

Міністерство освіти і науки, молоді та спорту України
Харківський національний університет радіоелектроніки

ЛЕЩИНСЬКА ІРИНА ОЛЕКСАНДРІВНА

УДК 519.711.3:519.68

**АЛГЕБРОЛОГІЧНІ МОДЕЛІ СТРУКТУР ПРИРОДНОЇ МОВИ ТА ЇХ
ЗАСТОСУВАННЯ В СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ**

05.13.23 – системи та засоби штучного інтелекту

А В Т О Р Е Ф Е Р А Т
дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2012

Дисертацією є рукопис.

Роботу виконано у Харківському національному університеті радіоелектроніки, Міністерства освіти і науки, молоді та спорту України.

Науковий керівник: доктор технічних наук, професор
Шабанов-Кушнарєнко Сергій Юрійович,
професор кафедри прикладної математики
Харківського національного університету
радіоелектроніки МОНмолодьспорту України,
м. Харків

Офіційні опоненти: доктор технічних наук, професор
Шостак Ігор Володимирович,
професор кафедри інженерії програмного
забезпечення ЕОМ Національного аерокосмічного
університету ім. М.Є. Жуковського «ХАІ»
МОНмолодьспорту України, м. Харків;

кандидат технічних наук, доцент
Ситников Дмитро Едуардович,
завідувач кафедри інформаційно-документних
систем Харківської державної академії культури
Міністерства культури України, м. Харків.

Захист відбудеться “___” _____ 2013 р. о ___ годині на засіданні спеціалізованої вченої ради Д64.052.01 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

З дисертацією можна ознайомитися в бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

Автореферат розісланий “___” _____ 2012 р.

Вчений секретар
спеціалізованої вченої ради

Є.І. Литвинова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Розробка нових технологій, підкріплених якісно виконаним програмним забезпеченням персональних комп'ютерів та інших застосувань, що дозволяють користувачеві задовольняти свої інформаційні запити та вирішувати поставлені перед ним завдання будь-якого ступеня складності, є одним з найважливіших в галузі перспективних напрямків розвитку комп'ютерних наук.

Для ефективного використання можливостей комп'ютерної техніки (насамперед в інформаційно-пошукових, експертних системах різних галузей застосування, у бібліографічних системах, у системах машинного перекладу) необхідно, щоб мова обміну інформацією «людина-система» була максимально наближена до природної. Це значно полегшить процес навчання під час роботи з інтелектуальними системами, спростить і саму роботу. Важливу роль під час розробки математичного забезпечення інформаційно-інтелектуальних систем відіграють реляційні та логічні засоби подання знань. Одним з ефективних універсальних математичних засобів для опису інформації є алгебри предикатів і предикатних операцій. Мовою цих алгебр легко й зручно описувати різну інформацію, що формалізується, моделювати інтелектуальну діяльність людини.

Алгебра скінченних предикатів є узагальненням апарату булевих функцій і апарату багатозначної логіки. Використання цієї алгебри дозволило формально описувати абстрактні поняття, якими користується людина у своїй інтелектуальній діяльності. При цьому виникла необхідність у розробці алгебр скінченних предикатів більш високого порядку, які дозволили б у набагато більш компактній і природній формі математично описувати численні й різноманітні поняття високого ступеня абстракції, що широко використовуються людським інтелектом – алгебри підстановочних операцій, предикатних операцій, прикладна й фундаментальна алгебри предикатних операцій, кванторна алгебра.

Останні розробки в теорії інтелекту, в галузі розвитку математичного апарата алгебри предикатів і предикатних операцій, пов'язані із проектуванням реляційних мереж. Реляційні мережі є схемною реалізацією формул алгебри скінченних предикатів. Методи синтезу реляційних мереж базуються на принципі паралельної обробки символічної інформації. Основна ідея побудови реляційних мереж полягає у бінарній кон'юнктивній декомпозиції багатомісних предикатних моделей, тобто в переході від системи багатомісних предикатів до системи бінарних предикатів. Це дозволяє паралельно обробляти завдання, що вимагають обчислень великого обсягу або обчислень в реальному темпі часу. Прикладом таких завдань є обробка зв'язного тексту – переклад з однієї мови на іншу і розпізнавання мовлення.

Розвиток теорії лінійних логічних перетворень є актуальним завданням для подальшої розробки теорії реляційних мереж, оскільки функціонуюча реляційна мережа є системою взаємодіючих лінійних логічних перетворень. У даній роботі розвивається математичний апарат лінійних логічних перетворень за рахунок введення дуальних лінійних логічних перетворень для забезпечення ефективної роботи реляційних мереж.

Значний внесок у постановку і розв'язання завдань моделювання природної мови внесли Ю.Д. Апресян, Г.Г. Белоногов, Т. Виноград, М.Ф. Бондаренко, А.В. Гладкий, А.С. Наріньяні, Р.Т. Піотровський, Д.А. Поспелов, Дж. Слокум, М. Хомський, Ю.П. Шабанов-Кушнарєнко, Р. Шенк, В.А. Широков та багато інших вчених.

Актуальність даної роботи визначається перспективністю використання отриманих алгебологічних моделей структур російської мови, що дозволяють істотно розширити клас завдань, розв'язуваних на ЕОМ у реальному темпі часу. Зокрема, для формалізації багатьох інформаційних процесів, у тому числі й розробки інтелектуальних систем спілкування з комп'ютером природною мовою, а також для розвитку апарату алгебри предикатів і предикатних операцій як універсального засобу формального опису об'єктів різної природи.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційну роботу виконано відповідно до плану науково-технічних робіт Харківського національного університету радіоелектроніки в рамках держбюджетної теми № 232-2 «Розробка теорії й принципів побудови мозкоподібних ЕОМ з реляційними мережами» (№ ДР 0109U001646).

Автором під час виконання цієї теми досліджено, розроблено, запропоновано й реалізовано алгебологічні методи та апаратно-програмні засоби синтезу реляційних мереж для побудови інтелектуальних систем з природно-мовним інтерфейсом користувача.

Мета і задачі дослідження. Метою дослідження є розробка комплексу методів та моделей для формалізації лінгвістичних структур природної мови на базі формального апарату алгебри скінченних предикатів та реляційних мереж для розширення можливостей інтелектуального користувальницького інтерфейсу, наступна реалізація отриманих моделей в інтелектуальних природно-мовних системах.

Мета дисертаційної роботи була досягнута шляхом розв'язання таких основних поставлених задач дослідження:

- розробка методу формульного опису відношень, що базується на заміні їх відображеннями за допомогою імплікативного розкладання предикатів;
- розробка методу формульного опису відношень, що базується на методі Гільберта формульного опису відношень та алгебрі предикатів;
- розробка моделей лінійних логічних операторів першого й другого роду, що описують функціонування реляційних мереж;

-розробка методів та моделей лінгвістичної алгебри – математичного апарата для аналізу структури природної мови – методу побудови предиката речення шляхом введення доповнюючих та несуттєвих предметних змінних;

-розробка методу побудови формул алгебри булевих функцій, що моделюють речення природної мови;

-розробка методу побудови логічної асоціативної структури реляційної мережі, що дозволяє підвищити швидкодію паралельного асоціативного пошуку рішень за мінімальним числом таблиць;

-застосування отриманих методів розв'язання систем логічних рівнянь для побудови інтелектуальних систем.

