

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
“ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”

**Омар А.Х. Авадала**

УДК 004.891.3

**МЕТОДИ І МОДЕЛІ СТВОРЕННЯ ПРИРОДНО-МОВНОГО ІНТЕРФЕЙСУ  
ЕКСПЕРТНОЇ СИСТЕМИ МЕДИЧНОЇ ДІАГНОСТИКИ**

Спеціальність 05.13.06 – Автоматизовані системи управління та  
прогресивні інформаційні технології

**АВТОРЕФЕРАТ**

дисертації на здобуття наукового ступеня  
кандидата технічних наук

Харків – 2006

Дисертацією є рукопис.

Робота виконана в Національному технічному університеті „Харківський політехнічний інститут” Міністерства освіти і науки України.

Науковий керівник кандидат технічних наук, професор  
**Судаков Борис Миколайович**,  
Національний технічний університет „Харківський політехнічний інститут”, професор кафедри систем інформації

Офіційні опоненти: доктор технічних наук, професор  
**Раскін Лев Григорович**,  
Національний технічний університет „Харківський політехнічний інститут”, професор кафедри економічної кібернетики та маркетингового менеджменту

доктор технічних наук, професор  
**Бідюк Петро Іванович**,  
Національний технічний університет України «Київський політехнічний інститут», професор кафедри математичних методів системного аналізу

Провідна установа Національний аерокосмічний університет  
ім. М.Є. Жуковського „Харківський авіаційний інститут”,  
кафедра інформаційно-управляючих систем, м. Харків.

Захист відбудеться “16” лютого 2006р. о 14-30 годині на засіданні спеціалізованої вченої ради Д64.050.07 у Національному технічному університеті “Харківський політехнічний інститут ” за адресою: 61002, м. Харків, вул. Фрунзе, 21

З дисертацією можна ознайомитись у бібліотеці Національного технічного університету “Харківський політехнічний інститут ”.

Автореферат розісланий “ \_\_\_\_ “ \_\_\_\_\_ 2006р.

Вчений секретар  
спеціалізованої вченої ради

Гамаюн І.П.

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми досліджень.** Автоматизація процесів медичної діагностики, як один з найважливіших напрямків медицини, відіграє значну роль у підвищенні надійності і точності (вірогідності) діагностики захворювань. Аналіз існуючих автоматизованих систем медичної діагностики (МД) показав, що вони не повною мірою задовольняють вимогам до рішення задач, що вимагають складних логічних умовиводів в умовах високого ступеня невизначеності, неповноти і суперечливості вихідних даних. Вихід з положення, що створилося, бачиться в інтелектуалізації цих систем на основі нових інформаційних технологій і, зокрема, у застосуванні концепції експертних систем (ЕС), що допомагають людині при рішенні задач, які важко формалізувати. Одним з елементів ЕС є підсистема взаємодії з користувачем. Основу взаємодії складають мовні засоби, оскільки тільки за допомогою мови (формальної або природної) можна досягти визначеної мети у процесі спілкування комунікантів.

Існуючі мови взаємодії і подання знань в інтелектуальних системах, як правило, мають вузькоспеціалізовану спрямованість. Вони базуються на відомих логіко-математичних моделях (численні предикатів, системах продукцій, фреймах, семантичних мережах) і не дозволяють враховувати такі особливості досліджуваної предметної галузі як неповнота і суперечливість знань і даних, динамічність і невизначеність об'єктів предметної галузі. Крім того, відсутність єдиних теоретичних підходів до розробки всіх компонентів системи взаємодії, різноманіття користувачів (експерти, когнітологи, особи, що приймають рішення (ОПР)), яким необхідно надати різноманітні мовні засоби для рішення різних задач, труднощі освоєння формальних мов непрограмуємими користувачами визначають **актуальність рішення науково-технічної задачі** – розробки методів і моделей створення природно-мовного інтерфейсу (ПМІ) експертної системи медичної діагностики (ЕСМД) захворювань, що дозволяють підвищити ефективність процесу взаємодії користувачів із системою.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційна робота виконувалася відповідно до плану наукових досліджень Національного технічного університету «ХПІ» у рамках державної теми «Дослідження технічних аспектів впровадження нових інформаційних технологій у навчальному процесі й обміну інтегрованими потоками даних між установами освіти» (ДР №0103U001541, МОН України), де здобувач був виконавцем окремих розділів.

