

УДК 519.62



Г.Г. Четвериков¹, О.С. Пузик², В.В. Курасова³, І.К. Божко

¹⁻⁴ХНУРЕ, г.Харьков, Украина, chetvergg@kture.kharkov.ua

УНІВЕРСАЛЬНИЙ БАГАТОЗНАЧНИЙ ФУНКЦІОНАЛЬНИЙ ПЕРЕТВОРЮВАЧ ЯК БАЗОВА КОМІРКА СЕМАНТИЧНОЇ МЕРЕЖІ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ

Статтю присвячено питанню організації даних та ефективного вибору програмних і технічних засобів при розробці мовно-інформаційних інтерфейсів та проєктів. Наведено аналіз тенденцій моделювання механізмів природної мови. Викладено також принципи та методи побудови багатовходових універсальних k -значних просторових структур мовних систем штучного інтелекту.

ЛОГІКА, ФУНКЦІОНАЛЬНИЙ ПЕРЕТВОРЮВАЧ, СКІНЧЕННИЙ ПРЕДИКАТ, АЛГЕБРА СКІНЧЕННИХ ПРЕДИКАТІВ, БАГАТОЗНАЧНА СТРУКТУРА, АСП-СТРУКТУРА, ПРИРОДНА МОВА

Вступ

Розвиток комп'ютеризації, проникнення обчислювальної техніки у всі сфери науки, промисловості, суспільного життя потребує від людини якісного описання та формалізації проблеми створення алгоритмів, розроблення програм аналізу результатів, рекурсивного видозмінювання постановки проблеми та всіх наступних компонентів. У зв'язку із запитами інтелектуалізації ставиться задача забезпечення її достатньо «високоорганізованим інтелектом», що базується на апаратних та програмних засобах, наближених за своїми можливостями до природної мови людини [1-3].

Однією з основних функцій інтелектуальної системи є функція спілкування природною мовою. Традиційна схема аналізу та синтезу тексту на природній мові містить наступні види оброблення: виділення слів та фраз у передредакторі, морфологічний аналіз, синтаксичний аналіз, семантичний аналіз, переведення у внутрішнє зображення, розуміння тексту. Процедури оброблення інформації природної мови утворюють комплекс, що дозволяє як аналізувати, так і синтезувати текст (який називають *лінгвістичним процесором*).

За своєю суттю системи штучного інтелекту (ШтІ) на сучасному етапі характеризуються алгоритмічною універсальністю, яка забезпечується можливістю послідовного багаторазового виконання трьох операцій: переадресації, умовного переходу й довільного пересилання даних з однієї комірки пам'яті в іншу. Отже, інформація та її опрацювання в цифрових мережах та системах розділені, і в системному відношенні величезні потоки інформації перетікають у прямому та зворотному напрямках між пам'яттю та центральним процесором. Всі інтелектуальні властивості привносяться ззовні програмістами, які до найдрібніших деталей продумують алгоритми поведінки машин і мереж. Природно, що продуктивність їх гранично низька, коли йдеться про роботу зі складними моделями знань, тобто з тим матеріалом, який характеризує всі задачі ШтІ.

Отже необхідно, щоб програми і дані не протистояли одне одному, а утворювали спільну нерозчленовану структуру опрацювання даних. Саме в інтегруванні, єдності програм та даних приховані потенційні можливості росту систем штучного інтелекту і, як наслідок, продуктивності цифрових систем та мереж. При цьому програма як така не повинна вводитися в систему - сама система служить породжуючою програму структурою. В кожний вузол мережі вбудовані універсальні функціональні перетворювачі з k -значним кодуванням (УБФП), що можуть гнучко переналагоджуватись па виконання будь-яких необхідних функціональних перетворень. Дані, що оточують універсальний перетворювач, надходять до нього, перетворюються у відповідності із закладеними перетвореннями й розповсюджуються далі комутаційною мережею, змінюючи заодно й саму мережу і породжуючи децентралізоване керування в результаті колективної взаємодії елементів. Кожний вузол семантичної мережі повинен здійснювати аналіз завдання (інтерпретувати його) та зв'язувати один універсальний перетворювач з іншим. Якщо на вході дані, що потребують опрацювання, - вузол ці дані опрацьовує, якщо ні - транслює мережею далі. Хвилі, породжені керуючими командами, поширюються активним середовищем з універсальних перетворювачів, взаємодіють між собою (інтерферують), отже дають можливість створювати нові алгоритми. В мережі залишаються сліди цих інформаційних припливів й відпливів, і наступні хвилі течуть своїм, несподіваним шляхом.

