

## ТЕХНІЧНІ НАУКИ

УДК 004.681

DOI: 10.15587/2313-8416.2017.118921

## МЕТОДОЛОГІЧНІ ЗАСАДИ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ЗНАНЬ

© Л. М. Бадьоріна

*Запропоновано універсальну інформаційну технологію інтелектуальної обробки знань, яка за рахунок теоретично обґрунтованої методології дозволить вилучати і обробляти “нову цінну інформацію”. Для професійної підготовки, зокрема в галузях діяльності, пов’язаних з використанням точної, семантично достовірної термінології, де спотворення формулювань, стандартизованих визначень термінів або недостатнє їх розуміння може призвести до відхилень у виконанні професійної діяльності, помилкам. Запропоновано методологічні засади обробки знань на рівні формалізації*

**Ключові слова:** інформаційні системи, інформаційні технології, моделі та методи в інформаційних системах, природно-мовні конструкції

**1. Вступ**

Аналіз світових тенденцій розвитку вищої школи, а також протиріч, що стримують його розвиток дозволили сформувати наукову проблему, що полягає в розв’язанні протиріччя між можливостями традиційних методів навчання (розробленими методами автоматизації та вилученні знань) і тенденціями розвитку сучасного суспільства заснованого на знаннях і інтелектуальних інформаційних технологіях, що вимагає систематизованого підходу до автоматизації навчання фахівців. Індустрія знань активізує дослідження з когнітивної лінгвістики, а також різних аспектів взаємостосунків між мовними та інформаційними системами опрацювання інформації. В Україні значні кроки зроблено на шляху створення навчальних технологій. За допомогою засобів відзначеного типу можлива комунікація через web-середовище. У світовій лінгвістиці підвищується значущість мовно-комунікативної проблематики, свідченням чого є наукові і технічні програми, присвячені дослідженню мережевих мовно-інформаційних систем та мережевих когнітивно-лінгвістичних процесів.

Цивілізаційне значення процесів функціонування знання в постіндустріальному суспільстві та роль мови в цих процесах набувають таких вимірів, що надають мовознавству технологічних якостей. У процесі конструювання сучасних інформаційно-комунікаційних технологій постала потреба у застосуванні властивостей природної мови для створення мовно-інформаційних систем, налаштованих на інтелектуальне опрацювання мови. Еволюція постіндустріального суспільства [1] привела передові країни світу до переходу від “інформаційного суспільства” до “суспільства знань”. Лінгвістичні технології, що виникли в ХХ столітті, докорінно змінили загальну ситуацію в інформаційному світовому просторі, структурували

лінгвістичний аналіз даних і вплинули на подальше формування як практики, так і теорії інформаційного середовища [2].

Лінгвістична технологія щодо засобу комунікації фіксує нові форми аналізу даних у сфері інформаційних технологій, визначає пріоритетні напрями розвитку електронних комунікацій та засобів масової інформації. Тому сьогодні, коли лінгвістика й інформаційні технології взаємопов’язані та взаємозалежні, постає проблема дослідження якісного рівня збалансованості комп’ютерної лінгвістики як у комунікативному, так і в соціально-економічному аспектах, характерних для українських лінгвістичних систем. Зрозуміло, що для цього потрібні й відповідні навчальні і дидактичні матеріали.

**2. Літературний огляд**

Методологічною основою досліджень були досягнення наукових шкіл в галузі розробки нових концепцій і сучасних технологій. Використовувались для дослідження наступні методи: обробка знань, їх формалізація; накопичення знань в певній предметній області; перевірка знань, що містяться в різномовних текстах на предмет функціональної повноти; семантична та логічна сумісність; загальні методи пізнання; методи системного аналізу; методи математичного моделювання тощо. Набули розвитку програмні продукти, що містять лінгвістичні модулі. Навички роботи з ними знадобляться тим, хто навчається, як при опануванні навчального матеріалу з різних дисциплін, так і у подальшій практичній діяльності, у будь-якому інформаційно-мовному середовищі. За результатами вивчення наукових джерел, над проблематикою розвитку інтелектуальних інформаційних технологій працює багато вітчизняних науковців. В умовах організації навчального процесу з використанням технічних

засобів, зокрема комп'ютерних інформаційних технологій виникає ряд теоретичних і практичних питань. Підхід, що лежить в основі створення інформаційної технології інтелектуальної обробки природно-мовної інформації на основі методу вилучення знань, які містяться в навчальних контентах. Такий підхід дозволить привести на новий рівень і концептуально-інструментальні засоби, і програмне забезпечення. Створення класу нових інтелектуальних інформаційних технологій, використання яких забезпечить якісно новий рівень взаємодії людини і комп'ютерними системами, розвиток нових функціональних можливостей цифрових технологій [3].

