



Г.Г. Четвериков

ХНУРЕ, м. Харків, Україна, chetvergg@kture.kharkov.ua

**АЛГЕБРОЛОГІЧНІ ТА ЛЕКСИКОГРАФІЧНІ АСПЕКТИ
МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ**

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

ПРИРОДНА МОВА, ЛОГІКА, СКІНЧЕННИЙ ПРЕДИКАТ, АЛГЕБРА СКІНЧЕННИХ ПРЕДИКАТІВ, БАГАТОЗНАЧНА СТРУКТУРА, АСП-СТРУКТУРА, ПРИКЛАДНА ЛІНГВІСТИКА, КОРПУСНА ЛІНГВІСТИКА, ЛІНГВІСТИЧНИЙ КОРПУС

Вступ

Для сучасного етапу розвитку науково-технічного прогресу характерне переваження неформалізованих чи погано формалізованих проблем, алгоритмічне рішення яких чи то не існує, чи то не може бути отримане на наявних сьогодні машинах. Орієнтація на досягнення високорівневої технології обробки інформації, зокрема природномовної, проявляється у спробах реалізувати цю обробку на фон-нейманівських комп'ютерах у галузі штучного інтелекту (ШТІ) [1–3].

Підсумовуючи процес аналізу перспективних шляхів створення надшвидкодійних структур для вирішення задач створення систем ШТІ відмітимо, що така концепція є апріорною й сформована на основі аналізу широкого спектру власних наукових досліджень та досліджень вчених, що працюють над створенням систем ШТІ, а також тих, що вивчають природний інтелект [1–12].

Сьогодні не викликає сумніву той факт, що інтелектуальні інформаційні технології та відповідні комп'ютерні системи, зокрема, онтолого-керовані інформаційні системи (ОКІС), будуть і далі інтенсивно розвиватися в напрямках, що відповідають найбільш істотним ознакам розумової діяльності людини, в тому числі розпізнавання, перетворення і розуміння знакових систем (включаючи природномовні – ПМ).

Результатом зазначених комплексних завдань є генерація смислу, закладеного в аналізовану знакову систему. При цьому під смислом розуміється те, що робить знакову систему текстом, а останній – джерелом знань. Розвиток комп'ютерної лінгвістики та онтологічного інжинірингу пов'язаний з удосконаленням методів комп'ютерної обробки ПМ-інформації та предметних знань. Можна відмітити, що добре опрацьовані сучасні методи та засоби обробки ПМ-інформації для вирішення прикладних задач у вузькоспеціалізованих предметних галузях недостатні для вирішення комплексних задач, пов'язаних з формально-логічним поданням ПМ-інформації, добуванням предметних знань, з

подальшим використанням у довільних предметних галузях. Це підтверджує актуальність і важливість розробки нових наукових методів і підходів до автоматичного аналізу та глибинно-семантичного розуміння ПМ-інформації, її формалізованого подання в рамках відповідної теорії, автоматичного виявлення і вилучення нових знань, технологій і, врешті – відповідних інструментальних засобів автоматизованої побудови баз даних та знань предметних галузей.

Отже сучасні, нові вимоги до технології обробки інформації можна звести до таких двох моментів: необхідність вирішення проблем, що погано формалізуються, та наявністю кінцевого користувача, який за своєю професійною та творчою діяльністю не є програмістом.

В роботі з усього розмаїття предметних знань особливий статус надається мовним знанням і відповідним онтологіям, побудованим на їх основі, зокрема всесвітньо відомій WordNet та іншим. Виконано достатньо глибокий аналіз вказаних онтологій, зокрема, дано підставу для розробки оригінальної лінгвістичної онтології – мовно-онтологічної картини світу, яка орієнтована, в першу чергу, на українську мову.

Аналіз природномовних текстів включає формування понятійних структур, які забезпечують автоматичне вилучення знань із текстів та їх прагматичну інтерпретацію в термінах прикладних задач. При цьому текст розглядається як об'єкт різних рівнів аналізу: як знакова система, як граматична система, система знань про світ і, зрештою, як джерело вихідних даних для вирішення прикладних задач. При цьому кожен рівень має свої особливості, свої засоби вираження, а отже, допускає присутність специфічних методів обробки.

Усі відомі системи обробки природномовних текстів, як правило, містять (чи можуть приєднувати відомі) різноманітні інструментальні засоби, які суттєво спрощують розробникам чи користувачам досягнення кінцевої мети. Важливо, що автор статті також пропонує свій, так званий

інструментальний комплекс онтологічного призначення, який реалізує дві самостійні інтегровані технології: алгебологічну та лексикографічну.

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

- задачі створення та розвитку української науково-технічної термінології;

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

- методи та способи автоматизованої побудови онтологічних баз знань предметних галузей, включаючи розробку інструментальних (лексикографічних) засобів для обробки великих об'ємів текстової 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: АСП-структури.

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

1. Алгебологічний аспект

1.1. Огляд розробок мовного інтерфейсу систем штучного інтелекту

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].

Задача інтелектуалізації ЕОМ не передбачає простого й швидкого вирішення. Виникають суттєві складнощі принципового характеру. Через ці складнощі не дають бажаних результатів роботи з автоматизації програмування, створення високоякісного перекладу з однієї мови на іншу, побудови систем для розпізнавання достатньо складних об'єктів, формування здатності машин зрозуміти побачене й почуте, спрямоване на рішення аналогічних питань тощо. Головна перепона, на яку нашовахуються всі спроби радикально підвищити інтелектуальні здібності ЕОМ, полягає у недосконалому машинних мов. Вони значно поступаються щодо промовистості природним мовам. Роста число прихильників точки зору, що створення «машинного інтелекту» вимагає вивчення й моделювання людського інтелекту та, у першу

чергу, природної мови [1–3], що лежить у його основі. Стає все очевиднішим, що при подальшому вдосконаленні машинних мов необхідно використовувати ширше організацію людської мови. У цьому зв'язку виникає необхідність створення математичних моделей лінгвістичних зв'язків (законів мови), що лежать в основі мовної діяльності людини. Важливо математично описати ієрархічну структуру природної мови, що проявляється під час словесного опрацювання українських текстів, зображених у графічній чи акустичній формі.