**Мета і задачі дослідження.** Метою дисертаційного дослідження є підвищення ефективності процесу взаємодії користувачів з ЕСМД на обмеженій природній мові (ОПМ) на основі розробки нових мовних засобів і транслятора або лінгвістичного процесора (ЛП).

Для досягнення поставленої мети були вирішені наступні задачі:

- 1) аналіз існуючих методів, моделей і інформаційних технологій побудови підсистеми взаємодії користувачів з ЕСМД;
- 2) розробка методів структуризації і подання знань для ЕСМД;
- 3) розробка моделі внутрішньої мови ЕСМД;
- 4) розробка моделі ОПМ взаємодії користувачів з ЕСМД;
- 5) розробка структури й алгоритмів лінгвістичного процесора ЕСМД;

- 6) розробка моделей знань задачі діагностики захворювань;  
 7) реалізація моделей ПМІ й оцінка якості лінгвістичного забезпечення (ЛЗ) ЕСМД.

**Об'єкт дослідження** - процес взаємодії користувачів з ЕСМД.

**Предмет дослідження** – природно-мовний інтерфейс, що реалізує взаємодію користувачів з ЕСМД.

**Методи досліджень.** Проведені дослідження ґрунтуються на використанні методів: системного аналізу, штучного інтелекту, теорії категорій і топосів, апарата формальних граматики для структуризації і подання знань про предметну галузь внутрішньою мовою ЕСМД, а також апарата розміченої системи синтаксичних груп (РССГ) для побудови ЛП, що дозволило найбільш повно, адекватно і коректно забезпечити взаємодію користувача на ОПМ з ЕСМД.

**Наукова новизна отриманих результатів** полягає в наступному:

1) одержали подальший розвиток:

- метод структуризації і подання знань у ЕСМД у вигляді логічного методу, теоретичною основою якого є багатозначне (чотиризначне) числення присутності (БЧП) і теорія категорій, що дозволяє будувати ієрархічні формалізовані аксіоматичні теорії (АТ) об'єктів предметної галузі, враховувати невизначеність і суперечливість знань;

- метод формальних граматики для опису структур внутрішньої мови в лінгвістичному процесорі ЕСМД, відмінність від відомих полягає в розширенні апарата формальних граматики семантичними продукційними правилами, що дозволяє виключити тупикові ситуації, врахувати семантику при розборі вхідного висловлення, події і динаміку процесів діагностики;

2) були удосконалені:

- модель мови взаємодії для спілкування користувачів з ЕСМД, у якій кожен текст мови взаємодії представляється у вигляді двох підмножин: підмножини ОПМ, що представляє в плані вираження поверхневу структуру тексту; підмножини внутрішньої мови системи, що описує в плані змісту семантико-прагматичний (значеннєвий) рівень мови взаємодії, відмінність від відомих складається у використанні семантичних формальних граматики для опису мови подання знань, у використанні у внутрішній мові лексичних одиниць, уживаних користувачами в процесі повсякденної діяльності, що значно полегшує процес інтерпретації виражень внутрішньої мови;

- модель внутрішньої мови ЕСМД, відрізняється від відомих використанням логіко-категорного підходу до подання знань на основі БЧП, що дозволяє більш адекватно представляти процеси предметної галузі, забезпечує декларативний спосіб подання знань, а також робить систему універсальною і дозволяє без істотної зміни алгоритмів використовувати її в інших предметних галузях;

- багаторівнева модель обмеженої природної мови взаємодії користувачів з ЕСМД, відмінність від існуючих полягає у використанні апарата системи синтаксичних груп (ССГ) і розміченої структури семантико-синтаксичних груп (СССГ) для опису поверхневої і глибинної структури мови, використанні єдиного словника, синтаксичних і семантичних знань про мову і предметну галузь на різних рівнях, що забезпечує можливість природним образом перейти від поверхневого представлення текстів взаємодії до структур внутрішньої мови ЕС;

3) уперше розроблені моделі знань ЕСМД на БЧП і на внутрішній мові для опису задач диференціальної діагностики захворювань, відмінність від існуючих полягає у використанні апарата БЧП і семантичних формальних граматики, що підвищують якість (вірогідність) процесу прийняття рішень по встановленню діагнозу.