З розвитком систем автоматизованого навчання актуалізувалася проблема побудови формальних моделей, що описують ті чи інші аспекти значеної галузі. Серед них особливо вирізняються моделі та засоби орієнтовані на проведення автоматизованого оцінювання результатів навчального процесу. Слід відзначити, що якщо побудова навчальних контентів та цілісних систем у відзначеній діяльності розроблені достатньо повно, то, власне, автоматизація процесів оцінювання поки що перебуває на початковому етапі. Мовна система являє собою складну ієрархію різнорівневих комплексів одиниць, об'єктів та відношень.

Для створення такої методології необхідно розвинути теоретичні положення, моделі та методи побудови інтелектуальних засобів, що забезпечать можливість їх функціонування при обмежених обчислювальних ресурсах та дозволять адаптуватись до якісної обробки різнорідних лінгвістичних даних.

З розвитком теоретичних положень, моделей подання знань і на основі проведених досліджень запропоновано модель подання знань і здійснено формалізацію навчального контенту. Подання знань в формі продукцій (або правил) забезпечує формальний спосіб запису правильних рішень [3]. На відміну від традиційних методів розпізнавання когнітивне розпізнавання текстових об'єктів формалізує вилучення і подання знань з предметної галузі, які містяться в природно-мовному тексті. Входом процедури розпізнавання є природно мовний текст, який може бути представлений різними мовами. Виходом (результатом) процедури розпізнавання є поняттєва структура тексту. Поняттєва структура представляє собою формалізоване подання знань про світ (предметну галузь), відбите в певному природно мовному тексті.

Процес розпізнавання та вилучення знань з природно-мовного тексту базується на моделюванні інтелектуальних функцій людини, а саме: на комп'ютерному моделюванні процесу розуміння людиною природно мовного тексту. При цьому термін "розуміння" визначається через такі критерії: вміння переказати текст "власними" словами, тобто іншими мовними (лексичними, синтаксичними) засобами передати зміст вхідного тексту; вміння відповісти на запитання стосовно певного тексту.

Процедура розпізнавання базується на засобах формалізації (тобто розробки шаблонних моделей) знань про певну мову і знань про світ (предметну галузь).

### 3. Мета та задачі дослідження

Метою дослідження є аналіз, створення, обгрунтування та практична реалізація методологічних заasad на основі інформаційних інтелектуальних технологій; створення нових форм представлення и способів зберігання знань (навчального контенту), стратегії активації та використання знань; створення нових стратегій навчання та вивчення начального матеріалу.

Для досягнення цієї мети були поставлені такі задачі:

1. Проаналізувати існуючі комп'ютеризовані системи оцінювання знань, системи комп'ютеризованої обробки природно-мовної інформації, яка має різний професійний рівень тексту, з метою максимізації ефективного вивчення матеріалу і отримання нових знань.

2. Дослідити моделі представлення знань на основі морфемної обробки термінів предметної області.

3. Побудувати лінгвістичну модель подання знань за допомогою прикладного морфаналізу без словника.

4. Розробити методологію обробки текстової відповіді на основі словника синонімів.

5. Обчислити збіги текстових висловів на основі по елементного розбору мовних одиниць, що дасть можливість порівнювати граматичні і лексико-семантичні текстові конструкції.

### 4. Автоматизація обробки природно-мовних текстів

Методологічною основою проведених досліджень були досягнення наукових національних шкіл в галузі розробки нових концепцій і сучасних методологій. Використовувались для дослідження наступні методи:

– обробки знань, їх формалізація та приведення до єдиного подання; накопичення знань в певній предметній області;

– перевірка знань, що містяться в різномовних текстах на функціональну повноту, логічну і семантичну сумісність;

– інтегрування і узагальнення знань з певної предметної галузі;

– загальні методи пізнання (спостереження, порівняння, класифікація);

– методи системного аналізу для вивчення предметної області та розвитку інформаційних технологій освітньої сфери;

– методи математичного моделювання; створення і поповнення лексиконів, які містять необхідну для аналізу морфологічну і граматичну інформацію.

В зазначеній постановці: як задачі автоматизації розпізнавання і вилучення з природно мовних текстів-знань з предметної галузі, відомі дослідження, які спрямовані на реалізацію діалогової підтримки. При цьому компоненти комунікації можуть бути представлені наступними чином:  $K_1, K_2, \dots, K_n$ ; комунікативний текст (діалог); процеси вербалізації і розуміння; позамовна ситуація; практичні цілі; комунікативні цілі. Найбільш фундаментальні когнітивні аспекти мови включають:

– механізми зберігання і модернізації знань в семантичній пам'яті комунікативного тексту;

– когнітивні структури тексту, механізми вербалізації замислу;

– механізми поєднання текстових знань з когнітивною картиною екстралінгвістичного контексту спілкування;

– стратегії досягнення комунікативних і практичних цілей через механізми управління діалогом.