Об'єкт дослідження – процес паралельної автоматичної обробки текстів природної мови.

Предмет дослідження – алгебрологічні моделі структур природної мови – речень, словосполучень, слів, а також реляційні мережі, які моделюють ці структури.

Методи дослідження. Як основний математичний апарат обрано логічний аналіз, алгебру скінченних предикатів – для логіко-математичних побудов, створення повноцінного апарата лінгвістичної алгебри; теорію лінійних логічних операторів – для розробки моделі лінійних логічних операторів першого й другого роду; апарат багат шарової декомпозиції предикатів – для розвинення теорії реляційних мереж; теорію компараторної ідентифікації – для обґрунтування структури лінгвістичної алгебри. Також використовувалися основні поняття булевої алгебри й теорії графів для формалізації дерев синтаксичного підпорядкування.

Теоретичними передумовами дисертаційної роботи є логічний аналіз, алгебра скінченних предикатів, математичні моделі семантичної структури природної мови; формальні методи побудови реляційних мереж.

Наукова новизна отриманих результатів.

1. Вперше запропоновано моделі лінійних логічних операторів першого й другого роду, які описують функціонування реляційних мереж та базуються на теорії лінійних логічних операторів, що дозволяють спростити розв'язання системи логічних рівнянь на кожному такті реляційної мережі.
2. Вперше розроблено метод побудови логічної асоціативної структури реляційної мережі, який базується на використанні апаратної пам'яті з асоціативним доступом, що дозволяє спростити доступ до даних та підвищити швидкість пошуку розв'язку.
3. Вперше розроблено метод побудови предикатної моделі речення природної мови шляхом введення доповнюючих та несуттєвих предметних змінних. Метод базується на апараті алгебри скінченних предикатів і лінгвістичних експериментах та забезпечує однозначність предикатної моделі речення.
4. Набула подальшого розвитку лінгвістична алгебра шляхом розробки методу

побудови формул алгебри булевих функцій, які моделюють речення природної мови, що дозволяє спростити формалізацію проєктивних речень.

Практичне значення отриманих результатів. Розроблені в дисертаційній роботі математичні методи формалізації структур природної мови, алгоритми й програмна система призначені для автоматизованих інформаційних систем із природно-мовним інтелектуальним інтерфейсом, для логічної підтримки проєктування інформаційних структур. Математичні й програмні результати роботи можуть бути використані в системах автоматичної обробки текстової інформації (біржі послуг, електронна пошта й т.д.).

Результати, отримані в ході даної дисертаційної роботи, знайшли своє практичне застосування в ОКБ Харківського державного приладобудівного заводу ім. Т.Г. Шевченка для підтримки конструкторських рішень під час розробки системи моніторингу розподільчих електромереж (СМ РЕМ), призначеної для моніторингу основних технічних параметрів нормального й аварійного режимів роботи електромереж 6 (10) кВ, метеорологічних умов (температури й відносної вологості повітря) у місцях експлуатації й оптимізації пошуку та локалізації місця ушкодження повітряних ліній (акт впровадження від 25.12.2008 р.); у ПП «Енергія» для розробки комп'ютерного комплексу створення єдиного інформаційного середовища керування основними фондами й ресурсами підприємств з метою автоматизації процесу контролю технологічних сервісів підприємств (договір №07-18 на створення науково-технічної продукції від 24.04.2007 р., акт впровадження від 19.09.2007 р.).

Теоретичні результати дисертації були використані у навчальному процесі для студентів 3-го та 4-го курсів спеціальності 06.050103 - «Програмна інженерія» факультету комп'ютерних наук ХНУРЕ під час підготовки лекцій з навчальних дисциплін «Теорія інтелекту» та «Біоніка інтелекту» (акт впровадження від 28.10.2011 р.).

Особистий внесок здобувача. Усі результати, наведені в роботі, отримані здобувачем особисто. У роботах, виконаних у співавторстві, здобувачеві особисто належать такі результати: у роботі [1] розроблено формальний опис поняття предметного простору мовою алгебри скінченних предикатів; у роботі [2] побудовано модель бінарної логічної мережі для булевого рівняння; у роботі [3] проаналізовано булеву алгебру предикатних операцій, алгебру предикатних операцій з константами й змінними, диз'юнктивно-кон'юнктивну алгебру предикатних операцій, фундаментальну алгебру предикатних операцій та прикладну алгебру предикатних операцій; у роботі [4] лінійні логічні оператори застосовано для розв'язання логічних рівнянь у реляційних мережах; у роботі [6] запропоновано метод моделювання дерева граматичних залежностей речень природної мови у вигляді реляційної мережі; у роботі [8] запропоновано лінійні логічні оператори першого та другого роду та відповідні їм реляційні мережі; у роботі [9] розроблено метод побудови логічної асоціативної структури

реляційної мережі; у роботі [10] розроблено лінгвістичну алгебру – модель семантичної структури тексту; у роботі [11] виконано бінарizaцію алгебологічної моделі відмінювання іменників російської мови. У тезах доповідей [12, 15 – 18] наведено апробацію розроблених алгебологічних моделей реляційних мереж.

Апробація результатів. Основні положення і результати дисертаційної роботи були повідомлені й апробовані на дев'яти науково-технічних конференціях [12 – 20]: 1) VIII Всеукраїнська (III Міжнародна) студентська наукова конференція з прикладної математики та інформатики (Львів, 14 – 15 квітня 2005 р.); 2) 9-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті» (Харків, 19 - 21 квітня 2005 р.); 3) 10-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті» (Харків, 10 – 12 квітня 2006 р.); 4) Міжнародна наукова конференція «MegaLing'2006 – Горизонти прикладної лінгвістики та лінгвістичних технологій» (Партевіт, 20 - 27 вересня 2006 р.); 5) Міжнародна наукова конференція «MegaLing'2009 – Горизонти прикладної лінгвістики та лінгвістичних технологій» (Київ, 20 – 27 вересня 2009 р.); 6) XV-th International Conference “Knowledge-Dialogue-Solution” (Sofia, Bulgaria, 2009); 7) Друга факультетська науково-практична молодіжна школа-семінар «Інформаційні інтелектуальні системи - 2009» (Харків, 2 – 4 грудня 2009 р.); 8) 14-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті» (Харків, 24 - 26 березня 2010 р.); 9) 17 Міжнародна конференція з автоматичного управління «Автоматика-2010» (Харків, 27 - 29 вересня 2010 р.).

Публікації. Основні результати дисертаційної роботи опубліковано у 20 наукових працях, зокрема 11 статей [1 – 11] у виданнях, що входять до переліку фахових наукових видань України (2 одноосібно), 9 публікацій у матеріалах та тезах доповідей міжнародних наукових конференцій (з них 4 – одноосібно, 1 – публікація за кордоном).