Практичні досягнення автоматизації мовної діяльності людини вражають бідні у порівнянні з важливістю проблеми та величезними зусиллями, що затрачені упродовж третини віку на їх реалізацію. Дуже мало створено програм лінгвістичного характеру, що виявилися придатними для вирішення реальних задач автоматизації праці людини. Причини такого положення для багатьох розробників тепер вже зрозумілі. 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 [1–12].

Як видно з блок-схеми однією з основних функцій інтелектуальної системи є функція підсистеми спілкування на природній мові. Традиційна схема аналізу та синтезу тексту на природній мові включає наступні види оброблення: виділення слів та фраз у передредакторі, морфологічний аналіз, синтаксичний аналіз, семантичний аналіз, перекладу у внутрішнє зображення, розуміння тексту (рис. 1 та рис. 2).

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 вузких спеціалізованих наукових задач автоматизації обробки текстів та практичної реалізації на ЕОМ 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 вибрано 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,

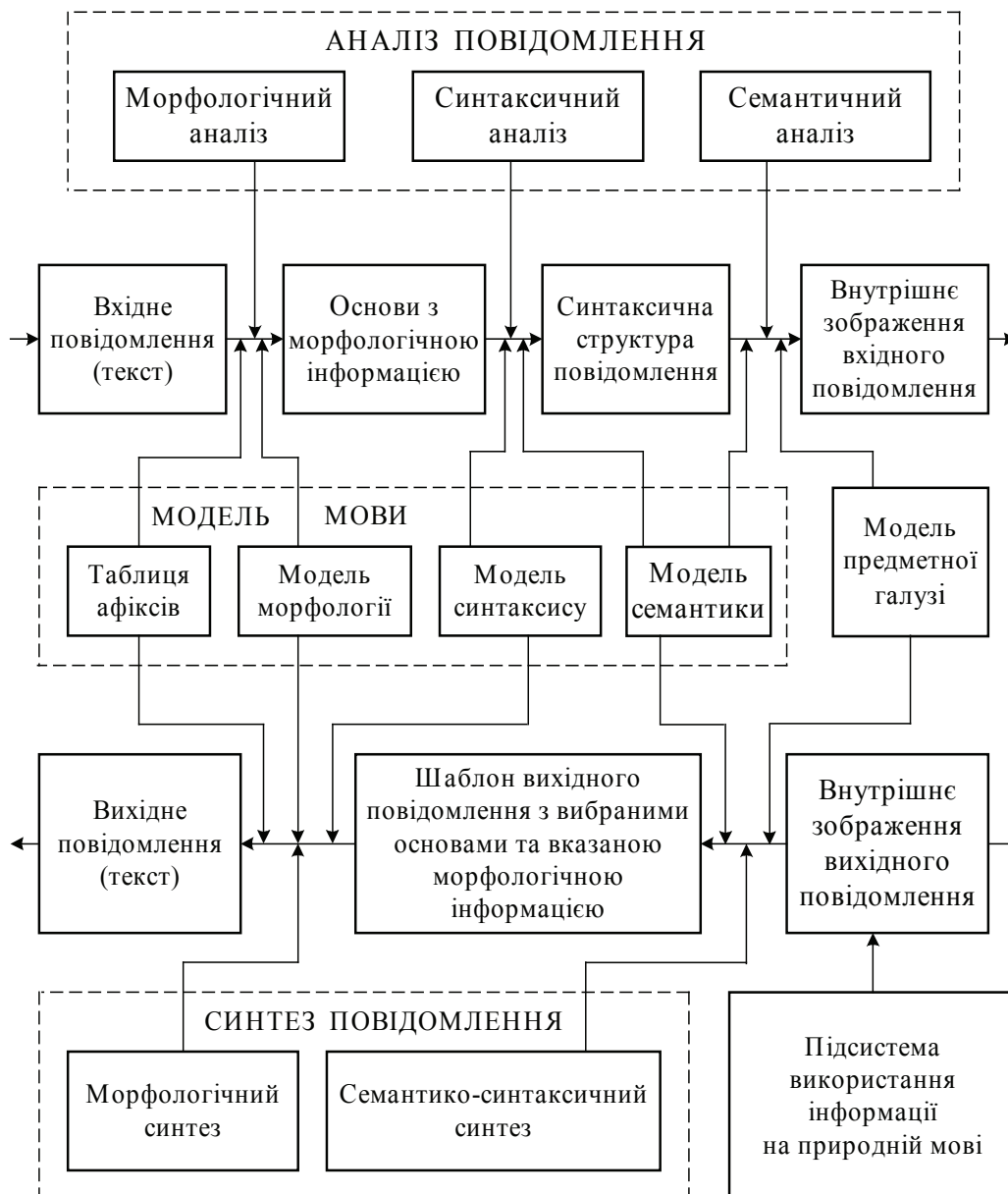


Рис. 1. Структура процедур опрацювання інформації на природній мові

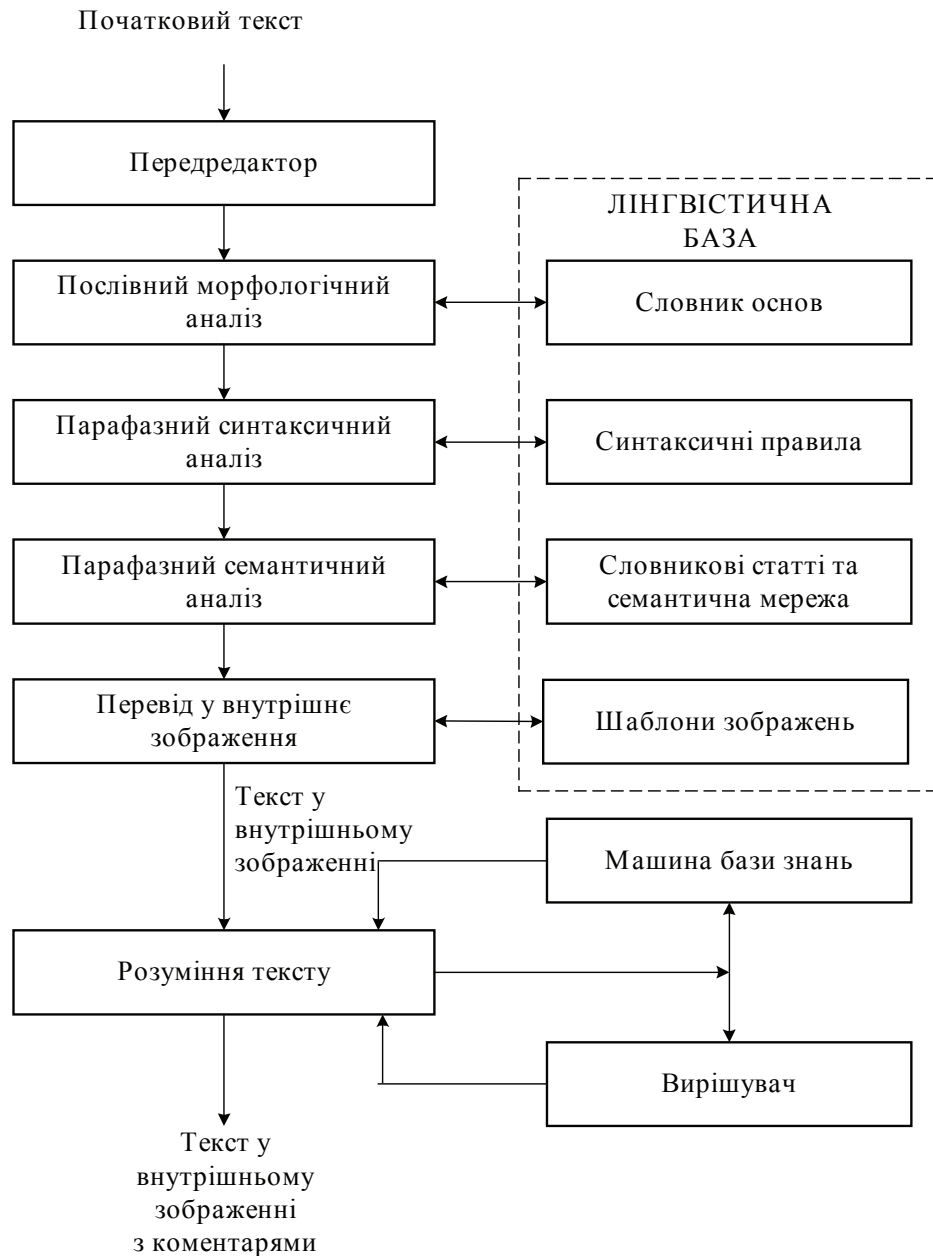


Рис. 2. Схема аналізу тексту

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.

Людська мова, як явище дискретне, очевидно, повинна описуватись засобами дискретної математики. Крім цього, природна мова багатозначна, що проявляється у вигляді омографічності слів, тобто неоднозначності їх змісту. Мови ж програмування та теорії алгоритмів можуть описувати тільки однозначні функції. Природна ж мова вимагає формальних засобів для опису багатозначних функцій, іншими словами, — відношень. Таку мову

утворюють логічні числення висловлювань та числення предикатів, які дають можливість описати мову за допомогою апарату відношень (рівнянь). Та щоб розв'язувати такі рівняння, необхідно довести числення предикатів до рівня алгебричної системи. Коли ж звернутись до апарату багатозначної (k -значної) логіки [2, 3], то в ній, на сьогоднішній день, не існує вчення про розв'язки рівнянь, а відповідний розвиток у цьому напрямку веде до алгебри скінченних предикатів [1]. Дійсно, для того щоб записувати найзагальніші рівняння k -значної логіки, у правій частині немає потреби ставити довільні формули, достатньо писати константи. По-друге, необов'язково використовувати всі k констант, достатньо взяти всього два знаки: 0 та 1, але як тільки це зроблено, то виникає поняття скінченного предиката й моделювання природної мови приходить до алгебри скінченних предикатів.

Наявність алгебри скінченних предикатів (АСП) [1–3, 11] 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 витерту 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 [1–6].