Когнітивне розпізнавання текстових об'єктів має свої особливості, а саме: на відміну від діалогу позамова ситуація (обставини) письмових текстів широкого призначення визначається лише знаннями з певної предметної галузі; засоби вербалізації цих знань спрямовані на певний рівень фахової підготовки; механізми поєднання знань в тексті з когнітивною картиною світу полягають у взаємодії представлення людиною знань про певну мову та знань про фрагменти реальної дійсності.

Природна мова, як явище інтелектуальної діяльності людини є дуже складним об'єктом. Але, маючи формальний опис природної мови, його можливо реалізувати комп'ютерними засобами і таким чином наділити машину здібністю володіти природною мовою. Алгебро-логічний апарат, що можна знайти в природній мові, дозволить розширити можливості розроблювача, який займається створенням нових інформаційних технологій. Отже, концептуально-методологічний підхід до природної мови, з математичної точки зору, дозволяє сприймати її як деяку алгебру, а тексти – як формули цієї алгебри.

При цьому зміст думок можна висловлювати реченнями і текстами, які розглядатимемо з точки зору їх математичної природи як предикати. Надалі відправною точкою в наших міркуваннях є те, що думки – це предикати. Таким чином, кожне речення виражає деяку функцію з двійковим значенням, тобто задає деякий предикат  $P(x) = \lambda$ . Незалежно змінною  $x$  даної функції буде змінна ситуація, залежною – істинна змінна  $\lambda$ . 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 зміст цього речення ситуації  $a$ , до якої воно віднесено. Нехай змінна ситуація в якості набору  $x = (x_1, x_2, \dots, x_m)$  предметних змінних  $x_1, x_2, \dots, x_m$ . Будь-яка постійна ситуація  $x = a$  має бути набором  $a = (a_1, a_2, \dots, a_m)$  деяких предметів  $x_1 = a_1, x_2 = a_2, x_m = a_m$ . Таким чином кожне речення можна виразити деяким предикатом  $P(x_1, x_2, \dots, x_m) = \lambda$ , що представляє залежність істинної змінної  $\lambda$  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 всю функцію  $P(x_1, x_2, \dots, x_m)$ , а тільки її ім'я  $P$ . І це так, бо щоразу, коли людина перетворює те чи інше речення у відповідну до нього думку вона добудовує його до предиката. При цьому вона додає до нього (як до ім'я предиката) 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 [3].

Таким чином, алгебра розглядається як інструмент дослідження, а не його предмет. Це ефективний засіб математичного подання інформації та розв'язання логічних завдань для розширення можливостей і підвищення ефективності обробки природно-мовної інформації. Завдяки якості обробки природної інформації можливо спростити процес формалізації мовних одиниць інформаційних систем та підготовки первинної інформації до етапу програмної реалізації задач [4].

Задача обробки словосполучень математичними методами вирішується в такій постановці:

Метод декомпозиції моделі  $\langle M, P \rangle$  за набором змінних на множині моделей

$$I = \left\{ \left\langle M_{\sigma_1, \sigma_2, \dots, \sigma_i}, P_{\sigma_1, \sigma_2, \dots, \sigma_i} \right\rangle, \sigma_k \in A, k = \overline{1, i} \right\}$$

засновано на теоремі про розкладання. Нехай  $x_1, x_2, \dots, x_i$  – змінні предикату  $P(x_1, x_2, \dots, x_i, \dots, x_m)$ ,  $\dots$  – їх значення відповідно. Предикат, який відповідає відношенню  $M_{\sigma_1, \sigma_2, \dots, \sigma_i}$ , має вигляд

$$M_{\sigma_1, \sigma_2, \dots, \sigma_i}(x_{i+1}, \dots, x_m) = M(\sigma_1, \sigma_2, \dots, \sigma_i, x_{i+1}, \dots, x_m). \quad (1)$$

Предикат моделі  $\langle M_{\sigma_1, \sigma_2, \dots, \sigma_i}, P_{\sigma_1, \sigma_2, \dots, \sigma_i} \rangle$  знаходимо за формулою

$$P_{\sigma_1, \sigma_2, \dots, \sigma_i}(x_{i+1}, \dots, x_m) = P(\sigma_1, \sigma_2, \dots, \sigma_i, x_{i+1}, \dots, x_m). \quad (2)$$