Структура й обсяг дисертаційної роботи. Дисертаційна робота складається із вступу, чотирьох розділів, висновків, списку використаних джерел, трьох додатків. Повний обсяг роботи – 146 сторінок, з них основного тексту – 132 сторінки. Дисертація містить 34 рисунки на 26 сторінках, 25 таблиць на дев'яти сторінках, 3 додатки на трьох сторінках, список використаних джерел із 135 найменувань на 11 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі висвітлено основні проблеми проектування та розробки інтелектуальних систем різного призначення, обґрунтовано актуальність теми, сформульовано мету і задачі дослідження, наведено відомості про зв'язок роботи з науковими програмами, планами і темами організації, у якій

виконувалися дослідження. Надано стислу характеристику об'єкта, предмета і методів дослідження, наукової новизни, практичного значення і використання отриманих результатів в інтелектуальних системах штучного інтелекту.

У **першому розділі** дисертації викладено огляд основних наукових досягнень в галузях теорії інформатизації, логічного аналізу та формального опису механізмів природної мови. Проаналізовано алгебологічні результати в галузях формалізації семантики природної мови та методів побудови реляційних мереж. Обґрунтовано вибір напрямку і сформульовано задачі дослідження.

У **другому розділі** описано способи формульного подання відношень і дій над ними. Розглянуто методи розв'язання логічних рівнянь реляційних мереж за допомогою лінійних логічних операторів. Розроблено моделі лінійних логічних операторів першого й другого роду, що описують функціонування реляційних мереж. Сформульовано й доведено основні властивості логічних операторів першого й другого роду.

Метод Гільберта формульного подання відношень поширюється на будь-які відношення. Співвідношення

$$P(x_1, x_2, \dots, x_m) = \begin{cases} 1, & \text{якщо } (x_1, x_2, \dots, x_m) \in P; \\ 0, & \text{якщо } (x_1, x_2, \dots, x_m) \notin P \end{cases}$$

встановлює бієктивну відповідність між усіма відношеннями множини L всіх відношень на A і всіма предикатами множини M усіх предикатів на A .

Перехід від предиката P до відповідного йому відношення P відбувається за правилом:

$$\begin{aligned} \text{якщо } P(x_1, x_2, \dots, x_m) = 1, & \text{ то } (x_1, x_2, \dots, x_m) \in P; \\ \text{якщо } P(x_1, x_2, \dots, x_m) = 0, & \text{ то } (x_1, x_2, \dots, x_m) \notin P. \end{aligned}$$

Можливий також інший спосіб опосередкованого формульного опису відношень, заснований на заміні їх відображеннями. Формульне подання відображень засноване на імплікативному розкладанні предиката:

$$F(x, y) = \bigwedge_{a \in A} (x^a \supset F(a, y)). \quad (1)$$

Згідно з (1) будь-яке відображення $F(x)=y$, яке діє з A в B , що відповідає предикату $F(x, y)$ на $A \times B$, можна записати системою умов $x^a \supset F(a, y)$, $a \in A$.

Багатомісні відображення отримуємо, переходячи від предиката $F(x, y)$ на $A \times B$ до предиката $F(x_1, x_2, \dots, x_n, y)$ на $A_1 \times A_2 \times \dots \times A_n \times B$ та від відображення $F(x) = y$, що діє з множини A в множину B , до відображення $F(x_1, x_2, \dots, x_n) = y$ з $A_1 \times A_2 \times \dots \times A_n$ в B . Імплікативне розкладання багатомісного предиката має вигляд:

$$F(x_1, x_2, \dots, x_n, y) = \bigvee_{\substack{a_1 \in A_1 \\ a_2 \in A_2 \\ \dots \\ a_n \in A_n}} (x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \supset F(a_1, a_2, \dots, a_n, y)).$$

Максимальним образом множини $P \subseteq A_1$ щодо відповідності $K(r, s) = 1$ називається множина $Q_{\max} \subseteq A_2$, що є об'єднанням образів всіх предметів $r \in P$. Він обчислюється за формулою:

$$\exists x_1 \in A_1 (P(x_1) \cdot K(x_1, x_2)) = Q_{\max}(x_2) \quad (2)$$

Мінімальним образом множини $P \subseteq A_1$ щодо відповідності $K(r, s) = 1$ називається множина $Q_{\min} \subseteq A_2$, що є перетинанням образів усіх предметів $r \in P$. Він обчислюється за формулою

$$\forall x_1 \in A_1 (P(x_1) \supset K(x_1, x_2)) = Q_{\min}(x_2). \quad (3)$$

Перетворення $F(P) = Q_{\max}$ вигляду (2) адитивне $F(P_1 \vee P_2) = F(P_1) \vee F(P_2)$ щодо операції диз'юнкції й однорідне $F(\alpha P) = \alpha F(P)$ щодо операції кон'юнкції, $\alpha \in \{0, 1\}$. Воно називається лінійним логічним оператором першого роду.

Перетворення $\Upsilon(P) = Q_{\min}$ вигляду (3) адитивне $\Upsilon(P_1 \wedge P_2) = \Upsilon(P_1) \wedge \Upsilon(P_2)$ щодо операції кон'юнкції й однорідне $\Upsilon(\alpha \vee P) = \alpha \vee \Upsilon(P)$ щодо операції диз'юнкції. Воно називається лінійним логічним оператором другого роду. Можна довести, що будь-який лінійний логічний оператор першого роду виражається у вигляді (2), а другого – у вигляді (3) при належному виборі предиката $K(x_1, x_2)$, що називається ядром лінійного логічного оператора.

Кожна модель реляційної мережі характеризується своїм предикатом моделі. Робота реляційної мережі полягає у розв'язанні системи логічних рівнянь виду $K(x, y) = 1$ на кожному такті. Розглянемо логічне рівняння $K(x, y) = 1$. Обмежимо можливі значення змінної x якою-небудь множиною $P \subseteq U$. Тоді значення змінної y , що задовольняє це рівняння, також виявляться обмеженими деякою множиною $Q \subseteq U$. Множину Q можна виразити через множину P такою залежністю:

$$\exists x \in U K(x, y) P(x) = Q(y), \quad (4)$$

де $P(x)$ – предикат, що відповідає множині P ; $Q(y)$ – предикат, що відповідає множині Q . Запишемо предикатну операцію (4) у стислій формі – $F_K(P) = Q$.

Твердження 1. Будь-яка предикатна операція F_K однорідна щодо операції кон'юнкції.

Твердження 2. Будь-яка предикатна операція F_K адитивна щодо операції диз'юнкції.

Перетворення виду (4) називається лінійною предикатною операцією $F_K(P) = Q$.

Твердження 3. Будь-яка адитивна й однорідна предикатна операція $F_K(P) = Q$ виражається у вигляді (4).

Множина $Q = F_K(P)$ називається образом множини P за відношенням K . Множина P' , обумовлена формулою

$$\exists y \in U K(x, y)Q(y) = P'(x), \quad (5)$$

називається прообразом множини Q за відношенням K . Предикатна операція (5) стисло записується у вигляді $F'_K(Q) = P'$, вона називається дуальною стосовно операції $F_K(P) = Q$.

Перетворення виду (4) - (5) називаються лінійними предикатними операціями першого роду або лінійними логічними операторами першого роду $F_K(P) = Q$, $F'_K(Q) = P'$. $K(x, y)$ називається ядром лінійного логічного оператора.

Існують предикатні операції другого роду, які виражаються у такому вигляді: $F_K^*(P) = \neg F_K(\neg P)$.

