

ЛОГІКО-ЛІНГВІСТИЧНІ МОДЕЛІ РЕЧЕНЬ ЯК ЗАСІБ ПОРІВНЯННЯ ТЕКСТОВИХ ДОКУМЕНТІВ ЗА ЗМІСТОМ

Анотація. Проаналізовано основні методи інтелектуальної обробки текстової інформації. Застосовано метод резолюцій для пошуку логічних збігів у реченнях природної мови, представлених у вигляді логіко-лінгвістичних моделей. Досліджено основні етапи алгоритму методу шинглів для порівняння текстових документів за змістом. Запропоновано метод порівняння логіко-лінгвістичних моделей текстової інформації за змістом.

Ключові слова: інтелектуальна обробка, логіко-лінгвістична модель, метод резолюцій, метод шинглів, метод порівняння логіко-лінгвістичних моделей, текстова інформація.

Аннотация. Произведен анализ методов интеллектуальной обработки текстовой информации. Применен метод резолюций для поиска логических совпадений в предложениях естественного языка. Исследованы основные этапы алгоритма метода шинглов для сравнения текстовых документов по смыслу. Предложен метод сравнения логико-лингвистических моделей текстовой информации по смыслу.

Ключевые слова: интеллектуальная обработка, логико-лингвистическая модель, метод резолюций, метод шинглов, метод сравнения логико-лингвистических моделей, текстовая информация.

Abstract. The main methods of intelligent processing of text information were analyzed. The method of resolutions for searching logical coincidences in the sentences of natural language, represented in logical-linguistic models, was applied. The basic steps of the shingles method algorithm for comparing text documents according to the content were researched. A comparing method for the logical-linguistic models of textual information according to the content was proposed.

Keywords: intelligent processing, logical-linguistic model, resolution method, shingles method, method of comparison of logical-linguistic models, the text information.

1. Вступ

Нові можливості, які несе з собою інформаційна революція, створюють виклик традиційним системам генерування, розповсюдження та передачі знань, тобто системам науки й освіти. Потужні бази даних і знань відіграють роль гігантських «сховищ» для нескінченних фактів і базових даних у всіх сферах людської діяльності, а глобальні комп'ютерні мережі стають потужними інструментами для високошвидкісного доступу до цієї інформації з будь-якого куточку світу.

У зв'язку з цим істотно зростає роль методологічних, системних, міждисциплінарних знань людини, необхідних для раціонального й осмисленого оперування різноманітними знаннями і даними з метою вирішення нових, нестандартних проблем. У цій новій парадигмі найголовніше місце відводиться аналітичним здібностям ученого чи педагога, тобто його спроможності шукати і знаходити необхідну інформацію, точно формулювати проблеми і гіпотези, вбачати в сукупностях даних певні закономірності, знаходити розв'язок складних міждисциплінарних задач [1].

Застосування інформаційних технологій, зокрема, систем автоматизованого вилучення знань, систем порівняння електронних документів за змістом та повнотекстових пошукових серверів, дасть змогу полегшити роботу людини, підвищуючи відсоток релевантності отриманої інформації.

2. Постановка задачі

Сьогодні з величезної кількості інформації, яка надається людині, дуже важко виокремити головне, знайти необхідні дані для прийняття рішень. Це створює головну проблему су-

часних комунікацій – вилучення дійсно цінних даних з інформаційних потоків, тобто отримання знань з інформації.

Розроблені на основі статистичного та лінгвістичного аналізу, а також методах штучного інтелекту, технології Text Mining призначені для проведення змістового аналізу, забезпечення навігації та пошуку в неструктурованих текстах. Можливості сучасних систем Text Mining можуть застосовуватися при управлінні знаннями для виявлення шаблонів у тексті, автоматичного анотування інформації або її розподілу за певними профілями.

Дослідники пропонують декілька підходів, на які спирається технологія Text Mining, вони розпадаються на дві категорії. В основі підходів, що не спираються на знання, лежить відмова від додавання нових правил для кожної нової прикладної області знань або мови. Ще один підхід, який спирається на знання, передбачає, що якщо вдається зрозуміти значення тексту, то скоротити його стає набагато простіше, відповідно, отримана в результаті анотація буде більш якісною. Такий підхід передбачає використання бази знань значного об'єму, що складається з правил, які включаються, підтримуються і потім адаптуються до нових додатків та мов. Ці два підходи не виключають один одного, тому можна використовувати гібридні підходи.