З усіх отриманих таким чином відношень  $M_{\sigma_1, \sigma_2, \dots, \sigma_i}$  та предикатів  $P_{\sigma_1, \sigma_2, \dots, \sigma_i}$ , ( $\sigma_1, \sigma_2, \dots, \sigma_i \in A$ ) складаємо моделі  $\langle M_{\sigma_1, \sigma_2, \dots, \sigma_i}, P_{\sigma_1, \sigma_2, \dots, \sigma_i} \rangle$ , утворюючи з них систему

$$I = \left\{ \left\langle M_{\sigma_1, \sigma_2, \dots, \sigma_i}, P_{\sigma_1, \sigma_2, \dots, \sigma_i} \right\rangle, \sigma_k \in A, k = \overline{1, i} \right\}.$$

Необхідно відзначити, що не обов'язково розкладати математичну модель за першими  $i$ -змінними. Цей метод застосовується для предиката, визначеного на всьому просторі  $U^m$ . Використовуючи метод декомпозиції моделей за набором змінних, математичну модель природної мови розкладено за набором змінних до необхідних для моделювання мовних одиниць. Основними компонентами знань з точки зору їх формалізованого подання є поняття, відношень між ними, характеристики понять і відношень, а також модальності цих характеристик. Отже, обробка вхідного тексту має бути спрямованою на виявлення (розпізнавання) в тексті основних компонент знань і встановлення логіко-семантичних відношень між ними з метою формування поняттєвої структури змісту вхідного тексту. Інформація, яка міститься в текстових джерелах, може бути поданою різними мовами, що обумовлює необхідність перет-

ворення різномовної вхідної інформації до єдиного її подання в базі знань. Воно також є основою і для синтезу опису змісту, який відображається в формалізованому поданні. До формалізованого подання знань висуваються наступні вимоги: по-перше, воно має бути представлено в такому вигляді, який забезпечить можливість коректної логіко-семантичної обробки знань; по-друге, воно має містити всю необхідну інформацію для вирішення певної задачі, тобто максимально повно на змістовому рівні зберігати текстове представлення елементів знань [5].

Представлені методи і моделі обробки текстових фрагментів в автоматизованих системах обробки знань дозволили розробити модель визначення релевантності довільної відповіді. Нехай  $S$  – множина всіх термів національної мови. Оскільки кількість термів у мові не є нескінченною, множина  $S$  є кінцевою і числовою, а також неупорядкованою:

$$S = \{s_i : 1 \leq i \leq n\}, \quad (3)$$

де  $s_i$  – терм;  $n$  – ціле число

Подамо сукупність термів будь-якої предметної галузі як множину  $T$ . Ця множина – частина множини термів мови  $S$ , а отже має всі її властивості:

$$T \subset S; T = \{t_i : 1 \leq i \leq m\},$$

де  $m$  – ціле число;  $m \leq n$ .

Запровадимо такі умовні позначення:  $p(T)$  – сукупність всіх можливих підмножин множини  $T$ ;  $e$  – елемент підмножин сукупності  $p(T)$ , який найбільш повно відповідає описуваному об'єктові або ситуації.

Будемо вважати, що один терм не може бути використаний для опису декількох об'єктів або ситуацій, тобто  $e_i \neq e_j$  [6].

Із сукупності  $p(T)$  можна виділити такі підмножини  $D_e$ , елементи яких характеризуються семантичною близькістю, що відповідає синонімічному зв'язку слів один з одним:  $D_e \subseteq p(T)$ .

Елемент  $e$ , щодо якого утвориться множина (синонімічний ряд)  $D_e$ , назвемо базовим термом, інші елементи множини  $D_e$  (слова, об'єднані синонімічним зв'язком з базовим термом) – залежними термами [5]. У рамках кожної множини  $D_e$  введемо відношення між його елементами, що позначає наявність синонімічного зв'язку між базовим термом і залежними термами. Позначимо цей зв'язок функцією  $\theta(e, t)$ . На кожній підмножині  $D_e$  вона задає частковий порядок, такий що:

$$t \leq s \leftrightarrow \theta(e, t) \leq \theta(e, s); \forall t \in D_e; \forall s \in D_e. \quad (4)$$

Цей частковий порядок можна задавати у числовому вигляді відображенням пари термів на множину дійсних чисел:  $\theta: T \times T \rightarrow R$ ,

Тоді для двох довільних термів  $e$  і  $t$  можна вказати кількісний показник їх синонімії:

$$k = \theta(e, t); e, t \in D_e,$$

значення якого знаходиться в інтервалі  $[-1, 1]$ .