Проте задача сприймання та видавання інформації на природній мові не вирішена й до сьогоднішнього часу оскільки спроби обробки природної мови є копіюванням задач, методів та моделей дво-значної 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].

1.3. Задачі створення та розвитку української науково-технічної термінології

Однією з сучасних задач в Україні є вирішення завдання створення, розвитку й відродження української науково-технічної термінології. Українську мову офіційно забороняли (укази Петра Першого, Валуєвський циркуляр, Емський указ), оголошували придуманою німцями, вважали діалектом російської чи польської мов, дивилися як на «хлопську», призначали її тільки для «домашнього вжитку». Її свідомо принижували, не допускали у сферу науки, техніки, на фабрики, заводи, до армії, в спорт – до тих сфер, що важливі для діяльності та повноцінного життя нації [2]. Разом з тим, фонетична розкіш, лексичне та фразеологічне багатство, синтаксична гнучкість, величезні словотворчі можливості дістали загальне визнання у Парижі (1934 р.) на всесвітньому конкурсі краси мов, де українська зайняла третє місце (на першому – французька, на другому – перська).

Найголовніше – виходити з тези: українська мова має бути мовою державною, інакшого шляху в народі, який себе поважає, нема. Українська Академія наук до її погрому (1933 рік) видала 35 українських словників з окремих галузей науки і техніки [2, 7–10].

Без перебільшення українська мова була і є однією з найбагатших у світі, зокрема й серед слов'янських. Наші академіки-словникарі не мали жодної потреби запозичувати в чужинців і склали науково-технічні словники з мовних скарбів українського народу. Українське наукове й технічне назовництво є найбільш слов'янське запоміж слов'янських, крім чеського [2].

В українському алфавіті порівняно з російським є суттєві відмінності [123]. Так, у російському алфавіті відсутні букви **ї**, **є**, **г** та апостроф; українській букві **є** відповідають **э** та **е** (електрика – электричество, театр – театр); українській букві **и** в російській графіці відповідає буква **ы** (син – сын, рити – рыть).

Наявна в російському алфавіті та відсутня в українському буква **ё** передає звукосполучення [jo] або м'якість попереднього приголосного та голошний [o]: ёлка – [jo'лка], ёжик – [jo'жик], тётка – [t'отка].

Буква «г» в українській фонетиці позначає гортанний звук, для задньо-язикового використовується буква «ґ»: грати (на бандурі); (міцні) грати, а в російській мові наявний лише задньоязиковий [г]: играть, голова, игнорировать, який передається буквою «г».

Як суттєву відміну слід зауважити тверду вимову приголосних перед буквою «е» в українській мові проте м'яку в російській: день [д'ен'] – день [д'ен'], тепер [тепе'р] – теперь [т'эп'р'].