До основних елементів Text Mining відносяться [2]:

1) класифікація – використовує статистичні кореляції для побудови правил розміщення документів у наперед заданій категорії; задача класифікації – це задача розпізнавання, коли система відносить новий об'єкт до тієї чи іншої категорії;

2) кластеризація – це розбиття множини документів на кластери (групи документів зі спільними ознаками), які представляють собою підмножини, смислові параметри яких заздалегідь невідомі; числові методи кластеризації базуються на визначенні кластера як множини документів:

а) значення семантичної близькості між будь-якими двома елементами якого не менше визначеного «порогу»;

б) значення близькості між будь-яким документом множини і центроїдом (вектором, який обчислюється як середнє арифметичне векторів усіх документів кластера) цієї множини не менше визначеного «порогу»;

3) автоматичне реферування представляє собою створення коротких викладів матеріалів, анотацій, дайджестів, тобто вилучення найбільш важливих відомостей з одного або декількох документів і генерація на їх основі лаконічних та інформаційно-ємних звітів; на сьогодні існує два основних напрямки автореферування: квазіреферування (засноване на екстрагуванні фрагментів документів, тобто виділенні найбільш інформативних фраз і формування з них квазірефератів) і коротке викладення змісту первинних документів (дайджести).

Процес реферування розпадається на три етапи: аналіз початкового тексту, визначення його характерних фрагментів та формування відповідного висновку. Більшість сучасних робіт концентруються навколо розробленої технології реферування одного документа.

У методі створення дайджестів робиться акцент на виділенні характерних фрагментів, речень. Для цього методом співставлення фразових шаблонів виділяються блоки найбільшої лексичної та статистичної релевантності. Створення нового документа у даному випадку – це з'єднання вибраних фрагментів.

У більшості методів застосовується модель лінійних вагових коефіцієнтів. Основу аналітичного етапу в цій моделі складає процедура призначення вагових коефіцієнтів для кожного блоку тесту у відповідності до таких характеристик, як розташування цього блоку в оригіналі, частота появи у тексті, частота використання у ключових реченнях, а також показники статистичної значущості. Сума індивідуальних ваг визначається після додаткової модифікації у відповідності з спеціальними параметрами налаштування, пов'язаними з

кожною вагою, дає загальну вагу всього блоку тесту. Крім того, для призначення вагових коефіцієнтів у цій моделі враховується показник статистичної важливості, який обчислюється на основі даних, отриманих у результаті аналізу автоматичної індексації, при якій дослідники виявляють та оцінюють цілий ряд метрик, що визначають вагові коефіцієнти терміна. Ці метрики дозволяють виділити документ із числа інших у певному наборі документів [2].

Як видно, жоден із цих методів не здатен вилучати із текстової інформації знання і базується в основному на статистичних методах досліджень та завчасно створених шаблонах. Таким чином, якщо предметною областю досліджень буде вся природна мова, а не текст певної тематики, задовільне рішення задачі – вилучення знань з електронних документів – не буде знайдено. Неможливим буде і порівняння текстових документів за змістом, спроби чого проводяться шляхом застосування різних методів, основні з яких – метод резолюцій та метод шинглів.

3. Метод резолюцій як метод пошуку логічних збігів

У методі резолюцій множина речень розглядається як складений предикат, що містить декілька предикатів, з'єднаних логічними функціями і кванторами існування та узагальнення. Оскільки однакові за змістом предикати можуть мати різний вигляд, то речення перетворюються в клаузальну форму – різновид кон'юнктивної нормальної форми, з якої видалені квантори існування, узагальнення, символи імплікації, рівнозначності і т.д.

У роботах таких радянських та зарубіжних вчених, як Е.В. Попов, Д.А. Поспелов, В.М. Вагін, Д. Джарратано, С. Осуга та Х. Уено метод резолюцій розглядають, інтерпретуючи висловлювання S як сукупність речень C_1, C_2, \dots, C_n [3–5]. Вважається, що в S є такі речення, що якщо в одне з них входить деякий літерал, то в інше входить його заперечення (контрарна пара) (1):

$$\begin{aligned} C_1 &= (L_1, L_2, L_4), \\ C_2 &= (L_3, \neg L_4). \end{aligned} \quad (1)$$

Їх викреслення формує новий диз'юнкт із складових частин диз'юнктів (2). Новий сформований диз'юнкт C_3 називається резольвентою диз'юнктів C_1 і C_2 :

$$C_3 = C_1 \cdot C_2 = (L_1, L_2, L_4) \cdot (L_3, \neg L_4) \rightarrow (L_1, L_2, L_3). \quad (2)$$

Тобто, резольвента C_3 , отримана з двох диз'юнктів C_1 і C_2 , є логічним наслідком цих двох диз'юнктів. Тоді, якщо інтерпретувати метод резолюцій на два довільних речення природної мови, можна стверджувати, що резольвентою двох протилежних за змістом речень буде пуста множина і навпаки.