З огляду на викладене вище, стає очевидною актуальність задачі створення основ теорії синтезу надшвидкодійних k -значних структур для мовних систем штучного інтелекту з використанням математичних моделей української мови, що базуються на методах теорії інтелекту і базовій основі - алгебрі скінченних предикатів (АСП).

Серед питань, які потребували глибоких досліджень в роботі, можна виділити наступні:

– застосування формальної теорії (так звана теорія інтелекту на підставі алгебри скінченних предикатів) з метою формалізації подання інформації на різних стадіях обробки мовної інформації;

– методи системної інтеграції знань різних предметних галузей, що представлено в онтологічній формі і шляхи автоматизації обробки мовної інформації;

– методи розробки апаратних засобів підтримки трудомістких процедур лінгвістичного аналізу ПМ-інформації (оброблення інформації природної мови): АСП-структури (на підставі УБФП).

Вищенаведений аналіз головних фрагментів змісту даної роботи показує її безсумнівно актуальність.

1. Алгебологічний аспект моделювання природної мови

1.1 Аналіз проблеми формалізації та програмно-апаратної реалізації мовного інтелектуального інтерфейсу

Запити інтелектуалізації вимагають від ЕОМ здатності розуміння людської мови, достатньої «кмітливості», здатності мислення й творчості. У зв'язку з цим, перед вченими й творцями засобів обчислювальної техніки ставиться задача забезпечення її достатньо «високоорганізованим інтелектом», що базується у першу чергу на програмних, а в другу – на апаратних засобах наближених за своїми можливостями до природної мови людини [1–4].

Людська мова як явище дискретне повинна описуватись засобами дискретної математики, крім цього вона багатозначна (k -значна). Мови програмування та теорії алгоритмів можуть описувати тільки однозначні функції, природна ж мова вимагає формальних засобів для опису багатозначних функцій, чи відношень. Таку мову утворюють логічні числення висловлювань та числення предикатів, які дають можливість описати мову з допомогою апарату рівнянь. Розроблення алгебри скінченних предикатів (АСП) відкриває можливість переходу від алгоритмічного опису інформаційних процесів до опису їх у вигляді рівнянь, а рівняння задають відношення між змінними. Усі змінні в рівнянні рівноправні: будь-які з них можуть виступати як в ролі незалежних, так і в ролі залежних. Алгоритми та програми описують необоротний процес роботи систем від входів до виходів, а рівняння АСП забезпечують у процесі роботи систем й оборотний процес: входні сигнали можна подати на будь-які полюси системи і, відповідно, зняти результуючий сигнал теж можна з будь-яких полюсів. При цьому рівняння дають ту перевагу перед алгоритмами, що можна розрахувати реакцію системи навіть при неповному визначенні входних сигналів, у той час як неповністю розроблений алгоритм є непрацездатним. По-друге, за умов зміни знань про об'єкт система рівнянь АСП, покладених на структуру системи, завжди

готова до використання, а алгоритм часто вимагає докорінної зміни її структури [2,3].

Єдиним недоліком опису системи за допомогою рівнянь АСП є те, що коли число змінних у рівняннях велике (а в рівняннях інтелектуальних процесів воно без сумніву буде великим), то комбінаторно зростає й число способів розподілу входних та вихідних сигналів між полюсами системи. За цих умов практично неможливо створити повний набір алгоритмів, кожний з яких обчислював би реакції системи при певному способі розподілу входних та вихідних сигналів між полюсами системи [1-3].