**Практична значимість отриманих результатів.** Розроблений у дисертації комплекс методів і моделей є науково-методичною основою для створення інструментальних засобів діалогової взаємодії користувачів з ЕСМД на ОПМ, що описаний алгоритмічно і реалізований у вигляді комп'ютерної знання-орієнтованої програми ЛП природно-мовного інтерфейсу ЕСМД. Універсальність логіко-категорного підходу подання знань для опису процесів фізичної реальності і формалізмів для опису ОПМ взаємодії користувачів із системою дозволяє використовувати результати досліджень при створенні ЛЗ інтелектуальних систем для різних предметних галузей.

Результати дисертаційних досліджень впроваджені:

- 1) у Харківському інституті загальної та невідкладної хірургії, кафедра госпітальної хірургії (акт про впровадження від 15.04.2005 р.) при розробці комплексних методик для експертної системи медичної діагностики;
- 2) у Національному технічному університеті «Харківський політехнічний інститут», кафедра системи інформації (акт про впровадження від 4.04.2005 р.), в рамках курсу «Програмне та лінгвістичне забезпечення інтелектуальних систем» (2 лекції (4 години) за темою 2.1 «Теоретичні принципи побудови лінгвістичного забезпечення інтелектуальних систем») та курсу «Природномовні інтелектуальні системи» (4 лекції (8 годин) за темами 1.2 «Морфологічний аналіз», 1.3 «Синтаксичний аналіз», 1.4 «Семантичний аналіз»).

**Особистий внесок здобувача.** Всі основні наукові положення, результати, висновки і рекомендації дисертаційної роботи здобувачем отримані самостійно. Серед них: методи структуризації і подання знань для ЕСМД; моделі внутрішньої і обмеженої природної мови ЕСМД; опис задач диференціальної діагностики захворювань з використанням запропонованих методів і моделей; реалізація моделей ПМІ шляхом розробки ЛП і оцінка якості ЛЗ ЕСМД.

**Апробація результатів роботи.** Основні положення дисертаційної роботи повідомлені, обговорені і схвалені: на II міжнародній науково-технічній конференції «Електронні інформаційні ресурси: проблеми формування, збереження, обробки, розповсюдження, захисту і використання - 2001» (м. Київ, 2001р.); на міжнародній науково-практичній конференції «Динаміка наукових досліджень» (м. Дніпропетровськ, 2002 р.); на II міжнародній науково-технічній конференції «Проблеми інформатики і моделювання» (м. Харків, 2002 р.); на 13 міжнародній науково-практичній конференції «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я» (м. Харків, 2005 р.).

**Публікації.** За матеріалами дисертаційної роботи опубліковано 4 статті у фахових наукових виданнях ВАК України та 4 тези доповідей на науково-технічних конференціях.

**Структура й обсяг роботи.** Дисертаційна робота складається з вступу, 4-х розділів, висновків і додатків. Результати досліджень викладені на 235 сторінках друкованого тексту, з них:

13 рисунків на 13 сторінках, 6 таблиць на 6 сторінках; бібліографія з 145 найменувань на 12 сторінках; 8 додатків на 58 сторінках.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

**Вступ** дисертаційної роботи включає: актуальність теми і наукової задачі; зв'язок роботи з науковими програмами, планами, темами; мету і задачі дослідження; об'єкт, предмет і методи дослідження; наукову новизну і практичну значимість отриманих результатів; особистий внесок здобувача; інформацію про реалізацію, апробацію і публікації результатів.