Замість апострофа, властивого українській графіці, у російській – використовується твердий знак «ъ»: з'їзд – съезд, об'явити – объявить (але в'ялий – вялый, кров'ю – кровью оскільки в російській мові губні приголосні є м'якими).

Підкреслимо ще одну особливість, яка розмежує українську мову від російської з її прийменником «у» або «в» [2]. У російській мові прийменники «у» та «в» до того ж самого кореня дають слова не однакового, а різного значення, тоді як в українській мові такі слова мають одне значення й від зміни прийменника «у» на «в» й навпаки їх значення залишається те ж саме. Наприклад, по українськи: улетіти і влетіти – куди? – в клітку, в кімнату; уїхати, в'їхати – куди? – в двір; внести й унести – куди? – в хату тощо. А в російській мові справа стоїть цілком інакше, наприклад, улететь – откуда? – з клітки, з кімнати, а влететь – куда? – в клітку, в кімнату; уехать – откуда? – з двору, въехать – куда? – в двір; унести – откуда? – з хати, а внести – куда? – в хату тощо. Отже, для прикладу, в словах «Україна» та «Вкраїна» початкове «у» і «в» не корінні звуки, а прийменники до кореня «КРА» і можуть змінювати одне другого. Український народ належить до Адріатичної або Динарійської слов'янської раси, тому ці слова з спільнослов'янським коренем «КРА» є і в сербів, словаків, боснійців, словенів, хорватів, оскільки вони належать до тієї ж Адріатичної слов'янської раси. Отже, слова край, країна, Вкраїна і Україна – мають своїм коренем старослов'янське слово «КРА» і звідси пішло слово «УКРАЇНА», а не від «окраїна», як це твердять деякі російські ідеологи. Але, з іншого боку, значення слів «україна» та «окраїна» абсолютно різні й отожнювати їх неможливо.

Подамо наглядні приклади: вбити-убити...обити, впасти-упасти...опасти, вбігти-убігти...обігти, вскубти-ускубти...оскубти, вгнути-угнути...огнути, вмити-умити...омити, втікати-утікати...отікати, всипати-усипати...осипати, вкраїна-україна...окраїна, впадати-упадати...опадати, вкопати-укопати ... окопати, вжити-ужити...ожити, впустити-упустити ...опустити, встригти-устригити...остригити, вкряяти-укряяти...окряяти тощо.

Кожна мова позичає собі слова з інших, але розвинені мови запозичують лише нові наукові та вузько-технічні слова. А російська – мала можливість позичати звичайні, щоденні: її лексика має багато слів та виразів німецьких, французьких та англійських. Так російська взяла з німецької, наприклад: абрикос (жерделя), бархат (оксамит), брендмейстер (пожежник), брюки (штани), брухт (покидьки), бунт (повстання), бурт (купа), вакса

(мастило), вал (окіп), ванна (купіль), вахта (варта), гауптвахта (вартівня), горн (ріжок), гастроль (виступ), глетчер (льодовик), глазурь (полива), камердинер (покойовий), кант (облямівка), краги (холявки), кран (затичка), лозунг (гасло), лакей (прислужник), маклер (посередник), мундштук (цибух), орден (відзнака), офіцер (старшина), обер-, унтер-офіцер (над-, підстаршина), пакгауз (комора), парта (лавка), патронташ (ладунка), пилигрим (прочанин), пушка (гармата), ранец (наплечник), траур (жалоба), фальшивий (підроблений), флюс (пухлина), флаг (прапор), флагшток (щогла), фон (тло), форпост (передова варта), фрахт (перевізні), футляр (коробка), шанцы (окопи), шахта (копальня), швейцар (двірник), шельма (крутій), шина (обруч), шкатулка (коробочка), штемпель (тавро), штиблети (черевики), штука (одиниця), штурм (наступ), штурман (стерновий), шулер (шахрай) та десятки тисяч подібних.

Крім того, російська мова взяла десятки тисяч французьких слів щоденного вжитку, наприклад [2]: азарт (запал), акомпанемент (супровід), ампула (роль), ангажемент (запрошення), анонс (оповістка), антракт (перерва), аплодисменты (оплески), апломб (самовпевненість), армія (військо), артиллерист (гармаш), артист (митець), ассортимент (набір), атака (напад), афиша (оголошення), балюстрада (поруччя), баланс (рівновага), бандаж (перев'язь), барьер (перепона), бассейн (сточисьце, водойма), бивуак (табір), блондин (русявий), бокал (келих), бонна (нянька), бордюр (окрайка), ботфорти (чоботи), брошюра (книжечка), брюнет (смаглявий), бульйон (юшка), бутылка (пляшка), бюджет (кошторис), бюст (погруддя), визит (відвідини), вуаль (серпанок), галимаття (нісенітниця), гарантія (запорука), гардина (завіса), гарнизон (залога), гувернантка (вихователька), девиз (гасло), декаданс (занепад), департамент (відділ), деталь (подробиця, частина), диссонанс (розлад), експлуататор (визискувач), експорт (вивів), елегантний (чепурний), етажерка (полиця), етикетка (наліпка), жетон (значок), интерес (зацікавлення), интимный (близький), кавалерія (кіннота), канделябр (світильник), каприз (примха, вередування), карьер (каменище), квартира (помешкання), квитанція (посвідка), колонист (поселенець), команда (наказ), коммероант (торгівець), компанія (товариство), компот (узвар), компрометація (знеславлення), коммунікат (повідомлення), константировать (стверджувати), контур (обрис), конфеты (цукерки), костюм (убрання), кошмар (жах), кулінар (кухар), куліси (лаштунки), курьер (посланець), магазин (крамниця, склад), манеры (поведінка), манкировать (ухилятися), манто (накидка), марш (похід), маршрут (розклад подорожі), медаль (відзнака), мемуари (спогади), модистка (швачка), мотив (спонука),

оранжерея (теплиця), ордер (наказ), пансион (бурса), портмоне (гаманець), пейзаж (краєвид), пьедестал (підніжжя), персона (особа), поза (постава), престиж (повага), приз (нагорода), район (округа), рапорт (звіт), режим (лад), результат (наслідок), реноме (слава, розголос), репрезентабельный (показний), репрессия (утиск), ресторан (харчевня), ресурси (засоби, запас), секрет (таємниця), серьезный (поважний), сигнал (знак), спектакль (вистава), стаж (досвід), табуретка (стілець), тираж (наклад), трактир (корчма), фаворитка (улюблениця), фамильярность (панібратство), шеф (начальник), шофер (водій) тощо.

