понять, пов'язаних між собою співвідношенням «*is_a*» – бути елементом класу. Відношення *is_a* має завідомо фіксовану семантику і дозволяє організувати структуру понять у вигляді дерева. Отже, семантичний словник S можна задати у вигляді об'єднання множин $V_i(x)$ (функції значення слова) для всіх $x \in W(\varnothing)$ (де $W(\varnothing)$ – кінцева множина всіх слів):

$$S = \bigcup_{j=1}^N \bigcup_{i=1}^m V_i(x_j). \quad (3)$$

Записи семантичного словника S_r будуть мати наступний вигляд:

$$S_r = V_i(x) = x \cup f_m(x_1^1, \dots, x_n^1) \cup \alpha \cup k \cup M(u), \quad (4)$$

де x – заглавне слово; $f_m(x_1^1, \dots, x_n^1)$ – суперпозиція семантичних функцій, що відповідають значенням слова x ; α – посилання на ієрархію типів; k – посилання на концептуальний граф чи порожню множину; $M(u)$ – множина посилань на словник стійких словосполучень чи порожню множину.

Для побудови моделі БЗ на основі семантичного словника S_r з записами (4) потрібно визначити складне відношення \mathfrak{R} для всіх елементів усіх класів предметної області, тому на першому етапі пропонується побудувати квадратну матрицю Q із статистично визначеними вагами (пропорційними кількості) зв'язків між словоформами обраного тексту. Для реалізації моделі БЗ будемо використовувати нечіткі відношення.

Після ряду перетворень, на основі синтаксичного розбору кожного речення вхідного тексту вносимо зв'язки між значимими словоформами до БЗ Q .

Отже, запропоновано застосувати семантичну модель бази знань (4) з урахуванням квадратної матриці, в якій по рядках вказане головне слово, а по стовпцях – залежне. Елементами матриці є сили зв'язків між цими словоформами, тобто чисельне вираження частоти зв'язку між головним та залежним словом. У побудові матриці беруть участь лише значимі частини мови обраного тексту, а для формалізації відповідного нечіткого відношення бази знань будується функція належності (5).

$$\overline{\mu_Q} = \frac{1}{m} \sum_{j=1}^m \mu_{Q_j} = 0,5, \quad (5)$$

де m – кількість ненульових елементів матриці Q ; μ_Q – сигмоїдальна функція.

Приведемо приклад проаналізованого речення. Нехай маємо речення: *(Does) (the) people (from) (the) ship mayflower have come friends (with) (the) indian?* На рисунку 1 представлено аналітичний граф.

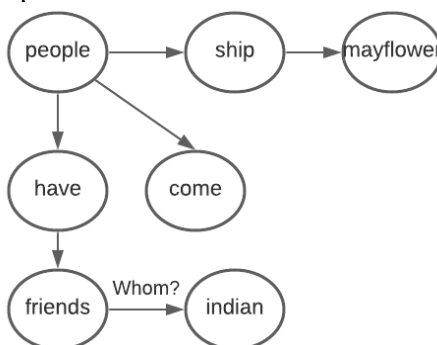


Рис. 1. Дерево графа аналізу речення

Як бачимо, структуру графа утворюють лише значимі словоформи, які і формують базу знань.

Література

1. Марченко А. В. ОБДЗ. Тема 11 – Загальна характеристика баз знань. URL: <https://ocw.sumdu.edu.ua/content/811#node86174>
2. Марченко А. В. ОБДЗ. Тема 12 – Моделі знань. URL: <https://ocw.sumdu.edu.ua/content/811#node86180>
3. Моделі подання знань. Мережеві моделі: фрейми, семантичні мережі. URL: http://baklaniv.at.ua/PSAI/lekcija_9-10_2016.2.pdf
4. Словарь по кибернетике / [ред. академик В. М. Глушков]. К.: Главная редакция украинской советской энциклопедии, 1979. 624 с.
5. Карпіловська Є. А. Термінологічний підфонд у складі морфемно-словотвірному фонду української мови / принципи формування та можливості використання. *Україномовне програмне забезпечення. Матеріали 4-ої та 5-ої Міжнарод. науково-практ. конф. «УкрСофт»*. Львів, 1995. С. 161–162.
6. Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка М.: Наука, 1981. 467 с.

УДК 004.942+656.052.1

ЛОГІСТИЧНА РІВНОВАГА НЕША І МІСЬКИЙ ТРАФІК

П. К. Ніколюк

Однією із актуальних проблем сьогодення є проблема заторів на вулицях великих міст. В цьому напрямку ставиться задача прокладання оптимального по часу проїзду по маршруту всіх автомобілів, що приймають участь у міському трафіку. Іншими словами, міський трафік знаходиться у нерівноважному стані і тому його потрібно перевести у стан рівноваги Неша. В такій ситуації досягається оптимальний розподіл транспортних засобів (ТЗ) по транспортних артеріях міста. Крім того, досягається стан мінімізації сумарного часу проїзду всіх ТЗ, що приймають участь у трафіку [1].

Стратегічна мета заключається в побудові оптимальних маршрутів для кожного ТЗ та синхронізації потоків ТЗ. Ставиться задача провести кожен ТЗ по місту по оптимальному маршруту з урахуванням можливої зміни такого маршруту. Отже, система регулювання трафіка всієї сукупності автомобілів на трасах міста повинна прокладати динамічний – в режимі реального часу – і оптимальний маршрут кожному ТЗ, що замовляє прокладання маршруту центру керування трафіком (ЦКТ).

ЦКТ, використовуючи спеціалізовану комп'ютерну програму, передає інструкції кожному водієві у вигляді голосових команд щодо маршруту руху до заявленого водієм пункту призначення як при звичайній GPS-навігації. Особливість полягає у тому, що програма аналізує динамічну ситуацію на кожному перехресті і по всьому місту і відповідно прокладає маршрут з урахуванням ситуації на кожен конкретний момент часу. При цьому використовується комп'ютерна програма, що реалізує спеціальний алгоритм щодо знаходження оптимального шляху. Остаточною метою даного дослідження є синхронізація транспортних потоків, оптимальне використання транспортних артерій всього міста, запобігання утворенню заторів а також супровід кожного ТЗ до місця призначення з таким розрахунком, щоб затрачений на поїздку час був мінімальним.

Сформулюємо стадії вирішення проблеми оптимального режиму керування міським трафіком. Першим кроком у цьому відношенні є моделювання транспортної мережі міста за допомогою зваженого орієнтованого планарного мультиграфа. Принциповим моментом є те, що у такому графі кожній дузі співставляється у відповідність динамічна величина – вага, величина якої відповідає реальній завантаженості автомобілями міської дорожньої смуги руху. Мовою теорії графів згаданий граф представляє собою бієктивне (максимально об'єктивне) відображення транспортної мережі міста [2].