Властивості та обмеження функції  $\theta(e, t)$  такі:

- 1) якщо  $k = \theta(e, t) = 1$ , то терм  $t$  є повним синонімом терму  $e$ ;
- 2) якщо  $\theta(e, t) \leq \theta(e, s)$ , то  $t \leq s$ ;  $e, t, s \in D_e$ ;
- 3) якщо  $\theta(e, t) - \theta(t, s) > 0$ , то  $t < s$ ;  $e, t, s \in D_e$ ;
- 4) якщо  $\theta(e, t) > 0$  і  $\theta(t, s) = 1$ , то  $\theta(e, s) = \theta(e, t)$ ;

Деякі залежні терми з множини  $D_{e_i}$  можуть бути базовими й у множинах  $D_{e_j}$ . Це означає наявність непрямого синонімічного зв'язку між базовим термом множини  $D_{e_i}$  і залежними термами множини  $D_{e_j}$ . Установимо відповідність між термами еталонного визначення і термами відповіді, спираючись на поняття синонімічної відповідності термів, і на основі цієї відповідності обчисли не його предмет. Це ефективний засіб математичного подання інформації та розв'язання логічних завдань для розширення можливостей і підвищення ефективності обробки природно-мовної інформації. Завдяки якості обробки природної інформації можливо спростити процес формалізації мовних одиниць інформаційних систем та підготовки первинної інформації до етапу програмної реалізації задач.

Лінгвістична модель подання та обробки знань на основі синонімії термінів предметної галузі, яка дозволить оцінити збіг елементів речень складних конструкції, поданих у вигляді природного тексту з довільною довжиною [6]. Автоматизація процесу лінгвістичної обробки інформації представлена математичною формалізацією взаємопов'язаною сукупністю формальних систем, які відбивають зміст обробки на кожному рівні. Ця сукупність організована у відповідну ієрархічну структуру формальних теорій, рівні якої узгоджені між собою за інтерфейсом. Теорію  $i$ -го рівня можна представити кортежем:  $T_i = \langle A_i, S_i, P_i, R_i \rangle$ , де  $A_i$  – вхідна інформація для  $i$ -ої теорії, яка може бути представленою в текстовій формі, множиною ознак, множиною словників і таке інше;  $P_i$  – множина правил  $i$ -ої теорії лінгвістичної обробки;  $A_i, R_i$  – результати використання правил  $P_i$  для обробки  $A_i, S_i$  – формалізовані синтаксичні правила подання всіх елементів  $A_i, P_i, R_i$ .

Множина  $P_i$  в загальному випадку може включати дві групи правил – правила безпосередньо лінгвістичної обробки і правила використання правил першої групи. Аналогом прикладу правила другої групи може бути правило використання синтаксичного правила управління. Враховуючи, що елементи множини  $R_i$  в результаті ітеративного використання правил із множини  $P_i$  можуть використовуватись в якості елементів множини  $A_i$ , в загальному випадку  $A_i \cap R_i \neq \emptyset$ . Виникає необхідність в розробці кількох

модифікацій  $T_{ij}$  деякої  $T_i$ . Враховуючи специфіку природно-мовних текстів, характерну для конкретної предметної галузі, особливостями практичних задач обробки текстової інформації виникає необхідність незначної зміни правил лінгвістичного аналізу або використання цих правил та зміни змісту інших складових теорії  $T_i$ .

Загальна теорія  $T$  представляє собою множину  $\{T_{ij}\}$  з визначеним на ній відношенням строгого порядку ієрархічного підпорядкування. З іншого боку, теорія  $T$  має бути представлена як загальна формальна система  $T = \langle A_T, S_T, P_T, R_T \rangle$ , де  $A_T$  – вхідна інформація для теорії  $T$ ,  $P_T$  – правила використання теорій  $T_{ij}$ .

$R_T$  – результати виконання правил  $P_T, S_T$  – формалізовані синтаксичні правила “зовнішнього” представлення теорій  $T_{ij}$ . Необхідність представлення теорії  $T$  у вигляді формальної системи обумовлена наступним. По-перше, за наявності модифікацій теорій  $T_i$  необхідно чітко формулювати правила використання певної  $T_{ij}$  при вирішенні конкретних задач. Процес лінгвістичного аналізу в цілому є ітеративним, тому необхідно чітко формулювати правила використання правил залучення теорій  $T_{ij}$  для обробки інформації.

В процесі лінгвістичних досліджень створюється і досліджується на коректність інформаційна база, яка забезпечує автоматизацію задач обробки текстової інформації. Ця база може включати різного призначення словники (словники квазіпрефіксів, квазізакінчень, перекладні словники, словники понять в певній предметній галузі, словники службових слів, тощо) [7], які під час експлуатації відповідних інформаційних систем поповнюються. Практичний аналіз знакового рівня організації природно-мовного тексту обмежується виділенням синтаксичних розділових знаків від слова, виділенням абревіатур, скорочень тощо.