Порівняльний аналіз морфологічної та деривативної системи української та російської мов [2] показує наявність не дуже значної різниці між ними. Проте, спільність задач, що їх вирішує будь-яка мова та частково спільні витоки, як вказувалось вище, дозволяють використати розроблені в теорії інтелекту для російської мови основні математичні моделі словозміни та словотворення в достатньою ефективністю й для української мови.

Як російській мові, так й українській властиве яскраве демутивно-емоційне стилістичне зафарбування. Ця ж особливість зберігається й для словотворчого процесу даних мов. Найпродуктивніше в ній проявляється суфіксний спосіб словотворення для іменників та прикметників, у меншій мірі – для дієслів та числівників. Не дивно, що в іменників та прикметників української мови така різноманітна система суфіксних морфем, тоді як у дієслів різноманітніша система префіксної деривації.

Отже, проблемною задачею на сучасному етапі є вирішення завдання створення, розвитку й відродження української науково-технічної термінології, що в силу історичних обставин, є у значній мірі зрусифікованою, а також спотвореною навіть на рівні елементарної фонетики. *Мова ж є однією з головних прикмет нації, а культ нації – другим після релігії джерелом сили народу.* Розвиток та широке застосування рідної мови є тим важелем, що формує й породжує національну культуру і науку, виробництво та економіку. Нав'язування і використання чужомовної термінології веде до несприйняття суспільством новітніх технологій, занепаду культури та розпаду державності. Все це викликає потребу переосмислення нагромаджених фактів, зосередження зусиль на створенні нового теоретичного каркасу, у рамках якого отримали б розвиток та пояснення всі наявні знання у галузі інтелектуалізації ЕОМ на користь України.

1.4. Шляхи автоматизації обробки мовної інформації

У даний час в системах штучного інтелекту машинний словник та комплекс програм (тезауруси) використовуються, як правило, для виконання будь-якої однієї функції: у вигляді аналізу деякого

повідомлення, його синтезу, виправлення помилок, нормалізація тощо. Розрізнене моделювання несхожих функцій породжує серйозні труднощі зі стиковкою процедур й знань у рамках підсистеми інтелектуального інтерфейсу (ІнІн) повної системи ШТІ, що виконує як аналіз так і синтез повідомлень на ПМ. Як правило, стає неможливим використання процедур однієї системи у складі інших систем. Вказані обставини приваблюють увагу до створення комплексних моделей, які реалізують здібності людини до сприйняття та видавання інформації у вигляді текстів на ПМ. Остане можливо завдяки запропонованим та модифікованим у роботі оборотним k -значним перемикальним ланцюгам просторового типу.

Отже, місце даного дослідження, щодо проблеми у цілому є наступне: представити укрупнену структуру ІнІн та виявити підсистеми (методи, структури, алгоритми), розробка яких забезпечить її функціонування.

Один із таких підходів до побудови подібних моделей передбачає виявлення та опис лінгвістичних відношень на фонетичному, морфологічному, синтаксичному та семантичному рівнях. Опис таких відношень робимо у вигляді формул (рівнянь) диз'юнктивно – кон'юнктивно алгебри предикатів. Тоді рішення конкретних задач аналізу, синтезу текстів та інші інтерпретуємо як підставлення відомих значень змінних до рівнянь і визначення невідомих [1–6, 11, 12].

Таким чином, будемо природну мову уявляти як математичний об'єкт (якась алгебра). При цьому смисл (зміст) думок можна висловити реченнями і текстами, що з точки зору їх математичної природи будемо розглядати як предикати. Наша відправна точка у даних міркуваннях: думки – це предикати. Таким чином, кожне речення виражає деяку функцію з двійковим значенням, тобто задає деякий предикат $P(x) = \lambda$. Незалежною змінною x даної функції буде змінна ситуація, залежною – істинна змінна λ . Після підставлення замість змінної x конкретної постійної ситуації $x = a$ задане речення стає істинним ($\lambda = 1$) або хибним ($\lambda = 0$). Це залежить від того чи відповідає чи ні зміст цього речення ситуації a , до якої воно віднесено. А що означає змінна ситуація x ? Вона повинна удавати з себе набір $x = (x_1, x_2, \dots, x_m)$ предметних змінних x_1, x_2, \dots, x_m , як запропоновано вище. Будь-яка постійна ситуація $x = a$ повинна бути набором $a = (a_1, a_2, \dots, a_m)$ деяких предметів $x_1 = a_1, x_2 = a_2, \dots, x_m = a_m$. Таким чином кожне речення повинно висловлювати деякий предикат $P(x_1, x_2, \dots, x_m) = \lambda$, що представляє залежність істинної змінної від предметних змінних x_1, x_2, \dots, x_m . Проте будь-яке речення за природно-мовною формою відрізняється від математичної формули тим, що виражає не усю функцію $P(x_1, x_2, \dots, x_m)$, а тільки її ім'я P . І це так, бо кожен

раз, коли людина перетворює те чи інше речення у відповідну до нього думку вона добудовує його до предикату. При цьому вона додає до нього (як до ім'я предикату) відсутні предметні змінні. Тільки після цього речення стає доступним для розуміння. Та, навпаки, перетворюючи деяку думку у речення, людина виключає з неї предметні змінні, що дозволяє передавати іншим людям не саму думку, а лише її ім'я. Продемонструємо сказане на прикладі, у якому зробимо доповнення речення предметними змінними.

Нехай маємо деяке речення, наприклад:

«На майдані стоїть церква». (1)

Визначимо число предметів про які йде мова у реченні (1). Очевидно, що таких предметів два: один характеризується словом «майдан», а інший – словом «церква». З множини Q виберемо будь-які дві предметні змінні, наприклад q_1 і q_2 , та введемо їх до речення (1) після зазначених слів. В результаті отримаємо наступне твердження:

«На майдані q_1 стоїть церква q_2 ». (2)

Воно висловлює зараз не лише ім'я предиката, як початкове речення (1), але й сам предикат з його аргументами q_1 і q_2 . Речення, що доповнено предметними змінними, будемо називати висловленням. Саме так називають будь-яке твердження в математичній логіці, що виражається предикатом.