Отримаємо загальний вид предикатної операції другого роду:

$$\begin{aligned} F_K^*(P) &= \neg F_K(\neg P) = \neg \exists x \in U K(x, y) \neg P(x) = \\ &= \forall x \in U \neg (K(x, y) \wedge \overline{P(x)}) = \forall x \in U (\overline{K(x, y)} \vee \overline{\overline{P(x)}}) = \\ &= \forall x \in U (\overline{K(x, y)} \vee P(x)) = \forall x \in U (K(x, y) \supset P(x)). \end{aligned}$$

Предикатна операція другого роду $F_K^*(P) = Q$ має такий вигляд:

$$\forall x \in U (K(x, y) \supset P(x)) = Q(y). \quad (6)$$

Твердження 4. Предикатна операція $F_K^*(P) = Q$ (4) однорідна щодо операції диз'юнкції.

Твердження 5. Предикатна операція $F_K^*(P) = Q$ (6) адитивна щодо операції кон'юнкції.

Перетворення (6) називається лінійною предикатною операцією F_K^* .

Твердження 6. Будь-яка адитивна й однорідна предикатна операція F_K^* має вигляд (6).

Множина P' , обумовлена формулою

$$\forall y \in U (K(x, y) \supset Q(y)) = P'(x), \quad (7)$$

називається прообразом множини Q за відношенням K . Предикатна операція (7) стисло записується у вигляді $F'_K(Q) = P'$, вона називається дуальною відносно операції F_K^* .

Під час розв'язання логічних рівнянь реляційної мережі за допомогою

лінійних предикатних операцій першого або другого роду у другому напівтакті кожного такту реляційна мережа знаходить загальну частину $P_i(x)$ всіх знань $P'_1(x), P'_2(x), \dots, P'_l(x)$ про значення кожної зі своїх предметних змінних x (l – число галузей, що підходять до полюса x) за формулами, відповідно:

$$P'_1(x) \wedge P'_2(x) \wedge \dots \wedge P'_l(x) = P_i(x); P'_1(x) \vee P'_2(x) \vee \dots \vee P'_l(x) = P_i(x).$$

У **третьому розділі** набула подальшого розвитку лінгвістична алгебра, що має два яруси – семантичний і синтаксичний. Перший ярус представлений одним з варіантів алгебри предикатів, другий – алгебри предикатних операцій. Вперше розроблено метод побудови предикатної моделі речення природної мови шляхом введення доповнюючих та несуттєвих предметних змінних. Розроблено метод побудови формул алгебри булевих функцій, що моделюють речення природної мови.

Кожне речення виражає деякий предикат $P(x_1, x_2, \dots, x_m) = \xi$, що становить залежність істинності змінної ξ від предметних змінних x_1, x_2, \dots, x_m . Однак, якщо звернутися до конкретних речень природної мови (наприклад, російської), то ніяких предметних змінних у них виявити не вдається. Пояснюється це тим, що речення природної мови, на відміну від математичної формули, виражає не всю функцію $P(x_1, x_2, \dots, x_m)$, а тільки її ім'я P . Людина, перетворюючи те або інше речення у відповідну йому думку, добудовує його до предиката, додаючи до нього (як до ім'я предиката) відсутні предметні змінні. Тільки після цього речення стає доступним для розуміння. І навпаки, перетворюючи деяку думку в речення, людина виключає з неї предметні змінні, передаючи іншим людям не саму думку, а тільки її ім'я. Наприклад, у реченні “На стоянці стоїть машина” вводимо дві предметні змінні. У результаті отримуємо наступне твердження “На стоянці x_1 стоїть машина x_2 ”. Використовуючи речення як ім'я отриманого предиката, останній можемо записати у вигляді «На стоянці стоїть машина (x_1, x_2)». Додамо до цього речення ще одне: “Поруч із нею стоїть мотоцикл”, що утворює з ним єдиний текст. Слово “мотоцикл” вводить третій предмет, що відрізняється від перших двох та вимагає введення ще однієї предметної змінної x_3 . У результаті одержуємо висловлення «Поруч із машиною x_2 стоїть мотоцикл x_3 », що виражає предикат «Поруч із машиною стоїть мотоцикл (x_2, x_3)». Розглянуто питання про несуттєві змінні в тексті мови та у предикатній моделі.

Речення будуються з окремих слів. Показано, що в ролі базисних елементів у лінгвістичній алгебрі виступають слова. Оскільки будь-які елементи носія лінгвістичної алгебри – предикати, тому і слова, розглянуті як базисні елементи, також є предикатами. Показано справедливість цього твердження для різних частин мови – іменників, прикметників, дієслів і т.ін. Розглянуто питання про базисні операції лінгвістичної алгебри. Показано, що ними є заперечення, кон'юнкція й диз'юнкція, а лінгвістична алгебра є булевою алгеброю. З будь-якого речення можна утворити його заперечення, поставивши перед ним частку

“не”. Аналогічним чином розглядаємо сполучники «і» та «або», за допомогою яких можна з'єднувати будь-які речення, отримуючи в результаті нові речення (у загальному випадку – тексти). Застосування цих операцій показано у таких формулах:

$$\begin{aligned}(\neg P)(x_1, x_2, \dots, x_m) &= \neg(P(x_1, x_2, \dots, x_m)); \\ (P \wedge Q)(x_1, x_2, \dots, x_m) &= P(x_1, x_2, \dots, x_m) \wedge Q(x_1, x_2, \dots, x_m); \\ (P \vee Q)(x_1, x_2, \dots, x_m) &= P(x_1, x_2, \dots, x_m) \vee Q(x_1, x_2, \dots, x_m).\end{aligned}$$

У природній мові тексти й відповідні їм змісти не зв'язані взаємно однозначно. Зміст того самого тексту може мінятися залежно від вибору предметних змінних і від контексту. Той самий зміст можна виразити різними текстами.

Речення природної мови є формулами алгебри предикатних операцій. Будь-яка логічна формула має цілком певну структуру, яку можна виразити у вигляді деревоподібного графа. Наприклад, зазначена на рис. 1 схема синтезує формулу $\bar{X}_1 X_2 \vee X_3 \bar{X}_4$ з її аргументів X_1, X_2, X_3, X_4 .

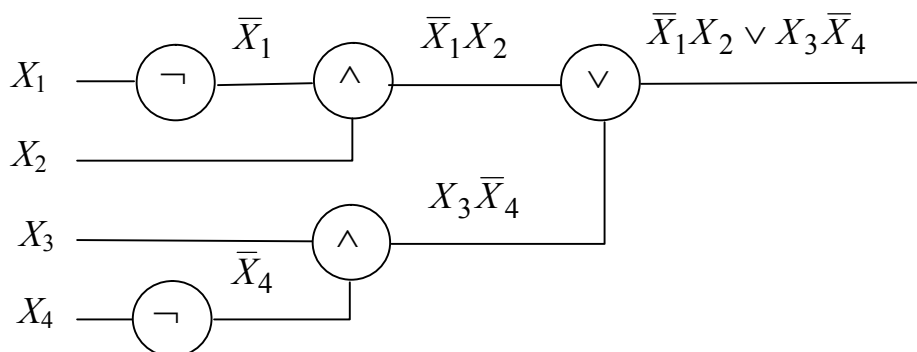


Рис. 1. Графічне подання формули алгебри булевих функцій

У граматиці для наочного подання структури речень використовуються дерева синтаксичного підпорядкування. Побудова дерев синтаксичного підпорядкування для великої кількості проєктивних речень показала, що зв'язки між словами завжди утворюють деревоподібну структуру, аналогічну схемам логічних формул.