Для автоматизованого вирішення задач обробки природно-мовного тексту суттєвим є визначення структури тексту, для виокремлення службової інформації, виділення абзаців, заголовків. Текст при цьому розглядається як певним чином організована послідовність рядків і граем. Розпізнаванням є побудова формалізованого подання граемної структури тексту та розробка формального апарату виділення і класифікації текстів на множині рядків і граем. Загальний алгоритм розпізнавання працює при наступних обмеженнях на вхідний текст: текст є відформатованим по лівому і правому краю; текст не містить переносів; текст не містить таблиць, малюнків, формул, графічних символів; текст може бути поданим англійською, українською або німецькою мовами. Розпізнавання граемного рівня подання тексту є побудова графематичної структури тексту, яка включає виділення на множині рядків і граем вхідного тексту: фрагментів (дискурсів), речень, синтагм, лексем; визначення типів (класів) перелічених одиниць тексту та встановлення відношень між ними [7].

Результат граемного аналізу є вхідною інформацією як для лінгвістичного рівня тексту (мовні лексеми, мовні речення), так і для прагматичного інтерпретатора. Задача автоматизованого аналізу зводиться до двох параметрів: якості, яка визначається точністю (рівнем помилок в побудованих синтаксич-

них конструкціях речень, повнота (ступінь покриття синтаксичними зв'язками, або зв'язність графа речення), і швидкість, яка натеper недостатня для прикладних задач.

Сутність процедури розпізнавання полягає у приписуванні кожній мовній лексемі вхідного тексту відповідної змістової інформації. Така інформація включає: лексико-граматичні класи (іменник, прикметник, чисельник, дієслово тощо), відповідні цим класам граматичні характеристики (наприклад, для іменників – це рід, число, відмінок), синтаксичні характеристики (еталонна модель передбачена лише для прикметників), семантичні характеристики (вони переписуються за певними правилами зі словника морфем і представляють собою відповідний набір семантичних ознак, які привносять до кореню префікси і суфікси, наприклад: розмір, стан, процесуальність тощо). Еталонна модель визначає всі притаманні певній морфемі характеристики. Потім здійснюється аналіз їх взаємодії у заданій лексемі, при цьому за допомогою певної множини правил розв'язання перевіряються набори ознак на суперечливість, сумісність. Результатом розпізнавання є інтегрований набір ознак, відповідний певній лексемі тексту.

Синтаксичне розпізнавання поділяється на два етапи. Задачею першого етапу є усунення граматичної омонімії морфологічного рівня і побудова словосполучень. Об'єктом розпізнавання цього рівня є закономірності взаємодії лексем у межах мовної синтагми. Вхідними даними є результати попередніх етапів розпізнавання та еталонні моделі синтаксичних правил, які описані в попередніх роботах і визначають ознаки синтаксичного з'єднання лексем у словосполучення.

Семантичне розпізнавання також поділяється на два етапи. Задачею першого етапу є усунення синтаксичної омонімії (якщо вона має місце) та формування понять, відношень і їх характеристик (властивостей) в межах одного речення. Крім того, на цьому етапі проходить заміщення анафоричних зв'язків (коли замість поняття підставляється займенник), узагальнення понять і відношень, пропусків понять і відношень, що притаманно природній мові. Об'єктами розпізнавання, на відміну від синтаксичного розпізнавання, є всі речення (а не тільки мовні), речення-заголовки, які підлягають обробці на другому етапі семантичного розпізнавання, коли будується власне поняттєва структура вхідного тексту.

Вхідними даними семантичного аналізатора є результати графемного та синтаксичного розпізнавання, а також еталонні моделі, які включають тезаурус понять, тезаурус відношень та тезаурус логіко-семантичних відношень. Структура тезаурусу, як правило, орієнтована на прикладні задачі обробки текстів. Для побудови поняттєвої структури тезаурус визначає такі відношення: рід – вид, частина – ціле, синонімії, асоціації. Ці відношення на Випадок граматичної багатозначності. З формальної точки зору випадок багатозначності можна записати як:  $A_i \rightarrow \{Q\}$ , де  $A$  – поточний образ (в даному випадку слово або словосполучення);  $i$  – унікальний набір граматичних ознак, який характеризує  $A$  в поточному контексті;  $Q$  – множина еталонних зразків, які підходять до  $A$  [8].

В природно-мовній системі граматичну багатозначність (омонімію) розрізняють таким чином: омонімію граматичних характеристик словоформ (наприклад: для словоформи *divizii* граматична характеристика числа має значення однина та множина, граматична характеристика відмінка має значення: родовий, давальний, місцевий для однини, та називний і знахідний для множини); лексико-граматичну омонімію, яка виникає на рівні лексико-граматичних класів [8].