**У першому розділі** проведений огляд літератури й аналіз стану питання по застосуванню інформаційних технологій та інформаційних систем у медицині, проаналізований процес взаємодії користувачів з ЕСМД і проблеми розробки ЛЗ для ЕС, сформульований й обґрунтований основний напрямок дисертаційних досліджень по розробці природно-мовного інтерфейсу ЕСМД.

Дослідження з автоматизації процесів медичної діагностики, як один з найважливіших напрямків медицини, показали, що найбільш ефективний його розвиток досягається шляхом застосування інтелектуальних ЕСМД. При цьому інтелектуалізація задач припускає не тільки формалізацію процесу рішення задач обробки медичної інформації, але і надання більш ефективних засобів взаємодії користувачів з інтелектуальною системою. Показано, що організація взаємодії різних груп користувачів (експертів, когнітологів, ОПР) здійснюється з використанням ПМІ, основу якого складає ЛЗ, під яким розуміється сукупність засобів для формалізації природної мови, побудови і сполучення інформаційних одиниць у процесі взаємодії різних груп користувачів з інтелектуальною системою.

Рішення задачі розробки природно-мовного інтерфейсу ЕСМД має ряд особливостей, основними з яких є наступні:

1. Особливості досліджуваної предметної галузі, зокрема, необхідність враховувати невизначеність і суперечливість знань і даних, враховувати модальності типу «можливо», «необхідно» і т.д. у процесі розпізнавання захворювань привели до використання у якості формальної основи побудови ЕСМД чотиризначного числення присутності, що істотно відрізняється від формалізмів, застосовуваних у відомих інтелектуальних системах.

2. Наявність великого числа користувачів, яким необхідно надавати різні мовні засоби.

3. Відсутність єдиних теоретичних підходів до розробки всіх компонентів взаємодії.

У силу зазначених особливостей дане дослідження зв'язане з розробкою методів структуризації і подання знань, а також з розробкою моделей внутрішньої мови (ВМ) ЕСМД на основі формально-логічного БЧП і ОПМ для забезпечення взаємодії користувачів із системою.

Аналіз особливостей природних і формальних мов показав, що з погляду розв'язуваних задач у ЕСМД доцільно мати мову взаємодії (МВ), що складається з двох рівнів. Один рівень ОПМ, на якому взаємодіють ОПР і експерти. Інший – формальна мова, що є внутрішньою мовою системи, що може бути використана когнітологами. Перехід з одного рівня мови на інший здійснює транслятор або лінгвістичний процесор.

Виходячи з цього здійснена постановка задачі дослідження.

**Другий розділ** присвячений розробці методу структуризації і подання знань, а також методу семантичних формальних граматик для опису структур внутрішньої мови та моделі

внутрішньої мови для ЕСМД відповідно на базі логіко-категорного підходу, БЧП і розширеного апарата формальних граматик.

У результаті проведеного аналізу встановлено, що відомі логічні числення малоприматні для формалізації знань ЕСМД. Обґрунтовано доцільність використання для цієї мети категорного підходу до структуризації і представлення об'єктів предметної галузі. Зроблено висновок про те, що для обліку виявлених особливостей знань найбільшою мірою підходить формальна система, що побудована на основі теорії категорій – логічне БЧП. Категорії описують концепти (поняття), що відповідають класам реальних або абстрактних об'єктів, або представляють приватні приклади цих концептів, що позначають приклади об'єктів реальності. Концепт (поняття) – сукупність знань, що відбивають відмітні й істотні ознаки, властивості предмета, явища дійсності.

Кожна з категорій відбиває визначений аспект знань про концепти (інтенціонал, стан, простір, час і т.п.). Повний опис концепту можна представити у вигляді добутку категорій  $Kat_1^* \dots Kat_n^*$ .

Склад категорії міняється в залежності від розглянутого «можливого світу»  $\omega$ . У якості інформації, що характеризує об'єкт, можуть виступати ознаки, що є у свою чергу концептами. Об'єкти, істотні ознаки яких є у свою чергу концептами, надалі будемо іменувати складеними об'єктами (СО), а об'єкти, описані тільки через індивідуальні ознаки - термінальними об'єктами (ТО).