Отже, метод резолюцій дає можливість знайти певний елемент із заданих висловлювань, в яких вказані певні початкові умови.

Якщо для логіки висловлювань знаходження контрарних пар не викликає труднощів, то для логіки предикатів це не так. Принцип резолюцій має важливу властивість – повноту, яка встановлюється такою теоремою: множина диз'юнктів S не виконується тоді і тільки тоді, коли існує вивід із S порожнього диз'юнкта.

У силу нерозв'язності логіки предикатів першого порядку для істинної множини диз'юнктів S процедура, що базується на принципі резолюцій, буде працювати нескінченно довго.

Спроба демонстрації методу резолюцій для логіки предикатів першого порядку не на прикладі продукційної моделі представлення знань, що відображає певну ситуацію, дає такі результати.

Нехай є два простих речення, кожному з них відповідає логіко-лінгвістична модель.

Перше речення: «Літак пролітає над нашим будинком», логіко-лінгвістична модель для нього представлена у вигляді (3)

$$P(x_1, x_2 \{c_{12}\})$$

Пролітає (літак, будинком {нашим})

(3)

Друге речення: «Літак пролітає над великим містом», логіко-лінгвістична модель для такого речення буде мати вигляд (4)

$$P(x_1, x_2 \{c'_{12}\})$$

Пролітає (літак, містом {великим})

(4)

Множина атомів A буде містити літерали з однаковими предикатами [6]:

$$A = P(x_1, x_2 \{c_{12}\}), P(x_1, x_2 \{c'_{12}\}).$$

Згідно з алгоритмом уніфікації, шукаємо підстановку $Q = \{x_2 / x'_2, c_{12} / c'_{12}\}$. Тоді резольвента для цих двох речень буде мати вигляд

$$P(x_1, x_2 \{c_{12}\}), P(x_1, x_2 \{c_{12}\}).$$

Те, що резольвентою є не пуста множина, доводить, що речення не протилежні за змістом, а однакові. Проте насправді речення не однакові за змістом, у них лише однакові предикат та предикатна змінна – суб'єкт, а предикатна змінна – об'єкт та предикатна константа, яка його характеризує, різні, що безпосередньо впливає на зміст речення.

Демонстрація конкретного випадку порівняння двох речень (власне, призначення методу), а також дослідження особливостей здійснення алгоритму, дозволяють виявити ряд недоліків методу резолюцій:

1. Вибір літералів для множини атомів A відбувається тільки з літералів з однаковими предикатами, проте при цьому не враховуються синоніми, рівнозначність активної та пасивної форм дієслів, що за певних умов також дають однозначність предикатів за змістом.
2. Підстановка Q не бере до уваги зміст предикатних змінних та констант, що приводить до невірних результатів при порівнянні.
3. Велика кількість ітерацій для отримання кінцевого значення підстановки Q для складних та поширених речень.
4. Якщо предикати різні, до них не можна застосувати метод резолюцій, хоча в цей же час предикати можуть бути синонімами.

4. Метод шинглів як найпоширеніший метод порівняння текстів

Ще один із найбільш відомих способів порівняння текстових документів – метод шинглів. Він полягає в тому, що для кожного рядка тексту, який аналізується (шингла), обчислюється контрольна сума. Контрольна сума – це унікальне число, якому ставляться у відповідність деякий текст та функція його обчислення. Потім із всієї множини контрольних сум (їх кількість дорівнює кількості слів у документі мінус $(w - 1)$, де w – число слів у шинглі) відбираються лише ті, які діляться на певне вибране завчасно число. Оскільки значення контрольних сум розподілено рівномірно, критерій відбору ніяк не пов'язаний з особливостями тексту. Один шингл, який співпав при відборі, приблизно відповідає наперед заданому числу однакових частин у повному тексті.

На відміну від описаного методу шинглів існує базовий метод шинглів, в якому беруться до уваги всі шингли, отримані з тексту, а не лише кратні наперед заданому числу, що забезпечує більш широку вибірку для коротких текстів.