Механізм природної мови, не дивлячись на видиму легкість користування ним людиною, надзвичайно складний й до того ж слабо вивчений. Отже розроблення систем обробки мовної інформації дуже часто базуються на недостатньо міцному лінгвістичному фундаменті. Багато хто з дослідників, що займалися автоматизацією мовної діяльності, перейшли до поглибленого вивчення, дослідження й моделювання окремих сторін механізму природної мови [1–7].

Зауважимо, що процедури оброблення інформації природної мови утворюють комплекс, що дозволяє як аналізувати, так і синтезувати текст, чому й називають його лінгвістичним процесором. Якщо літерами подано речення чи зв'язаний текст, то процес переходу від знаку до смислу (концепту) умовно можна розбити на три етапи аналізу: морфологічного, синтаксичного та семантичного. Відповідно виділяють три рівні мовного знаку та мовної системи.

На *морфологічному* рівні аналізуються словоформи поза зв'язком з контекстом. Визначаються їх граматичні ознаки (рід, число, відмінок тощо) та лексичне значення. Останнє важливе через те, що граматичні ознаки слів можуть збігатись і розрізнити їх можна лише за лексичним значенням.

На *синтаксичному* рівні встановлюються синтаксичні зв'язки між словами речення, що відображають смислові зв'язки між поняттями. Для встановлення синтаксичних зв'язків використовуються морфологічні відомості про словоформи.

На *семантичному* рівні значення тексту подається на внутрішній формальній мові системи ШтІ. Тобто створюють відображення смислу тексту природної мови в термінах, що зрозумілі машині.

Завершуючи огляд інженерних та наукових розробок у галузі автоматизації мовної діяльності людини відмітимо, що тут існують декілька тенденцій: перехід до постановки вужчих спеціалізованих наукових задач автоматизації обробки текстів та практичної реалізації на обчислювальній техніці у програмному та програмно-апаратному вигляді мовної діяльності людини і вивчення глибинних механізмів природної мови.

Вибір задач, що вирішуються у роботі базується на цих тенденціях. Українське слово як об'єкт

моделювання вибране тому, що воно є найпростішим компонентом нашої мови. При цьому досліджується його структура на різних рівнях: морфологічному, коли з букв утворюються морфи (корінь, префікс, суфікс та закінчення), вплив морфів один на одного, послідовність морфів у слові, яка визначається смислом. Морфи несуть різноманітне граматичне та смислове навантаження. Слово, крім цього, пов'язане з контекстом і всією сумою знань людини складними семантичними залежностями. Усі ці закони лінгвістикою вивчені неповно, а моделювання перебуває у початковому стані. Нижче наводиться аналіз стану досліджень моделювання механізмів природної мови та їх машинного моделювання.

1.2. Аналіз тенденцій моделювання механізмів природної мови

Формальні мови, за допомогою яких формально описуються та реалізуються перелічені вище методи аналізу та синтезу людської мови, використовують різні мови програмування, апарат теорії графів, мову теорії алгоритмів та логічні числення. Тривалі й малоефективні спроби створення систем спілкування на природній мові показали, що всі ці засоби погано пристосовані для задач формального опису людської мови та ще й створюють непереборні труднощі під час її моделювання. Оскільки природна мова лежить у основі інтелектуальної діяльності, то й моделювання ієрархії інших шарів інтелекту наведеними засобами виявиться також неефективним.

Наявність алгебри скінчених предикатів [1-3] відкриває можливість переходу від алгоритмічного опису інформаційних процесів до опису їх у вигляді рівнянь, а рівняння задають відношення між змінними. Усі змінні в рівнянні рівноправні, будь-які з них можуть виступати як у ролі незалежних, так і в ролі залежних. Алгоритми та програми описують необоротний процес роботи систем від входів до виходів, рівняння ж АСП – забезпечують у процесі роботи систем й зворотній процес: вхідні сигнали можна подати на будь-які полюси системи і, відповідно, зняти сигнал теж можна з будь-яких полюсів. При цьому рівняння дають ту перевагу перед алгоритмами, що можна розрахувати реакцію системи навіть при неповній визначеності вхідних сигналів, у той час як неповністю розроблений алгоритм є непрацездатним. По-друге, за умов зміни знань про об'єкт система рівнянь АСП, покладених на структуру системи, завжди готова до використання, а алгоритм часто вимагає докорінної зміни її структури.