Додамо до речення (1) ще одне

«Поряд з нею квітне бузок». (3)

що утворює спільний текст. У реченні (3) мова йде також про два предмета. Перший з них зазначено займенником «нею», другий – іменником «бузок». Із контексту (контекстом для речення (3) є речення (1)) очевидно, що слово «нею» може бути замінювачем слова «церква», що стосується предмета q_2 , який фігурує у реченні (1). Слово «бузок» вводить третій предмет, що відрізняється від двох перших. У якості нової предметної змінної беремо q_3 . В наслідок отримаємо висловлення

«Поряд з церквою q_2 квітне бузок q_3 ».

Зазначимо, що речення «Поряд з церквою квітне бузок» стає двозначним, якщо його розглядати поза контекстом та без змінних q_2 і q_3 . У даному випадку не можна однозначно визначити про яку конкретно церкву йде мова: про ту ж саму, що й в реченні (1) (тобто предмет q_2) або про деяку іншу (наприклад, предмет q_3). Ще більш багатозначним сприймається речення (3). У ньому слово «нею» можна віднести до будь-якого предмету, а не лише до церкви. Але якщо ввести предметні змінні, то подібні неоднозначності повністю можна усунути. Цей факт наявно демонструє необхідність доповнення речень предметними змінними для можливості їх однозначного розуміння. Однак наш перехід від речень до відповідних висловлювань повністю опирається на інтуїцію людини, що володіє українською

мовою. Формальний перехід також можливий: лише на основі аналізу тексту речення і навколишнього контексту без звернення до їх смислу. Ця задача дуже складна і очікує свого розглядання. Проте без її рішення неможлива автоматизація процесу розуміння текстів природної мови та побудова інтелектуального інтерфейсу взагалі [1–3, 11].

2. Лексикографічний аспект

2.1. Лексикографічні засоби моделювання природної мови

Мова – основа мислення, інтелектуальна діяльність без мови неможлива. Прояви мови вкрай складні і різноманітні, їх вивчають багато наукових дисциплін – морфологія, синтаксис, лексикологія, семіотика, стилістика та ін. вибираємо Для математичного опису вибираємо морфологію мови, тобто учення про граматику окремо взятого слова. Слово – це “атом” мови, щось первинне і найпростіше в мові. І проте, навіть окремо взяте слово надзвичайно складний об'єкт для математичного опису.

Лінгвістичні праці містять неформалізовані описи морфологічних явищ, що складно перекласти на математичну мову. Людина здатна розуміти їх «з півслова». Крім того, людина, що не приступала до вивчення морфології, фактично уже володіє мовою, морфологічні явища їй знайомі з власної мовної практики. Математичні описи морфології мови, які адресовані обчислювальній машині, повинні бути зрозумілі їй. Однак рівень «інтелекту» обчислювальної машини в порівнянні з людським надзвичайно малий. Обчислювальна машина в сенсі мовних здібностей – “чиста дошка”, ніякими мовними навичками вона не володіє. Це веде до того, що неформалізовані описи, що містяться в посібниках з морфології мови, обчислювальній машині зовсім незрозумілі, недоступні їй. До цих описів повинно бути додано ще дуже багато чого, перш ніж машина зможе їх сприйняти й освоїти [7].

Один із розділів, що вивчає застосування математичних моделей для опису лінгвістичних закономірностей являється комп'ютерна та прикладна лінгвістика. Лінгвістику в широкому сенсі слова (пізнання мови і передача результатів цього пізнання іншим людям) можна розділити на наступні напрямки:

– теоретична лінгвістика (наукова), передбачає побудову лінгвістичних теорій, розглядає всі аспекти і проблеми, пов'язані з мовою, склад і вживання, загальні закономірності пристосування та розвиток мови;

– прикладна лінгвістика, спеціалізується на вирішенні практичних завдань, пов'язаних з вивченням мови, а також на практичному використанні лінгвістичної теорії в інших областях. *Наприклад*, методика навчання рідної та нерідної мови, лексикографія, переклад, дешифрування,

орфографія, транслітерація, розробка термінології і тому подібне;

– практична лінгвістика, є сферою, де реально проводяться лінгвістичні експерименти, що мають на меті верифікацію положень теоретичної лінгвістики та перевірку ефективності продуктів, що створюються прикладною лінгвістикою. Наприклад, навчання дітей рідній мові, вивчення іноземної мови, переклад, викладання рідної та іноземної мови, літературне редагування, коректура, практична логопедія, побутова і художня словотворчість, мовна політика, створення нових писемностей і навчання грамоті, і тому подібне [7].

Основними напрямками прикладної лінгвістики, які пов'язані з практичними впровадженнями, є наступні:

- 1) комп'ютерна лінгвістика;
- 2) лінгвістична експертиза;
- 3) наука про впорядкування та стандартизацію науково-технічної термінології.

Серед важливих напрямів прикладних робіт у галузі комп'ютерної лінгвістики слід назвати корпусну лінгвістику (створення та використання електронних корпусів текстів), машинний переклад, системи пошуку та класифікації, автоматичне розпізнавання символів (системи автоматичної корекції правопису, такі як автокоректори, спелчекери і так далі); автоматичне реферування текстів (системи авто реферування); створення електронних словників, тезаурусів, онтологій (електронні одномовні та перекладні словники загального вжитку: тлумачні, орфографічні, термінологічні тощо); автоматичне розпізнавання мови (системи акустичного розпізнавання і синтезу мовлення); оптичне розпізнавання символів (системи оптичного розпізнавання); логічний аналіз текстів (системи логічного аналізу змісту текстів та локалізації знань).

Розглянемо деякі з основних напрямів комп'ютерної лінгвістики більш детально і з'ясуємо, яке місце посідають у системах обробки природної мови граматичні (морфологічні) підсистеми.

Корпусна лінгвістика займається розробкою, створенням і використанням текстових (лінгвістичних) корпусів. Під лінгвістичним корпусом текстів розуміємо великий, структурований, розмічений, філологічно компетентний масив мовних даних, представлений в електронному вигляді й призначений для розв'язання різних лінгвістичних завдань [7- 10]. Лінгвістичний корпус текстів доповнюється системою керування даними – універсальними програмними засобами для пошуку різноманітної лінгвістичної інформації та зручного представлення її користувачеві [7]. Важливим при створенні лінгвістичних корпусів вважається поняття розмітки. Головною задачею розмітки є відмежування корпусів текстів від звичайної

сукупності електронних текстів за допомогою мітки (індикатора), яка буває різних типів:

- а) зовнішня (зазвичай, це елементи бібліографічного опису: видання, рік, автор, тощо);
- б) структурна (описують структуру тексту: розділ, абзац, речення ...);
- в) лінгвістична (морфологічна (включає не тільки ознаки частини мови, але й ознаки граматичних категорій, властивих даній частині мови);
 - 1) синтаксична;
 - 2) семантична;
 - 3) анафорична;
 - 4) просодична (мітки, що описують наголос та інтонацію)).