Реалізація алгоритму порівняльного аналізу електронних текстів передбачає декілька етапів:

1. Завантажується перший текст.
2. Для завантаженого тексту будується канонічна форма (відбувається видалення зайвих пробілів, знаків табуляції та спеціальних символів).
3. Визначається кількість слів у тексті.
4. Формується масив слів тексту.
5. Із масиву слів будується масив шинглів (виходячи з заданої довжини одного шингла).
6. Для кожного шингла обчислюється його хеш-функція. В залежності від кількості слів у тексті визначаються шингли, кратність контрольної суми яких береться до уваги, та створюється масив контрольних сум.
7. Аналогічно виконуються етапи для другого тексту.
8. Для кожного шингла першого тексту здійснюється пошук шингла другого тексту з однаковою контрольною сумою. Якщо такий шингл не знайдено, то лічильник однакових шинглів збільшується на одиницю.
9. Обчислюється процент збігів як відношення однакових шинглів до числа шинглів першого масиву.

Для конкретного прикладу порівняння електронних документів довільної структури та отримання певного відсотку збігів реалізація даного алгоритму буде складатися з таких етапів:

1. Розбиття кожного тексту, що розглядається, на токени – слова.
2. Визначення множини унікальних слів в одному з текстів, що порівнюються. Нехай $A_1 = \{a_i, i = \overline{0, n}\}$ – множина унікальних слів першого тексту, де a_i – деяке унікальне слово. Під унікальним словом розуміємо слово, що входить саме до конкретного тексту і не повинно обов'язково входити до інших електронних текстів.

3. Формування частотної таблиці залежності унікальних слів від частоти їх вживання у тексті (табл. 1).

Таблиця 1. Таблиця залежності унікальних слів від частоти їх вживання у першому тексті

| | 1 | 2 | ... | j | $j+1$ | ... | m |
|-----------|-----|-----|-----|-----|-------|-----|-----|
| a_1 | 1 | 1 | ... | 0 | 1 | ... | ... |
| a_2 | 0 | 0 | ... | 1 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| a_i | 0 | 1 | ... | 1 | 1 | ... | 1 |
| a_{i+1} | 0 | 0 | ... | 1 | 1 | ... | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| a_n | 1 | 1 | ... | 0 | 1 | ... | 0 |

Кожен рядок даної таблиці відповідає деякому унікальному слову, а кожен стовпчик – номеру рядка тексту, в якому це слово зустрічається ($j = \overline{1, m}$, де m – кількість рядків в електронному тексті). Елементами таблиці є нулі та одиниці в залежності від того, чи зустрічається певне унікальне слово a_i ($i = \overline{0, n}$) в даному рядку. Це дозволяє розглядати частотну таблицю як матрицю суміжності для неорієнтованого графа.

1. Визначення множини унікальних слів у другому тексті, що порівнюється. Нехай $A_2 = \{a_{i'}, i' = \overline{0, n}\}$ – множина унікальних слів другого тексту, де $a_{i'}$ – деяке унікальне слово другого тексту.

2. Формування частотної таблиці залежності унікальних слів від частоти їх вживання у тексті (табл. 2).

Кожен рядок даної таблиці відповідає деякому унікальному слову, а кожен стовпчик – номеру рядка тексту, в якому це слово зустрічається ($j' = \overline{1, m'}$, де m' – кількість рядків у другому електронному тексті). Елементами таблиці є нулі та одиниці в залежності від того, чи зустрічається певне унікальне слово $a_{i'}$ ($i' = \overline{0, n}$) в даному рядку. Отже, ми можемо розглядати частотну таблицю як матрицю суміжності для неорієнтованого графа.

1. На основі табл. 2 та табл. 3 складається узагальнена таблиця виявлення збігів унікальних слів у рядках двох електронних текстів, що порівнюються. Основою таблиці є дані по першому тексту, що порівнюється. Елементами таблиці є кількість унікальних слів, що співпали в обох текстах, розмежовані по рядках тексту.