Аналіз відомих робіт з моделювання механізму слова показує, що майже в усіх випадках моделі описуються на мовах теорії алгоритмів чи мов програмування. При цьому моделюються, як правило, окремі режими мовної діяльності людини: аналізу, синтезу чи нормалізації мовного повідомлення. Але людина, що володіє природною мовою, крім цих режимів здатна реалізувати практично незліченне число

інших режимів мовної поведінки. Так, вона може відновити текст нечітко вимовленого мовного повідомлення, заповнити зтерту букву в тексті, виправити друкарську помилку. Вона також може відновити значення частини ознак при неповному завданні тексту словоформи та відомих (наявних) значеннях деяких інших ознак.

Для здійснення повноцінної автоматизованої мовної діяльності необхідно моделювати велику кількість різних режимів оброблення слів. З ситуації, що виникла бачиться тільки один вихід: необхідно перейти від алгоритмічного опису окремих видів мовної діяльності до реляційного опису закономірностей мови, тобто відношень, які є основою керування мовної діяльності. Необхідна формальна мова, на якій можна було б записувати ці відношення у вигляді рівнянь. Нарешті, необхідні ефективні машинні методи розв'язку систем таких рівнянь [1-3].

Проте задача сприймання та видавання інформації на природній мові не вирішена й до сьогодення часу оскільки спроби обробки природної мови є копіюванням задач, методів та моделей двозначної логіки й схемотехніки, що були відомі й до створення апарату АСП. Зокрема, математичні моделі природної мови описуються паралельними алгоритмами АСП, а потім послідовно програмно обробляються на процесорах фон-Неймана; усунуто з розгляду власне суть багатозначного (k -значного) характеру мовних повідомлень; виникають проблеми розмірності систем рівнянь АСП та методів їх розв'язування; виникають проблеми з задачами мінімізації в задачах формального синтезу схем [1-3].

2. Побудова багатозначних просторових структур (АСП-структур)

Виходячи з викладеного вище, побудова багатозначної обчислювальної структури чи системи передбачає створення базового набору типових уніфікованих компонент просторового типу, які складають наступний конгломерат засобів: пристрої зовнішнього обміну, що перетворюють двозначні коди на багатозначні; універсальні багатозначні функціональні перетворювачі; комутаційні елементи на кілька напрямків. У зв'язку з орієнтацією на мікроелектронне виконання компонентів мереж усі ці засоби можуть бути класифіковані згідно з наступними ознаками: вид сигналу, що несе інформацію (інформаційну ознаку), галузь застосування, вид схемотехніки та технологія виготовлення.

Теоретичні та експериментальні дослідження й ускладнення, що виникають під час створення систем ШТІ, сприяють висуненню концепції адекватності багатозначної логіки та структур завданню створення систем ШТІ з передбачуваними властивостями та можливостями підвищеного захисту. В цьому контексті, для розкриття шляхів побудови і паралельно-об'ємних k -значних структур, розглянемо концептуальну

структурно-функціональну модель багатозначної комірки [1].

Довільна система ШтІ на системному рівні характеризується набором функцій, що реалізуються нею та функціональними вузлами, які реалізують ці функції, а також інформаційним обміном під час функціональних перетворень. У відповідності до розв'язуваних задач структурно-функціональна комірка узагальненого виду на рівні системного підходу декомпозується на три ієрархічних рівні: функціональний (аналітико-синтетичний); тактичний (аналізаторно-координаційний); стратегічний (координаційний).