Розроблено метод побудови схеми формули речення. Приклад такої схеми показаний на рис. 2.

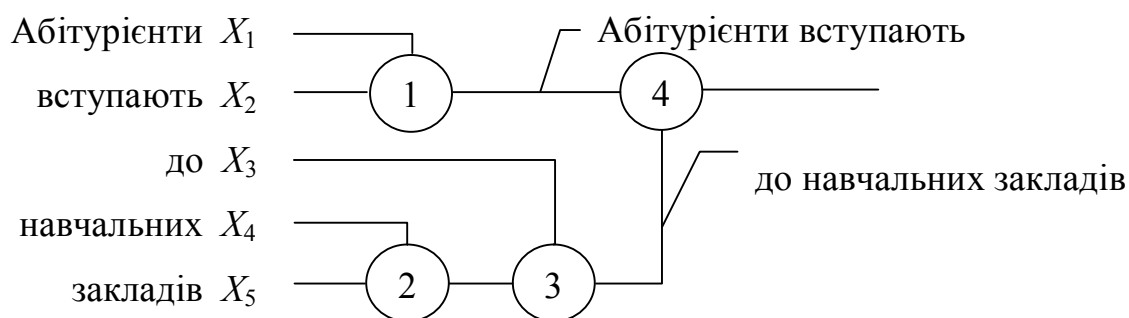


Рис. 2. Схема формули речення

Кружки, позначені номерами, зображують перетворювачі слів і словосполучень. Вони виконують операції з'єднання слів і словосполучень. Схема синтезує речення з окремих слів. Номери блоків указують послідовність виконання операцій перетворювачами схеми. Запропоновано спосіб визначення такої послідовності. За схемою можна побудувати формулу речення. У розглянутому прикладі формула матиме такий вигляд:

(Абітурієнти1(вступають))4(до3(навчальних2(закладів))).

Номери виконують у формулі роль імен операцій, дужки вказують черговість їх виконання і послідовність застосування кожної операції до слів, а форми слів є значенням аргументів формули. Перейдемо до аргументів формули від їх значень, замінюючи слово “Абітурієнти” на змінну X_1 , слово “вступають” – на X_2 , “до” – X_3 , “навчальних” – X_4 , “закладів” – X_5 . У результаті формула запишеться у вигляді $(X_1 1(X_2)) 4(X_3 3(X_4 2(X_5)))$.

Розглянуто питання про те, що конкретно становлять собою операції з'єднання слів і словосполучень. Змінні речення x_1, x_2, \dots, x_n спочатку зв'язуються присудком, що реалізує предикат $S(x_1, x_2, \dots, x_n)$, де n – число предметних змінних при дієслові S . Потім речення доповнюється тими або іншими словосполученнями, що відповідають на запитання щодо предметних змінних x_1, x_2, \dots, x_n . Цим досягається поступове наближення змісту тексту речення до необхідного.

Нехай $T_1(x_{i_{11}}, x_{i_{21}}, \dots, x_{i_{s_1}})$, $T_2(x_{i_{12}}, x_{i_{22}}, \dots, x_{i_{s_2}})$, ..., $T_r(x_{i_{1r}}, x_{i_{2r}}, \dots, x_{i_{s_r}})$ – предикати, що виражаються словами (або словосполученнями) речення, які приєднуються до його присудка. Тут r – число всіх слів (або словосполучень), що приєднуються до присудка речення. Аргументами кожного із предикатів T_j ($j = \overline{1, r}$) служать деякі з аргументів предиката $S(x_1, x_2, \dots, x_n)$ присудка S . Символом s_j позначене число істотних змінних предиката T_j . Тоді предикат речення виразиться у вигляді:

$$P(x_1, x_2, \dots, x_n) = S(x_1, x_2, \dots, x_n) T_1(x_{i_{11}}, x_{i_{21}}, \dots, x_{i_{s_1}}) \wedge \\ \wedge T_2(x_{i_{12}}, x_{i_{22}}, \dots, x_{i_{s_2}}) \dots T_r(x_{i_{1r}}, x_{i_{2r}}, \dots, x_{i_{s_r}}).$$

Остання рівність показує, як зміст речення P утворюється зі змісту присудка S шляхом обмеження його змістом словосполучень T_1, T_2, \dots, T_r .

У четвертому розділі розроблено метод побудови логічної асоціативної структури реляційної мережі, що дозволяє спростити доступ до даних та підвищити швидкість пошуку розв'язків. Описано концепцію паралельної обробки символічної інформації, сутність якої полягає у звертанні до механізмів асоціативної обробки інформації, родинних процесів аналізу й синтезу, що протікають у мозку людини. На прикладі реляційної мережі відмінювання повних неприсвійних прикметників російської мови побудована логічна асоціативна структура, подана у вигляді системи апіорних відношень.

Застосування в реляційних мережах пам'яті з асоціативним доступом дозволяє істотно підвищити швидкість вибірки й спростити доступ до даних у системах обробки природно-мовної інформації.

Логічна асоціативна структура має чотири вершини: $\Psi = f(P_1, P_2, P_3, P_4)$, $O = \{\cup, \cap\}$. Кожна вершина є групою ознак, що впливають на утворення словоформи. Вершина P_1 характеризує вплив контексту, що оточує відмінюване слово: рід, число, відмінок, істотність, архаїчність. Вершина P_2 характеризує саме відмінюване слово: остання буква основи, ударність основи, м'якість основи. Вершина P_3 характеризує словоформу, відповідну даному контексту та даному слову: закінчення форми слова. Вершина P_4 характеризує складений словник типових основ неprisвійних прикметників російської мови.

Для апаратної реалізації побудованої моделі реляційної мережі у вигляді пам'яті з асоціативним доступом складено таблиці, що відповідають трьом групам, у які можна природно об'єднати полюси реляційної мережі. Для кодування алфавітів змінних вершин логічної асоціативної структури використано позиційний та комбінаційний коди. Пошук інформації в пам'яті з асоціативним доступом ведеться, виходячи з деякої ознаки даних. При вибірці вхідні дані одночасно порівнюються з усіма полями ознаки у всіх комірках пам'яті.

Розроблена логічна асоціативна структура призначена для автоматичного розв'язання задач аналізу та синтезу словоформ повних неprisвійних прикметників. Схему для синтезу словоформ повних неprisвійних прикметників наведено на рис. 3.