Грамматична омонімія усувається за рахунок розроблених еталонних зразків синтаксичних зв'язків між словоформами в реченні (метод декларативного подання синтаксичних правил) – правил контекстного сполучення. Правила контекстного сполучення поділяються на правила узгодження, управління та прилягання. Розв'язання граматичної омонімії здійснюється на наступних рівнях. Так, для граматичної омонімії – на рівень розпізнавання синтаксичних сполук (тобто рівень словосполучень), для лексико-граматичної омонімії на рівень розпізнавання синтаксичної структури речення (тобто рівень речення). Формально сутність усунення граматичної омонімії в рамках правила граматичного узгодження полягає у здійсненні операції перетинання множин граматичних характеристик словоформ  $A$  і  $B$  :

$$\{G_A\}_1^n \cap \{G_B\}_1^k. \quad (5)$$

У формалізованому вигляді правило граматичного узгодження можна представити формулою:

$$P_{uzg} : N_A \otimes N_B \Rightarrow \{G\}_1^r, \quad (6)$$

де  $N_A$  – код лексико-граматичного класу словоформи  $A$ ,  $N_B$  – код лексико-граматичного класу словоформи  $B$ ,  $\{G\}_1^r$  – множина граматичних ознак, які є результатом здійснення операції узгодження граматичних характеристик для двох словоформ [9].

Лексико-граматична омонімія може виникати в результаті різного поділу обсягів понять, визначених певним словом або словосполученням. Задача

вибору на множині різних лексико-граматичних класів одного поняття виникає, наприклад, при автоматичному перекладі з однієї мови іншою [10].

## 5. Результати дослідження

Визначено недоліки існуючих технологій спричинені недосконалістю обробки різнорівневої природно мовної інформації на відміну від запропонованої технології та технології обробки знань. Проведено аналіз літературних джерел, щодо інтелектуальних інформаційних технологій, інтелектуальних технологій обробки знань, електронного навчання тощо. Обґрунтовано перспективність створення комплексної методології інтелектуальної обробки знань. Досліджено теоретико-множинну модель представлення знань на основі морфемної обробки термінів предметної області за допомогою операцій над множинами знаків; побудовано лінгвістичну модель подання знань за допомогою прикладного морфаналізу; розроблено методологію і методи обробки текстової відповіді на основі словника синонімів; визначено співпадіння текстових висловів на основі елементного розбору мовних одиниць.

## 6. Висновки

Інформаційна технологія інтелектуальної обробки природно-мовної інформації на основі методу обробки знань, які містяться в навчальних текстах дозволять перевести на новий рівень програмне і прикладне забезпечення. Отримано узагальнені результати проведеного дослідження, а саме:

1. Проаналізовано існуючі комп'ютеризовані системи оцінювання знань, системи комп'ютеризованої обробки природно-мовної інформації, яка має різний професійний рівень тексту;
2. Досліджено моделі представлення знань на основі морфемної обробки термінів предметної області;
3. Запропоновано лінгвістичну модель подання знань за допомогою прикладного морфаналізу без словника;
4. Сформовано методологічні складові обробки текстової відповіді на основі словника синонімів;
5. На прикладах показано яким чином можливо обчислити збіги текстових висловів.

## Література

1. Павлов, А. А. Информационные технологии и алгоритмизация в управлении [Текст] / А. А. Павлов, С. Ф. Телевик. – К.: Техника, 2002. – 344 с.
2. Дерещий, В. А. Об одном подходе к обработке естественно-языковых данных на основе анализа семантических сетей [Текст] / В. А. Дирецкий // Праці Першої міжнародної конференції з програмування УкрПРОГ'98. – К.: Кібцентр НАНУ, 1998. – С. 405–411.
3. Шенк, Р. Обработка концептуальной информации [Текст] / Р. Шенк. – М., 1980. – 360 с.
4. Bondarenko, M. F. Algebra of predicates and predicate operations [Text] / M. F. Bondarenko, Z. Dudar, Y. P. Kushnarenko-Shabanov, V. Chikin, N. T. Protsay, V. Cherkashin // Radioelectronics and Computer Science. – 2005. – Vol. 1. – P. 80–86.
5. Badyorina, L. M. Method of grammatical structure formalization of f natural language [Text] / L. M. Badyorina // Journal of Technical and Natural Sciences. – 2015. – Vol. 7-8, Issue 4. – P. 106–111.
6. Бадьоріна, Л. М. Метод кількісного оцінювання відповідей в системах тестування знань [Текст] / Л. М. Бадьоріна, І. В. Замаруєва // Системний аналіз та інформаційні технології. – 2011. – № 2. – С. 41–46.
7. Бадєрина, Л. Н. Метод лингвистической обработки текстовых ответов в системах тестирования знаний [Текст] / Л. Н. Бадєрина // Управляючі машини і системи. – 2011. – № 5. – С. 68–72.
8. Широков, В. А. Інформаційна теорія та системотехнічні засади комп'ютерної лексикографії [Текст]: автореф. дис. на ... д-ра техн. наук / В. А. Широков. – К., 1999. – 32 с.