Серед різних видів розмітки основною є морфологічна розмітка. Необхідними для її виконання є морфологічний аналізатор текстових слівформ на основі граматичного словника та алгоритми автоматичного усунення граматичної омонімії. Виконання цих етапів значно полегшує подальшу обробку текстів.

Машинний переклад – технологія перекладу текстів з однієї природної мови на іншу за допомогою спеціальної комп'ютерної програми. Під час машинного перекладу можна зустрітись з такими проблемами:

- проблема створення великих словників для систем;
- проблема навчання системи розпізнавання стійких виразів;
- проблема запису всіх правил побудови речення і перекладу у вигляді програми.

Щоб отримати текст для перекладу (текст-оригінал), слід почати з аналізу тексту. Аналіз тексту необхідно розбити на чотири модулі: передморфологічний, морфологічний, контекстний, синтаксичний. Після аналізу застосовуються алгоритми перекладу. Для перекладу використовуються автоматичний багатомовний перекладний словник та перекладний словник словосполучень. Наступний етап – це синтез вихідних слівформ та речення в цілому на вихідній мові. Синтаксис кожної мови має свої особливості. Наприклад, те, що в українському реченні - підмет, в іншій мові може (або повинно) бути виражено доповненням, а доповнення, навпаки, має перетворитися у підмет; те, що в одній мові позначається групою слів, перекладається на іншій всього одним словом. У зв'язку з цим в машинну пам'ять крім наборів синтаксичних правил для кожної мови «вкладають» і правила перетворення синтаксичних структур. До цього додають правила переходу від вже перетвореної структури до речення тієї мови, на яку роблять переклад.

Інформаційно-пошукова система забезпечує пошук і відбір необхідних даних у спеціальній базі з описами джерел інформації на основі

інформаційно-пошукової мови і відповідних правил пошуку. Основні функції:

- зберігання великих обсягів інформації;
- швидкий пошук необхідної інформації;
- внесення, видалення та зміна інформації, що зберігається;
- подання вихідної інформації в зручному для людини вигляді.

Розрізняють: автоматизовані (computerised); бібліографічні (reference); діалогові (online); документальні і фактографічні інформаційно-пошукові системи.

Всю сукупність методів аналізу тексту можна розділити на дві великі групи: лінгвістичний аналіз, статистичний аналіз. Перший орієнтований на отримання сенсу тексту за його семантичною структурою. Другий – за частотним розподіленням слів у тексті.

Лінгвістичний аналіз можна розділити на чотири взаємодоповнюючих аналізи:

- лексичний аналіз полягає в розділенні текстової інформації на окремі абзаци, речення, слова, визначенні національної мови викладення, типу речення, виявленні типу лексичних виразів і т.д.;

- морфологічний аналіз зводиться до автоматичного розпізнавання частин мови кожного слова тексту. Часто морфологічний аналіз використовується в статистичних методах аналізу для попередньої процедури обробки документів – приведення слів до базової форми;

- синтаксичний аналіз полягає в автоматичному виділенні семантичних елементів речення – іменних груп, термінологічних цілих, предикативних основ;

- семантичний аналіз полягає у визначенні інформативності текстової інформації та виділенні інформаційно-логічної основи тексту.

Дослідження впливу різних алгоритмів морфологічного аналізу на якість інформаційного пошуку в текстах на різних мовах дозволяють зробити висновок, що запровадження морфологічних модулів у пошукові системи дозволяє збільшити повноту і точність інформаційного пошуку [7 – 10]. Яскравим прикладом таких систем є програмні продукти серії Yndex (Языковой index) – набір засобів повнотекстової індексації й пошуку в текстових даних з урахуванням морфології російської мови [7].

Для більш ефективної роботи вище перерахованих алгоритмів скрізь використовували модуль морфологічного аналізу, основними функціями якого є:

- граматична ідентифікація словоформи;
- отримання канонічної форми для будь-якої словоформи;
- отримання повної словотвірної парадигми для будь-якого відмінюваного слова.

2.2. Морфологічний аналіз в системах обробки природної мови

Для мов флективного типу, які мають багату словозмінну парадигму, серед задач граматичної ідентифікації складним і праце містким завданням є формалізація процесів лексичної парадигматизації та лематизації, а саме отримання повного списку словоформ певної лексеми в усіх граматичних значеннях і виведення вихідної словоформи з будь-якої текстової словоформи [2, 7-10].

Морфологія описує не тільки формальні властивості слів та утворюючих їх морфем, але і ті граматичні значення, які виражаються всередині слова. Відповідно до цих двох задач, морфологію часто поділяють на дві області: «формальну» морфологію, або *морфеміку*, в центрі якої знаходяться поняття слова і морфеми, і *граматичну семантику*, що вивчає властивості граматичних морфологічних значень і категорій (морфологічно виражаємо словотворення і словозміну мов світу).

У розробці морфологічного аналізу виділяють кілька напрямків. Один з них моделює класичну схему аналізу шляхом поділу словоформи на основу і можливе закінчення з подальшою перевіркою на сумісність закінчення з основою, яка залишається. Інший напрям використовує інформацію, що міститься в кінцевих буквосполученнях. Ця інформація отримується в результаті попередньої статистичної обробки словника. Третій напрямок широко розвивається в останні роки. На цьому напрямку створюються універсальні математичні моделі морфології у формі відкритих систем рівнянь, які дозволяють шляхом обчислення, здійснювати нормалізацію словоформ, отримання граматичної інформації та синтез словоформ.

В основу побудови алгоритмів морфологічного аналізу покладено розділення всіх слів на класи, що визначають характер зміни літерного складу форм слова. Ці класи можуть бути названі морфологічними. Зміни форм слів може носити різний характер. Вони можуть бути пов'язані як зі зміною основи слова, так і зі змінами його закінчення. Зміна букв складу основ має місце, наприклад, в наступних парах: *око* – *оці*, *книга* – *книзі*, *садити* – *саджу*, *носити* – *ношу*, *ладжу* – *ладити*. Зміна закінчень є основним способом утворення різних форм слів.