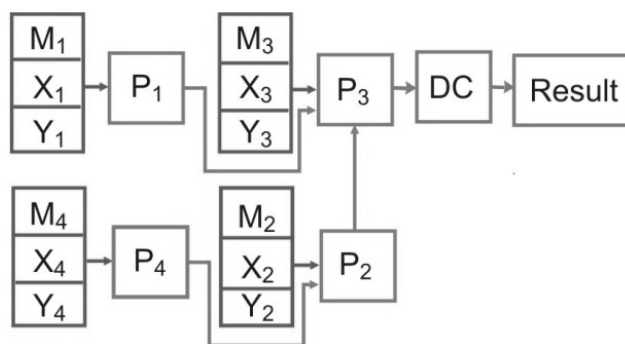


Рис. 3. Схема синтезу словоформ повних неprisвійних прикметників

Фактично дана схема визначає послідовність паралельних (конвеєрних) дій необхідних для обчислення результату - синтезу словоформи прикметника: 1) паралельна обробка взаємодії вхідних векторів з вершинами P_1 і P_4 ; 2) аналіз асоціативної таблиці P_2 ; 3) обробка вершини P_3 завершує процес формування аналізу графа асоціативних таблиць. Для кожної вершини логічної асоціативної

структури задається вектор-маска вхідних і вихідних змінних, а також вхідний вплив. Перспективою подальших досліджень є вибір стратегії аналізу перетворених таблиць, а також синтез подання таблиць у вигляді графів.

Реляційна мережа складається з полюсів і галузей, що з'єднують полюси. Кожному полюсу відповідає своя предметна змінна x_i з областю визначення A_i ($i = \overline{1, m}$). Пари полюсів x_i і x_j , з'єднаних галуззю $K(x_i, x_j)$, реалізують лінійний логічний оператор першого роду $\exists x_i \in A_i (P_i(x_i) K_{ij}(x_i, x_j)) = Q_{j \max}(x_j)$ або другого роду $\forall x_i \in A_i (P_i(x_i) \supset K_{ij}(x_i, x_j)) = Q_{j \min}(x_j)$.

Реляційна мережа є мережею *першого роду*, якщо в ній діють лише оператори першого роду. Аналогічно визначаються реляційні мережі *другого роду*. Якщо в реляційній мережі використовуються оператори обох видів, мережа називається *комбінованою*. Мережа відшукує розв'язок рівняння

$$K(x_1, x_2, \dots, x_m) = 1 \quad (8)$$

за обмежень, що накладаються на області зміни змінних x_i ($i = \overline{1, m}$) $x_i \in P_i$, де $P_i \subseteq A_i$. Якщо розв'язок рівняння відшукується при більш складному обмеженні $L(x_1, x_2, \dots, x_m) = 1$, тоді мережа добудовується таким чином, щоб вона відповідала рівнянню $K' = 1$, де $K' = K \wedge L$. Побудові мережі, що реалізує предикат K , передують *бінаризація* предиката K , тобто подання його у вигляді

$$K(x_1, x_2, \dots, x_m) = \bigwedge_{\substack{i=1 \\ j=1 \\ i \neq j}}^m K_{ij}(x_i, x_j),$$

яка може бути проведена різними способами. Розв'язання рівняння (8) мережею здійснюється за тактами. Протягом кожного такту одночасно спрацьовують всі лінійні логічні оператори мережі. У мережі першого роду після кожного такту здійснюється перетинання всіх множин $Q_{j \max}$, що сходяться з усіх боків до кожного з полюсів x_j . У мережі другого роду множини $Q_{j \min}$, навпаки, поєднуються. Мережа першого роду може формувати зайві розв'язки, а другого – може не знайти деякі з дійсних розв'язків. У процесі розв'язання рівняння (8) зі збільшенням номера такту роботи мережі множини $Q_{j \max}$ й $Q_{j \min}$ зближаються, причому завжди $Q_{j \min} \leq Q_{j \max}$. На деякому такті зближення множин $Q_{j \min}$ і $Q_{j \max}$ припиняється. Якщо це досягається одночасно на всіх полюсах, то на цьому процес розв'язання рівняння (8) закінчується. Якщо при всіх $j = \overline{1, m}$ виявляється, що $Q_{j \min} = Q_{j \max}$, то це означає, що мережа знайшла всі розв'язки, не пропустивши жодного, і не зарахувала до розв'язку жодного помилкового. Якщо ж така рівність не досягнута наприкінці роботи мережі, це означає, що мережа спрацювала не

цілком ефективно. Ця ознака може бути використана при оцінці ступеня доброякісності методу синтезу мережі, зокрема – методу бінаризації предиката K .

Для тестування роботи реляційних мереж у дисертаційній роботі було розроблено програмну реалізацію, що складається із двох частин. Перша частина є редактором реляційної мережі. Вона дозволяє будувати й тестувати реляційні мережі. Друга частина програмної реалізації є веб-додатком. У користувача є можливість задати множину тематичних слів, яка після підтвердження буде передана через веб-сервіс у реляційну мережу, що обробляє вихідні дані й видає результат. На підставі цього результату буде сформований ряд запитів у пошукову систему.

Розроблене програмне забезпечення має практичне значення при складанні семантичного ядра для оптимізації тематичних сайтів і підвищення їх рейтингу при ранжируванні у видачі пошукових систем. Принцип роботи системи у тому, що вона отримує множину тематичних слів, асоційованих з тематикою ресурсу. Система знаходить множину різних слів і на підставі отриманих результатів робить запити у пошукові системи (рис. 4).

| | студент | студента | студенту | студента | студентом | студенте | студенты | ст |
|---|---------|----------|----------|----------|-----------|----------|----------|----|
| студент | | | | | | | | |
| http://ru.wikipedia.org/wiki/%D0%A1%D1%82%D1%83%D0%B4%D0%B5%D0%BD%D1%82 | 24 | 3 | 0 | 3 | 2 | 0 | 7 | |
| http://st-student.com.ua/ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| http://ya-student.com.ua | 4 | 0 | 0 | 0 | 0 | 0 | 0 | |
| http://www.studrespublika.com/vmgo/student.php | 42 | 4 | 0 | 4 | 0 | 0 | 6 | |
| http://student-taxi.com.ua/ | 3 | 0 | 0 | 0 | 0 | 0 | 0 | |
| студента | | | | | | | | |
| http://www.twirpx.com/ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| http://www.sputnikstudenta.com/ | 0 | 1 | 0 | 1 | 0 | 0 | 0 | |
| http://www.ni.kmitd.com.ua/students/rights/ | 0 | ? | ? | ? | 0 | 0 | ? | |

Рис. 4. Результати аналізу статистики пошуку

Результати пошуку аналізуються за певними статистичними критеріями і відображаються у графічному вигляді. На підставі цього фахівець із оптимізації й просування сайтів коректує семантичне ядро сайту й формує завдання для копірайтерів та веб-розробників.

У додатках наведено акти впровадження теоретичних і практичних результатів дисертаційної роботи.

ВИСНОВКИ

У дисертаційній роботі вирішено актуальну науково-практичну задачу побудови комплексу методів та моделей для формалізації лінгвістичних структур природної мови на базі апарату алгебри скінченних предикатів та реляційних мереж для розширення можливостей природно-мовного інтелектуального користувальницького інтерфейсу. Отримано наступні наукові результати:

1. Нові моделі лінійних логічних операторів першого й другого роду, що описують функціонування реляційних мереж. Отримані моделі базуються на теорії лінійних логічних операторів. Введення двоїстих лінійних логічних перетворень спрощують розробку і тестування реляційних мереж.