У БЧП будь-який об'єкт описується через деякі ознаки, істотні для опису об'єкта у світі  $\omega$  відповідно до теорії  $T_i$ . Основу числення складає поняття присутності об'єкта у світі  $\omega$ , що приймає одне з двох значень присутності:  $S(\omega, T_i)$  – є присутнім об'єкт у світі  $\omega$  відповідно до теорії  $T_i$  і  $N(\omega, T_i)$  – відсутній.

Задачі медичної діагностики, пов'язані з невизначеністю, приводять до того, що частина ознак може бути недоступна для реєстрації ЕС, тому їх можна розділити на два класи: доступні для системи  $\pi_{ij}$  і недоступні -  $\xi_{ij}$ . Тоді формула термінального об'єкта прийме вигляд

$$z = \bigcup_{i=1}^N \left( \bigcap_{j=0}^{p_i} \pi_{ij} \right) \bigcap \left( \bigcap_{r=0}^{m_i} \xi_{ir} \right), \quad (1)$$

де  $p_i + m_i = n_i$ ,  $i = \overline{1, N}$ . При цьому  $\pi_{i0}$  і  $\xi_{i0}$  означають  $S(\omega, T_i)$  для всіх  $i = \overline{1, N}$ .

На множині формул термінальних об'єктів можна задати топологічну булеву алгебру або алгебру присутності

$$\{X, \cap, \cup, \Rightarrow, \neg, I, C\}, \quad (2)$$

де  $X$ -множина формул об'єктів, що включають множину морфізмів  $\mu_x: x \rightarrow y$ ;  $\cap$ -операція з'єднання по присутності;  $\cup$ - операція не виключаюче «АБО» по присутності;  $\Rightarrow$ - релевантне логічне слідування по присутності;  $\neg$ -доповнення по присутності;  $I, C$ - узяття внутрішності і замикання відповідно. У такій алгебрі об'єкту  $z$  відповідає відкрита формула, якій можна поставити у відповідність її внутрішність  $I$  і замикання  $C$ , а також вектор умов присутності  $\bar{z} = \langle Iz, C\neg z, Cz, I\neg z \rangle$ , оцінки якого приймають значення з наступної сукупності:

$\text{val } \bar{z} = \text{Pr} = \langle S, N, S, N \rangle$ -присутній;  $\text{val } \bar{z} = \text{Ab} = \langle N, S, N, S \rangle$ -відсутній;  $\text{val } \bar{z} = \text{Un} = \langle N, S, S, N \rangle$ -невизначеність;  $\text{val } \bar{z} = \text{Cn} = \langle S, S, S, S \rangle$ -протириччя.

Строгий опис взаємодії об'єктів предметної галузі можливо здійснити в рамках формальних теорій. Динаміка поведіння об'єкта у фізичній (абстрактній) реальності знаходить своє відображення в теорії об'єкта через зміну оцінок присутності морфізмів, пов'язаних з даним об'єктом. З цією метою вводиться поняття диференціала присутності морфізму  $\Delta_{\alpha\beta}^n \mu_i$ , що приймає при інтерпретації одне з двох значень –  $S(\omega, T_j)$ ,  $N(\omega, T_j)$  і означає зміну значення присутності морфізму з  $\alpha$  на  $\beta$  у результаті елементарного кроку деякого процесу, що визначається порядком диференціала  $n$ .

Правила побудови формул теорій об'єктів можна визначити на основі властивостей алгебри диференціалів присутності морфізмів.

Тоді з урахуванням (2) маємо

$$\{X, \Delta_{\alpha,\beta}, \cap, \cup, \Rightarrow, \neg, I, C\}. \quad (3)$$

Такий підхід до побудови теорій дозволяє описувати теорії об'єктів через вхідні і вихідні з них морфізми, будувати багаторівневі ієрархічні теорії, здійснювати об'єднання теорій об'єктів одного рівня з іншим і в такий спосіб будувати глобальну теорію предметної галузі.