9. Zaitseva, L. Course Development for Tutoring and Training Systems in Engineering Education [Text] / L. Zaitseva, J. D. Zakis // Global Journal of Engineering Education. – 1991. – Vol. 1, Issue 3.

10. Horizons of applied linguistics and linguistic technologies «MegaLing-2009» [Text]: abstracts of the IV International Conf. – Kyiv: UMF Academy of Sciences of Ukraine, 2009. – 104 p.

Дата надходження рукопису 31.10.2017

**Бадьоріна Любов Миколаївна**, доктор технічних наук, старший викладач, кафедра комп'ютерних наук, Київський національний університет культури і мистецтв, вул. Євгена Коновальця, 36, м. Київ, Україна, 01601  
E-mail: vada@ukr.net

УДК 539.21 : 541.182

DOI: 10.15587/2313-8416.2017.118801

## ИССЛЕДОВАНИЕ ФАЗОВОГО СОСТАВА И ЭЛЕКТРИЧЕСКИХ ХАРАКТЕРИСТИК НАНОКРИСТАЛЛИЧЕСКОГО ОКСИДА ЦИНКА, ЛЕГИРОВАННОГО ПЕРЕХОДНЫМИ МЕТАЛЛАМИ

©А. А. Внуков, А. Н. Головачев, А. В. Белая

*Рассмотрены предпосылки использования переходных металлов в качестве легирующих элементов при синтезе нанопорошка оксида цинка. Исследовано влияние легирования никелем, марганцем и кобальтом на фазовый состав и диэлектрическую проницаемость спеченных материалов на основе нанокристаллического оксида цинка. Результаты исследований могут быть использованы при разработке эффективных технологий получения современных материалов электротехнического назначения с повышенными эксплуатационными свойствами*

**Ключевые слова:** наноструктурные материалы, нанопорошок оксида цинка, допирование, фазовый состав, диэлектрическая проводимость

### 1. Введение

Последние годы в материаловедении были ознаменованы появлением большого интереса к наноструктурным материалам [1]. В настоящее время разрабатываются и находят применение различные виды наноструктурных материалов: нанопленки и гетероструктуры, наночастицы и кластеры, углеродные наноструктуры, субмолекулярные системы, наноструктурные покрытия, нанопористые структуры и др. [2]. Особое место среди них занимают объемные наноструктурные материалы – поликристаллы с размером зерен порядка 100 нм и менее [3]. Эти материалы имеют уникальные физико-механические свойства, значительно отличающиеся от свойств обычных, крупнозернистых материалов [4].

Одним из наиболее востребованных наноструктурных материалов является оксид цинка. Высокий интерес к оксиду цинка связан с уникальными электрофизическими свойствами данного материала. Оксид цинка является широкозонным полупроводником. Он может быть использован в качестве компонента газовых сенсоров, прозрачных полупроводников, солнечных батарей, полупроводниковых и пьезоэлектрических устройств [5].

Оксид цинка это функциональный полупроводниковый материал с уникальными свойствами. Одним из важных направлений исследований оксида цинка является изучение зависимости фазового состава и морфологии его частиц, электрических, оптических и магнитных свойств от условий синтеза и введения различных добавок (допирования). Допирование ионами переходных металлов является пер-

спективным способом регулирования как оптических, так и электромагнитных свойств ZnO.

### 2. Анализ литературных данных и постановка проблемы

Оксид цинка (ZnO) относится к бинарным соединениям типа  $A^{II}B^{VI}$ . Химическая связь между атомами цинка и кислорода является средней между ковалентной и ионной. В нормальных условиях ZnO имеет гексагональную кристаллическую структуру вюрцита.

Схематическое представление структуры ZnO показано на рис. 1.

Структура ZnO состоит из двух взаимопроникающих гексагонально-упакованных подрешеток. Каждая подрешетка включает в себя четыре атома на первичную ячейку. Каждый атом цинка окружен четырьмя атомами кислорода, которые располагаются в углах тетраэдра, и, соответственно, вокруг каждого атома кислорода располагаются четыре атома цинка. В структуре вюрцита нет центра симметрии. Это обуславливает наличие кристаллографической поляризации вдоль оси  $c$ . На рис. 1. показано чередование слоев атомов кислорода и цинка вдоль направления оси  $c$ .

Такое расположение приводит к возникновению положительного и отрицательного заряда на поверхности плоскостей (0001) и (000 $\bar{1}$ ). Свойства материала, такие как скорость травления, кинетика роста, образование дефектов, пьезоэлектрические свойства зависят от его кристаллографической ориентации [6].

Оксид цинка обладает целым комплексом свойств, делающих его чрезвычайно привлекатель-