2. Новий метод побудови логічної асоціативної структури реляційної мережі, що дозволяє підвищити швидкодію (на 15%) паралельного асоціативного пошуку рішень за мінімальною кількістю таблиць при складанні семантичного ядра для оптимізації тематичних сайтів і підвищення їх рейтингу при ранжируванні у видачі пошукових систем.

3. Новий метод побудови предикатної моделі речень природної мови шляхом введення доповнюючих та несуттєвих предметних змінних. Метод забезпечує однозначність при розв'язанні задач аналізу і синтезу речень, тобто їх розуміння та формалізацію.

4. Набув подальшого розвитку математичний апарат для аналізу структури природної мови – лінгвістична алгебра. Показано, що природна мова має двох'ярусну будову. Перший ярус представлений алгеброю предикатів, другий – алгеброю предикатних операцій. Досліджено можливості лінгвістичної алгебри як булевої алгебри.

5. Отримані методи розв'язання систем логічних рівнянь застосовано для побудови інтелектуальної інформаційної системи у ОКБ Харківського державного приладобудівного заводу ім. Т.Г. Шевченка для підтримки конструкторських рішень під час створення системи СМ РЕМ (акт впровадження від 25.12.2008 р.) та в ПП «Енергія» для створення єдиного інформаційного середовища керування основними фондами й ресурсами підприємств (договір №07-18 на створення науково-технічної продукції від 24.04.2007 р., акт впровадження від 19.09.2007 р.). Теоретичні результати дисертації були використані у навчальному процесі на кафедрі програмного забезпечення ЕОМ ХНУРЕ під час підготовки лекцій з навчальних дисциплін «Теорія інтелекту» та «Біоніка інтелекту» для спеціальності 06.050103 - «Програмна інженерія» (акт впровадження від 28.10.2011 р.).

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Бондаренко М.Ф. О мозгоподобных ЭВМ / М.Ф. Бондаренко, З.В. Дударь, И.А. Ефимова, В.А. Лещинский, С.Ю. Шабанов-Кушнарченко // Радиоэлектроника и информатика. – 2004. – № 2. – С. 89 – 105.
2. Ефимова И.А. О методе построения моделей бинарных логических сетей / И.А. Ефимова, В.А. Лещинский // Восточно-Европейский журнал передовых технологий. – 2005.– №4.– С. 121 – 124.

3. Ефимова И.А. О кванторной алгебре предикатных операций / И.А. Ефимова, В.А. Лещинский // Искусственный интеллект. – Донецк, 2006.– № 4. – С. 603 – 612.
4. Ефимова И.А. Синтез бинарных логических сетей и особенности их функционирования / И.А. Ефимова, В.А. Лещинский, В.В. Токарев, Г.Г. Четвериков // Бионика интеллекта. –2006. – №2 (65). – С.14 – 18.
5. Лещинская И.А. Линейные логические операторы первого и второго рода в реляционных сетях / И.А. Лещинская // Системи управління, навігації та зв'язку. – 2010. №2 (14). – С. 214 – 217.
6. Бондаренко М.Ф. Об алгебре одноместных предикатов / М.Ф. Бондаренко, Н.П. Кругликова, И.А. Лещинская, Н.Е. Русакова, Ю.П. Шабанов-Кушнарченко // Бионика интеллекта. – 2010. №2 (73). – С. 62 – 67.
7. Лещинская И.А. О методе построения направленных схем реляционных сетей на примере отношения эквивалентности / И.А. Лещинская // Системи обробки інформації – 2010. – № 1 (82). – С. 75 – 81.
8. Бондаренко М.Ф. О реляционных сетях / М.Ф. Бондаренко, Н.П. Кругликова, И.А. Лещинская, Н.Е. Русакова, Ю.П. Шабанов-Кушнарченко // Бионика интеллекта. – 2010. №3 (74). – С. 8 – 13.
9. Бондаренко М.Ф. Инфраструктура анализа логических ассоциативных отношений / М.Ф. Бондаренко, В.И. Хаханов, И.А. Лещинская, Н.Е. Русакова// Радиоэлектроника и информатика. – 2010. – № 2. – С. 43 – 54.
10. Лещинская И.А. Лингвистическая алгебра как аппарат формализации смысла предложений естественного языка / И.А. Лещинская, В.А. Лещинский, Л.Г. Петрова, С.Ю. Шабанов-Кушнарченко // Системи обробки інформації. –2011. – №2 (92). – С. 34 – 38.
11. Ефимова И.А. Моделирование механизмов естественного языка с помощью бинарных логических сетей / И.А. Ефимова, В.А. Лещинский // Вестник НТУ «ХПИ»: Сб. науч. трудов. Тематич. вып.: Новые решения в современных технологиях. – Харьков: НТУ «ХПИ», 2005. – № 57. – С. 3 – 10.
12. Єфімова І.О. Моделювання логічних мереж для булевих рівнянь / І.О. Єфімова, В.О. Лещинський // VIII Всеукраїнська (III Міжнар.) студентська наук. конф. з прикладної математики та інформатики: Зб. матеріалів. – Львів: ЛНУ, 2005.– С. 128 – 129.
13. Ефимова И.А. Об алгебро-логических моделях механизмов естественного языка / И.А. Ефимова // 9-й Междунар. молодежный форум «Радиоэлектроника и молодежь в XXI веке»: Сб. материалов. – Харьков: ХНУРЭ, 2005.– С. 377.
14. Ефимова И.А. О проблемах и перспективах формализации естественного языка средствами алгебры предикатов и предикатных операций / И.А. Ефимова // 10-й Междунар. молодежный форум «Радиоэлектроника и молодежь в XXI веке»: Сб. материалов. – Харьков: ХНУРЭ, 2006.– С. 197.

15. Ефимова И.А. Построение формальных моделей структур естественного языка с помощью алгебры предикатов и предикатных операций и их реализация в виде бинарных логических сетей / И.А. Ефимова, В.А. Лещинский // Междунар. науч. конф. «MegaLing'2006 – Горизонты прикладной лингвистики и лингвистических технологий»: Сб. докладов. – Симферополь: Изд-во «ДиАйПи», 2006.– С. 207.
16. Четвериков Г.Г. Концептуально-методологічний підхід до моделювання природної мови алгебро-логічними засобами / Г.Г. Четвериков, Т.М. Федорова, І.О. Лещинська, В.В. Токарев, І.Д. Вечирська // Міжнар. наук. конф. «MegaLing'2009 – Горизонти прикладної лінгвістики та лінгвістичних технологій»: Зб. матеріалів. – К., 2009.– С. 68.
17. Chetverikov G. Methods of synthesizing reversible spatial multivalued structures of language systems / G. Chetverikov, I. Leschinskaya, I. Vechirskaya // Information science & computing. – 2009. – N.15 – P. 32 – 39.
18. Лещинская И.А. Аппаратная реализация модели реляционной сети памятью с ассоциативным доступом / И.А. Лещинская, Н.Е. Русакова // Друга факультетська наук.-практ. молодіжна школа-семінар «Інформаційні інтелектуальні системи - 2009»: Сб. матеріалів. – Харків, 2009. – С. 182 – 187.
19. Лещинская И.А. О формальной модели реляционной сети синтаксической структуры предложения естественного языка / И.А. Лещинская // 14-й Междунар. молодежный форум «Радиоэлектроника и молодежь в XXI веке»: Сб. материалов. – Харьков: ХНУРЭ, 2010.– С. 197.
20. Лещинская И.А. Словосочетания естественного языка как реляционные сети / И.А. Лещинская // 17 Міжнар. конф. з автоматичного управління «Автоматика-2010»: Тези доповідей. Т.2. – Харків: ХНУРЕ, 2010. – С. 123 – 125.