Відповідно на функціональному рівні до складу k -значної об'ємно-просторової комірки входять: n -вимірний комутатор сигналів; комплекс порогових пристроїв, дешифратори просторових проміжних ознак та формувачі k -значних функцій.

Комутатор сигналів є керуючим пристроєм входу системи ШтІ, що визначає, з яким вхідним сигналом працює комірка: зовнішнім чи від стратегічного рівня.

Комплекс порогових пристроїв дозволяє здійснювати перетворення неперервних чи дискретних за часом та за рівнем k -значних сигналів (семантичне опрацювання вхідного сигналу системи), а також формування простору проміжних ознак (простору k -значних за суттю характеристичних функцій) як результату семантичного опрацювання. Проміжні ознаки дешифруються надалі у керуючі сигнали вихідного комплексу формувача k -значних функцій, що здійснює аналітичні функціональні перетворення. Результат перетворень на функціональному рівні надходить на вихід комірки, а також для оцінювання з точки зору семантичного змісту, на стратегічний рівень [2,3].

Тактичний рівень об'ємно-просторової комірки реалізується за допомогою аналізаторно-координаційного процесора, до завдань якого входить керування комутатором вхідних сигналів, налагодження порогових пристроїв функціонального рівня, налагодження дешифраторів проміжних ознак з метою вибору виду здійснюваного функціонального перетворення та синхронізації роботи тактичного, функціонального та стратегічного рівнів.

На стратегічному рівні процесор-супервізор дозволяє системі ШтІ здійснювати остаточний семантичний аналіз з участю оператора, обмін даними з оператором, входом та виходом комірки, з базою знань системи ШтІ, а також контролювати в автоматичному та діалоговому режимах процес розв'язання задач на тактичному рівні. Останнє дає можливість прослідкувати покроково процедури реалізації довільного алгоритму на всіх етапах його виконання і таким чином здійснити селекцію та нагромадження у базі знань ефективних алгоритмів з множини інших, менш ефективних.

Площинно-просторова комірка нарощується з допомогою комутаторів на функціональному рівні як за входами, так і за виходами, а також об'єднується з іншими комірками через входи дешифраторів проміжних ознак; на тактичному рівні - через аналізаторно-координаційний процесор; на стратегічному рівні – через процесор-супервізор та базу знань.

Концептуальна структурно-функціональна модель комірки базується на принципі симбіозу (нерозривного зв'язку та взаємодії) дво- та аналого-дискретних засобів опрацювання інформації, тому на стратегічному рівні в ній містяться комплекси перетворювачів форми відображення даних – перетворювачі двозначного коду в k -значний і навпаки. Застосування чи незастосування цих перетворювачів в системі ШтІ визначає, на якому функціональному рівні розв'язуються задачі, в якій логіці (в якому структурному алфавіті) та з якою швидкістю.

Підсумовуючи процес логіко-інтуїтивного емпіричного системного синтезу структурно-функціональної моделі узагальненої комірки для розв'язання задач систем ШтІ зауважимо, що така концепція є апіорною й сформована на основі аналізу широкого спектру власних наукових досліджень та досліджень вчених, що працюють над створенням систем ШтІ, а також тих, які вивчають природний інтелект.

Пропонується перейти від створення спектру часткових типів k -значних структур до єдиної моделі універсальної комірки, яка шляхом зміни налагоджень (але не структури) забезпечує переналагодження під час відтворення інтелектуальних властивостей й перетворень [1].

Висновки

Таким чином, приходимо до розуміння одного з варіантів створення систем ШтІ – це симбіоз двох шляхів: перший – це аналіз, моделювання та синтез мовного інтелектуального інтерфейсу за допомогою засобів k -значної логічної системи, зокрема, алгебри скінченних предикатів (АСП) та k -значних структур і кодування, а другий – це за допомогою лексиграфічних систем та технологій [7].

Теоретичне значення роботи полягає в тому, що в ній запропоновані та розроблені теоретичні основи побудови та проектування моделей ПМ-інформації з онтолого-керованою архітектурою, які застосовуються, в тому числі, для побудови онтологічних баз знань предметних областей, їх системної інтеграції з метою проведення складних міждисциплінарних досліджень та вирішення розкладних задач користувачів, що представляється значущим вкладом у розвиток прикладної теорії інженерії знань.