Таблиця 2. Таблиця залежності унікальних слів від частоти їх вживання у другому тексті

| | | | | | | | |
|-----------|-----|-----|-----|------|--------|-----|------|
| | 1 | 2 | ... | j' | $j'+1$ | ... | m' |
| a_1 | 1 | 1 | ... | 0 | 1 | ... | ... |
| a_2 | 0 | 0 | ... | 1 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| a_j | 0 | 1 | ... | 1 | 1 | ... | 1 |
| a_{j+1} | 0 | 0 | ... | 1 | 1 | ... | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| a_n | 1 | 1 | ... | 0 | 1 | ... | 0 |

Таблиця 3. Узагальнена таблиця збігів

| | | | | | | | |
|-------|-------|-------|-----|-------|-----------|-----|-------|
| | a_1 | a_2 | ... | a_i | a_{i+1} | ... | a_n |
| 1 | 1 | 2 | ... | 1 | 2 | ... | 2 |
| 2 | 1 | 3 | ... | 2 | 4 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| j | 2 | 0 | ... | 2 | 1 | ... | 1 |
| $j+1$ | 2 | 2 | ... | 1 | 2 | ... | 4 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| m | 2 | 1 | ... | 4 | 1 | ... | 1 |

2. Обчислення відсотку збігів для двох текстів:

$$R = \frac{\sum_{j=1}^m q_j}{n + n'}$$

де q_j – кількість знайдених збігів серед унікальних слів текстів, що порівнюються.

Якщо порівнюються не два тексти, а декілька, то формула набуде вигляду

$$R = \frac{\sum_{j=1}^m q_j}{\sum_{k=2}^l n_k}$$

де l – кількість текстів, що порівнюються.

5. Метод порівняння логіко-лінгвістичних моделей речень

Кожне просте речення природної мови можна представити у вигляді його логіко-лінгвістичної моделі типу (5):

$$P(x_1 \{ \bigwedge_{d_1 \in C_1(x_1)} c_{1d_1} \}, \bigwedge_{q_1 \in J(S)} [\bigwedge_{q \in X(S)} [x_q \{ \bigwedge_{d_2 \in C_2(x_q)} c_{qd_2} \}]]]), \quad (5)$$

де P – предикат, що відображає зміст речення;

x_1 – предикатна змінна (суб'єкт) знаходиться у предикативному відношенні з P ;

c_{1d_1} – предикатна константа, що вказує на ознаку суб'єкта;

d_1 – номер предикатної константи, що вказує на ознаку суб'єкта;

$C_1(x_1)$ – множина предикатних констант суб'єкта x_1 ;

x_q – предикатна змінна (аргумент);

q – номер предикатної змінної (аргументу), початкове значення якого $q = 2$;

$X(S)$ – множина предикатних змінних (аргументів);

c_{qd_2} – предикатна константа, що вказує на ознаку q -ї предикатної змінної (аргументу або об'єкта);

d_2 – номер предикатної константи, що вказує на ознаку предикатної змінної (аргументу);

$C_2(x_q)$ – множина предикатних констант предикатної змінної x_q ;

$J(S)$ – множина предикатних змінних, які виконують у реченні рівнозначну роль, $J(S) \in X(S)$;

q_1 – номер предикатної змінної із множини $J(S)$, якщо речення не має ієрархічної будови або в ньому не зустрічаються аргументи, рівносильні за своєю роллю, то $J(S) = \emptyset$ [7].

Основою принципу побудови логіко-лінгвістичної моделі (5) служить синтаксичний розбір речення, тобто визначення зв'язків між усіма його словами та встановлення їх синтаксичних ролей, що дає змогу зрозуміти зміст текстової інформації.

Таким чином, структура логіко-лінгвістичної моделі (5), а також синтаксична структура речень природної мови обумовлюють правила порівняння таких моделей.

Нехай є два речення «Стаття присвячена дослідженню існуючих методів виявлення логічних протиріч у текстових документах на основі використання формул логіки предикатів, зокрема, дослідженню методу резолюцій» та «Статтю присвячено дослідженню існуючих методів виявлення логічних протиріч у текстових документах на основі використання формул логіки предикатів, зокрема, дослідженню методу резолюцій». Ці речення мають однаковий зміст, проте одне з них в активному стані (суб'єкт виконує дію), друге – в пасивному (суб'єкт відсутній).

Логіко-лінгвістичні моделі заданих речень будуть мати вигляд (6) и (7):

$$\begin{aligned} \text{Присвячена} & \left(\begin{array}{l} \text{стаття} \text{ дослідженню} \left[\begin{array}{l} \text{методи} \{ \text{існуючих} \} \left[\begin{array}{l} \text{виявлення} \left[\begin{array}{l} \text{протиріч} \{ \text{логічних} \} \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{основі} \left[\begin{array}{l} \text{використання} \left[\begin{array}{l} \text{формул} \left[\begin{array}{l} \text{логіки} \left[\begin{array}{l} \text{предикатів} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \left[\begin{array}{l} \text{методу} \\ \text{[резолюцій]} \end{array} \right] \end{array} \right] \end{array} \right) \\ P_1 = (x_1, x_2 [x_3 \{c_{31}\} [x_4 [x_5 \{c_{51}\}]]], x_6 [x_7 [x_8 [x_9 [x_{10}], x_{11} [x_{12} [x_{13}]]]]) \end{aligned} \quad (6)$$