Найпоширеніші види морфологічного аналізу:

- морфологічний аналіз зі словником основ;
- морфологічний аналіз зі словником словоформ;
- морфологічний аналіз методом логічного множення;
- морфологічний аналіз без словника, за допомогою таблиць.

Найбільш поширеним видом автоматичного морфологічного аналізу є *аналіз зі словником основ*,

що використовується для більшості європейських мов. В цьому виді аналізу використовується словник основ слів і ряд допоміжних таблиць. У словник включені основи простих і складних слів без внутрішньої флексії. Якщо слово має кілька форм основ, то в словник, як правило, включені всі форми основ слів.

Морфологічний аналіз зі словником словоформ також досить поширений. Із загальних міркувань він застосовується тоді, коли морфологія даної мови досить бідна. Крім того, на перший погляд здається, що алгоритм аналізу зі словником словоформ простіший, ніж алгоритм роботи зі словником основ: не треба здійснювати членування вхідної словоформи на морфеми з послідовним пошуком за словником. Але насправді при аналізі зі словником словоформ залишається наступна проблема – аналіз не знайдених у словнику слів. Визначення деякої інформації для слова, не виявленого в словнику, є необхідним для подальшого аналізу. Якщо кожна словоформа буде виступати як самостійна лексична одиниця, то це істотно ускладнить весь подальший аналіз і синтез. Словоформи одного слова повинні бути позначені як такі. Це означає, що система морфологічного аналізу зі словником словоформ повинна мати список афіксів, коренів (основ) слів і інші необхідні атрибути для ідентифікації різних словоформ однієї і тієї ж лексичної одиниці.

Морфологічний аналіз методом логічного множення застосовується до флективних мов і передбачає наявність словника основ. Сутність методу і застосування його до конкретних мов можна побачити на прикладі алгоритму аналізу російських словоформ, запропонованого угорським фахівцем Д. Варгою. Спочатку відбувається пошук слова у словнику основ. Якщо слова, що мають закінчення, не знаходяться в словнику, тоді від кожного такого слова відкидається по одній букві праворуч і пошук повторюється. При негативній відповіді відкидається наступна буква і т.д. Відкинуті літери утворюють закінчення і фіксуються. Кожна відкинута буква вважається елементарною одиницею морфологічного аналізу. Їй приписується булевий вектор – сукупність нулів і одиниць, компонентів цього вектора. Число компонентів цього вектора дорівнює числу граматичних категорій, які можуть бути виражені закінченням, частиною якого є дана буква. Оскільки попередньо був проведений пошук за словником основ і встановлена частина мови аналізованого слова, є можливість однако-вим буквам, що входять до закінчення різних частин мови (наприклад літерам в закінченні іменника і прикметника) приписувати різні вектори.

Морфологічний аналіз без словника, або так званий «незалежний» аналіз, проводиться без звернення до словника, тільки за рахунок використання

таблиць афіксів і особливого списку слів, які не мають граматичного значення. Такий спосіб використовується досить рідко.

Сучасний стан морфологічного аналізу характеризується тим, що сильно збільшилися вимоги до якісних показників систем автоматичної переробки тексту.

Елементи морфологічного аналізу досить сильно виражені навіть в мовах з граматичним складом, що істотно відрізняються від складу європейських мов. Морфологічний аналіз китайської мови, наприклад, передбачає процедуру аналізу односкладових і двоскладових китайських слів на рівні, близькому до рівня морфем.

Зазначимо, що вибір представленої багаторівневої програмної архітектури та використання інструментальних засобів платформи .Net Remoting стали базою для створення в Українському мовно-інформаційному фонді НАН України лексикографічної системи). На базі розробленої системи виконуються різноманітні лінгвістичні дослідження, програмний комплекс використовується як додатковий інструмент при розробці нових лексикографічних систем [7–10].

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 лексикографічних систем та технологій [11].

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.

Практичне значення одержаних результатів полягає в наступному. Отримані в роботі наукові результати створюють методологічну базу для розробки ефективних підходів аналізу і синтезу ОКІС з обробкою предметно-орієнтованих знань, які дозволяють підвищити ефективність обробки великих обсягів текстової інформації, рівень автоматизації побудови онтологічних баз знань для довільних (у науково-технічній сфері) предметних галузей.

Вказані теоретичні результати використані при розробці цілого ряду онтологічних систем у

науково-технічній сфері та освіті та 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. Результати даної роботи використані при виконанні науково-дослідних проектів 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 принципів та 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. Корпусна лінгвістика / Широков В.А., Бугаков О.В., Грязнухіна

Т.О., Костишин О.М., Кригін М.Ю., Любченко Т.П., Рабулець О.Г., Сидоренко О.О., Сидорчук Н.М., Шевченко І.В., Шипнівська О.О., Якименко К.М. – К. – Довіра, 2005. – 473 с. 8. Широков В.А. Інформаційна теорія лексикографічних систем. – К.: Довіра, 1998. – 331 с. 9. Широков В.А. Феноменологія лексикографічних систем. – К.: Наук. думка, 2004. – 327 с. 10. Широков В.А. Елементи лексикографічних систем. – К.: Наук. думка, 2004. – 327 с. 11. Шабанов-Кушнарєнко Ю.П. Теорія інтелекта. – Ч.3: Проблеми і перспективи. – Харків: Вища школа, 1987. – 160 с. 12. Бондаренко М.Ф. Мозгоподобные структуры: справочное пособие. – К.: Наукова думка, 2011. – 460 с.

Надійшла до редколегії 24.07.2014

УДК 519.62

Алгебрологические и лексикографические аспекты моделирования естественного языка / Г.Г.Четвериков // Бионика интеллекта: научн.-техн. журнал. – 2014. – № 2 (83). – С. 3–14.

Статья посвящена вопросу организации данных и эффективного выбора программных и технических средств для разработки языково-информационных интерфейсов и проектов. 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.

Ил. 2. Библиогр.: 12 назв.

UDC 519.62

Algebra's logical and lexicographical aspects of modeling natural language / G.G.Chetverykov // Bionics of Intelligence: Sci. Mag. – 2014. – № 2 (83). – P. 3–14.

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.

Fig. 2. Ref.: 12 items.