Вказані теоретичні результати використані при розробці цілого ряду онтологічних систем у науково-технічній сфері та освіті і дозволяють на підставі

принципу симбіозу двох запропонованих підходів будувати тлумачні словники різноманітних ПМ-інформаційних баз даних, баз знань та відповідних просторових структур.

Типові рішення найбільш трудомістких процедур граматичного аналізу (морфологічного аналізу і побудови відповідних АСП-структур) та їх технічні рішення захищено патентами України та РФ [4-6]. Результати даної роботи використані при виконанні науково-дослідних проєктів в Українському мовно-інформаційному фонді НАН України та при підтримці забезпечення навчального процесу для окремих предметних дисциплін у Харківському національному університеті радіоелектроніки.

Список літератури: 1. Четвериков Г.Г. Формалізація методів побудови універсальних k -значних структур мовних систем штучного інтелекту // Доповіді НАН України. – 2001. – №1 (41). – С. 76–79. 2. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії синтезу надшвидкодійних структур мовних систем штучного інтелекту. – К.: Наукова думка (ІЗМН), 1997. – 264 с. 3. Бондаренко М. Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії багатозначних структур і кодування в системах штучного інтелекту. – К.: Наукова думка, 2003. – 336 с. 4. Пат. 14935 А. Україна, МКВ Н 03 К 19/08. Функціональний перетворювач / М.Ф.Бондаренко, З.Д. Коноплянко, Г.Г.Четвериков (Україна). – №96010250; Дата подання 22.01.96; Опубл. 30.06.97, Бюл. №3. – 4 с. 5. Пат. 20462 А. Україна, МКВ Н 03 К 19/02. Двовходовий багатозначний логічний елемент / М.Ф.Бондаренко, З.Д. Коноплянко, Г.Г.Четвериков (Україна). – №97031289; Дата подання 20.03.97; Опубл. 15.07.97, Бюл. №3. – 4 с. 6. Пат. 2147789 РФ, МКВ Н 03 К 19/02, Н 03 М 1/00. Функціональний преобразователь с многозначным кодированием / М.Ф.Бондаренко, З.Д. Коноплянко, Г.Г.Четвериков (Україна). – №97101717/09; Заявл. 04.02.97; Опубл. 24.04.2000, Бюл. №11. – 6 с. 7. Четвериков Г.Г. Алгебрологічні та лексикографічні аспекти моделювання природної мови // Бионика интеллекта. – 2014. – №2 (83). – С. 3–14.

Поступила до редколегії 15.05.2016

УДК 519.62

Универсальный многозначный функциональный преобразователь как базовая ячейка семантической сети интеллектуальной системы / Г.Г. Четвериков, А.С. Пузик, В.В. Курасова, И.К. Божко // Бионика интеллекта: научн.-техн. журнал. – 2016. – № 1(86). – С. 164–168.

Статья посвящена вопросу организации данных и эффективного выбора программных и технических средств для разработки языково-информационных интерфейсов и проєктов. Дан анализ тенденций моделирования механизмов естественных языков. В статье также изложены принципы и методы построения многоходовых универсальных k -значных пространственных структур языковых систем искусственного интеллекта.

Библиогр.: 7 назв.

UDC 519.62

Universal polysemantic functional converter as basic cell of semantic network of intelligence system / G.G. Chetverikov, O.S. Puzik, V.V. Kurasova, I.K. Bogko // Bionics of Intelligence: Sci. Mag. – 2016. – № 1(86). – P. 164–168.

The paper is devoted to the problems of data structure organization and effective choice of software and technical solution for developing of the linguistic-informational interfaces and projects. The trend analysis of modeling natural language was imposed. The principles and methods of designing universal multiple-valued space structures of artificial intelligence were set out in this paper.

Ref.: 7 items.