АНОТАЦІЯ

Лещинська І.О. Алгебологічні моделі структур природної мови та їх застосування в системах штучного інтелекту. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки МОНмолодьспорту України, Харків, 2012.

Мета дисертаційного дослідження – розробка комплексу методів та моделей для формалізації лінгвістичних структур природної мови на базі апарату алгебри скінченних предикатів та реляційних мереж для розширення можливостей інтелектуального інтерфейсу користувача. Основні результати: розроблено моделі лінійних логічних операторів першого й другого роду, що описують функціонування реляційних мереж; набула подальшого розвитку

лінгвістична алгебра за рахунок розробки моделей базисних елементів і базисних операцій лінгвістичної алгебри та методу побудови формул алгебри булевих функцій, що моделюють речення природної мови; розроблено метод побудови предикатної моделі речення природної мови шляхом введення доповнюючих та несуттєвих предметних змінних; розроблено метод побудови логічної асоціативної структури реляційної мережі, що базується на використанні апаратної пам'яті з асоціативним доступом та використанні мінімального числа таблиць і дозволяє спростити доступ до даних та підвищити швидкість пошуку рішень.

Ключові слова: інтелектуальна система, реляційна мережа, логічне рівняння, логічний оператор, алгебра скінченних предикатів, лінгвістична алгебра, лінійне логічне перетворення, асоціативний пошук.

АННОТАЦІЯ

Лещинская И.А. Алгебологические модели структур естественного языка и их применение в системах искусственного интеллекта. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. – Харьковский национальный университет радиоэлектроники Министерства образования и науки, молодежи и спорта Украины, Харьков, 2012.

Цель диссертационного исследования – разработка комплекса алгебологических средств формализации некоторых лингвистических структур естественного языка в виде набора математических моделей. Полученные результаты использованы для разработки программного обеспечения интеллектуальных систем. Исследованы лингвистические структуры разной степени сложности – предложения, словосочетания, слова и их модели в виде реляционных сетей. Разработаны методы параллельной автоматической обработки текстов естественного языка с помощью реляционных сетей. Выполнен анализ научных исследований в области теории информатизации, существующих средств формализации естественного языка, методов построения реляционных сетей. Обоснована необходимость разработки универсального алгебологического аппарата для описания естественного языка. Показано, что язык алгебры предикатов и предикатных операций эффективен и удобен для описания различной формализуемой информации, моделирования деятельности человека и формирования запросов в базах данных. Разработаны два метода формульного описания отношений: с помощью алгебры предикатов и метод, основанный на замене отношений отображениями. Рассмотрены методы решения логических уравнений с помощью линейных логических операторов и реляционных сетей.

Сформулированы и доказаны основные свойства логических операторов первого и второго рода. Разработан метод явного формульного выражения однозначных отображений. Получил дальнейшее развитие математический аппарат линейных логических преобразований за счет введения двойственных линейных логических преобразований для обеспечения эффективной работы реляционных сетей. Впервые предложены модели линейных логических операторов первого и второго рода, описывающие функционирование реляционных сетей. Разработан математический аппарат для анализа структуры естественного языка – лингвистическая алгебра. В этой алгебре имеются два яруса – семантический и синтаксический. Первый ярус представлен одним из вариантов алгебры предикатов, второй – алгебры предикатных операций. Рассмотрен метод экспериментальной проверки алгебрологических моделей языка. Разработан способ формульной записи смысла словосочетаний и предложений естественного языка. Исследованы возможности лингвистической алгебры как булевой алгебры. Разработана теоретическая схема дополнения предложений предметными переменными для возможности их однозначного понимания. Разработан метод построения графов (деревьев синтаксического подчинения) для предложений естественного языка. Разработан метод построения логической ассоциативной структуры реляционной сети, который позволяет повысить быстродействие параллельного ассоциативного поиска решений по минимальному числу таблиц. Полученные методы решения систем логических уравнений применены для построения интеллектуальных систем в ОКБ Харьковского государственного приборостроительного завода им. Т.Г. Шевченко для поддержки конструкторских решений во время создания системы СМ РЭМ (акт внедрения от 25.12.2008 г.) и в ЧП «Энергия» для создания единой информационной среды управления основным фондами и ресурсами предприятий (акт внедрения от 19.09.2007 г.). Теоретические результаты диссертации были использованы в учебном процессе на кафедре программного обеспечения ЭВМ при подготовке курсов лекций «Теория интеллекта» и «Бионика интеллекта» для специальности «Программная инженерия» (акт внедрения от 28.10.2011 г.).

Ключевые слова: интеллектуальная система, реляционная сеть, логическое уравнение, логический оператор, алгебра конечных предикатов, лингвистическая алгебра, линейное логическое преобразование, ассоциативный поиск.

ABSTRACT

Leshchynska I.O. Algebra-logical models of natural language structures and their application in artificial intelligence systems. - Manuscript.

PhD thesis (candidate degree of technical sciences) in specialty 05.13.23 – Systems and tools of artificial intelligence. – Kharkov National University of Radio Electronics of Ministry education and science, youth and sports of Ukraine, Kharkov, 2012.

The aim of the research is to develop algebra-logical tools of formalizing some natural language linguistic structures as a mathematical models set. The methods for solving logical equations in relational networks and linear logical operators were developed. Mathematical apparatus of linear logical transformation got further development by introduction of dual linear logical transformations for providing efficient work of relational networks. There was developed the mathematical apparatus named “linguistic algebra for analysis of natural language structure”. There are two levels in this algebra – semantic and syntactic. The first level is represented by one of the predicates algebra, the second one is the algebra of predicate operations. The method of experimental verification of algebra-logical models was considered. Also, there was developed the theoretical scheme of supplement to sentences by object variables to obtain their clear understanding and there was devised the method for constructing graphs (trees of syntactic subordination) for sentences of natural language. The method for constructing the logical structure of relational associative networks, which can improve performance of parallel associative search for solutions to the minimal number of tables, was developed.

Key words: intelligent system, relational network, logical equation, logical operator, finite predicates algebra, linguistic algebra, linear logical transformation, associative search.

Відповідальний випусковий **Є.В. Бодянський**

| | | |
|--------------------------|-----------------------|----------------------------|
| Підп. до друку 20.12.12. | Формат 60x84 1/16. | Спосіб друку – ризографія. |
| Умов. друк. арк. 1,2. | Облік. вид. арк. 1,0. | Тираж 100 прим. |
| Зам. № | Ціна договірна. | |

ХНУРЕ, 61166, Харків, просп. Леніна, 14

Віддруковано в навчально-науковому
видавничо-поліграфічному центрі ХНУРЕ
61166, Харків, просп. Леніна, 14