$$\begin{aligned} \text{Присвячено} & \left(\begin{array}{l} \text{дослідженню} \left[\begin{array}{l} \text{методи} \{ \text{існуючих} \} \left[\begin{array}{l} \text{виявлення} \left[\begin{array}{l} \text{протиріч} \{ \text{логічних} \} \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{основі} \left[\begin{array}{l} \text{використання} \left[\begin{array}{l} \text{формул} \left[\begin{array}{l} \text{логіки} \left[\begin{array}{l} \text{предикатів} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \left[\begin{array}{l} \text{методу} \\ \text{[резолюцій]} \end{array} \right] \end{array} \right) \\ P_2 = (x_2, [x_3 \{c_{31}\} [x_4 [x_5 \{c_{51}\}]]], x_6 [x_7 [x_8 [x_9 [x_{10}], x_{11} [x_{12} [x_{13}]]]]) \end{aligned} \quad (7)$$

Порівняння речень різного стану відбувається за такими принципами:

1. Відбувається порівняння предикатів речень P_1 та P_2 : якщо P_1 та P_2 – дієслова, їх час співпадає, P_2 – дієслово в пасивній формі, а P_1 – дієслово в активній формі, а також співпадають корені та суфікси предикатів, або P_1 та P_2 є синонімами, то предикати P_1 і P_2 можна вважати тотожними за змістом.

2. Якщо в одному із речень, що порівнюються, відсутній суб'єкт, а x_1 та x_2 іменники у називному та знахідному відмінках відповідно, а кількість предикатних змінних в од-

ному з речень на одну більше, ніж у іншому, x_1 та x_2 – спільнокореневі слова або є синонімами, то суб'єкти речень (віртуальний у випадку пасивного стану) можна вважати тотожними за змістом.

3. Якщо справджуються умови 1 та 2, то речення можна вважати однаковими за змістом.

6. Висновки

На даному прикладі продемонстровано можливість здійснення порівняльного аналізу окремих речень природної мови різної структури за змістом за допомогою використання логіко-лінгвістичних моделей цих речень [8]. Як видно з прикладу, це дуже важка і громізка процедура, в якій необхідно врахувати абсолютно всі (як синтаксичні, так і семантичні) особливості речень та зв'язки між словами та словосполученнями.

Принципи порівняння речень природної мови з використанням логіко-лінгвістичних моделей продемонстровано для конкретних типів речень. Для того, щоб уніфікувати принципи здійснення порівняльного аналізу речень довільної структури за змістом, необхідно створити базу знань у вигляді продукційної моделі представлення знань, що міститиме правила порівняння.

СПИСОК ЛІТЕРАТУРИ

1. Дракин В.И. Общение конечных пользователей с системами обработки данных / Дракин В.И., Попов Э.В., Преображенский А.Б. – М.: Радио и связь, 1988. – 288 с.
2. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа / Ландэ Д.В. – М.: ООО «Вильямс», 2005. – 272 с.
3. Кузин Л.Т. Основы кибернетики: в 2 т. / Кузин Л.Т. – М.: Энергия, 1979. – Т. 2: Основы кибернетических моделей. – 584 с.
4. Хант Э. Искусственный интеллект / Хант Э.; пер. с англ. – М.: Мир, 1978. – 558 с.
5. Вагин В.Н. Дедукция и обобщение в системе принятия решений / Вагин В.Н. – М.: Наука, 1988. – 384 с.
6. Вавіленкова А.І. Логіко-лінгвістична модель як засіб відображення синтаксичних особливостей текстової інформації / А.І. Вавіленкова // Математичні машини і системи. – 2010. – № 2. – С. 134 – 137.
7. Джарратано Д. Экспертные системы: принципы разработки и программирование / Д. Джарратано, Г. Райли; пер. с англ. К.А. Птицына. – [4-е изд.]. – М.: ООО «Вильямс», 2007. – 1152 с.
8. Вавіленкова А.І. Обробка текстової інформації через призму аналізу та інтерпретації елементів формальної системи / А.І. Вавіленкова // Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю. – Київ: ПММС НАНУ, 2009. – С. 198 – 201.

Стаття надійшла до редакції 09.06.2011