У загальному вигляді аксіоми теорій мовою БЧП представляються у вигляді

$$\Delta_{\alpha_1\beta_1}^n \mu_1 \cap \Delta_{\alpha_2\beta_2}^n \mu_2 \cap \dots \cap \Delta_{\alpha_n\beta_n}^n \mu_n \Rightarrow \Delta_{\alpha_0\beta_0}^n \mu_0, \text{ при } \alpha_0 \neq \beta_0. \quad (4)$$

На базі теорії категорій і числення присутності обґрунтовані необхідні рівні (страти) бази знань ЕСМД  $\langle M_j, D_j, C_j, PM_j, T_j \rangle$ , де  $M_j$  - множина описів об'єктів галузі  $\omega_j$ ;  $D_j$  – множина описів структур вихідних, проміжних і результуючих даних;  $C_j$  - опис множини цільових станів;  $PM_j$  – множина описів програмних модулів;  $T_j$  - множина теорій предметної галузі і задач, що описують необхідні процеси досягнення цільових станів, умови актуалізації цілей, обмеження на якість рішення задачі й ін.

Для формування множини  $M_j$  виділяються істотні властивості понять, що узагальнюються і структуруються. Для складених об'єктів визначаються підоб'єкти, їхні атрибути і ролі. При цьому кожній ролі приписується унікальне ім'я морфізму  $\alpha_n$ , підоб'єктам і атрибутам - імена понять і властивостей  $\_a_n$ . Структура  $\_$  для опису об'єктів має загальний вигляд  $\langle i, \omega, T; \alpha_1 : a_1; \alpha_2 : a_2; \dots \alpha_n : a_n \rangle$ .

Визначена правильно побудована формула теорії об'єкта, що має вигляд

$$\langle \mu(n, \text{ИОП}): i, \omega, T, S; \alpha(n, \text{ИОП}): j, \omega, T, S * \dots * \gamma(n, \text{ИОП}): k, \omega, T, S \rangle \quad (5)$$

де  $\mu$  - морфізм, що виходить з об'єкта  $i$ ;  $n$  - порядок диференціала присутності морфізму;  $\alpha, \gamma$  - морфізми, що входять в об'єкт  $i$ ;  $*$  - операції алгебри (3), ИОП - зміна оцінки присутності морфізму;  $i, j, k$  - унікальне ім'я об'єкта, по якому його можна розпізнати;  $\omega$  - "можливий світ", у якому розглядається теорія об'єкта;  $T$  - тип теорії об'єкта (процес, стан, простір, час і т.д.);  $S$  - службова інформація (коментарі, оцінка присутності й ін.).

Кожна локальна теорія об'єкта ставиться у відповідність теорії перетворення інформації, що досягається шляхом зіставлення кожній аксіомі з теорії фізичного (абстрактного) об'єкта множини аксіом у теорії інформації. Тоді кожному об'єктові в теорії фізичної реальності ставляться у відповідність сукупність перетворювачів інформації, що трансформують інформацію про вихідні



поняття, що задовольняють початковим умовам, в інформацію про результуючі поняття, що пов'язані новими умовами.

Кожен перетворювач інформації задається в наступному вигляді

$$\langle PM, \omega; \zeta_1 : z_1, \dots, z_n; \zeta_2 : v_1 \dots, v_m \rangle, \quad (6)$$

де PM – ім'я програмного модуля,  $\zeta_1$  і  $\zeta_2$  – перелік структур вхідних і вихідних даних. Самі структури даних  $z_n, v_m$  мають вигляд:

$$\langle D, \omega, T; \mu_1, T_1 : \bar{g}_1; \mu_2, T_2 : \bar{g}_2; \dots; \mu_n, T_n : \bar{g}_n \rangle, \quad (7)$$

де D – ім'я структури даних,  $\mu_n, T_n, \bar{g}_n$  - відповідно, найменування компонента структури даних, тип даних і значення даних, що можуть варіюватись.

Для опису приватних прикладів використовуються конструкції наступного вигляду

$$\langle i, \omega, T; \mu_1 : \bar{x}_1, val; \mu_2 : \bar{x}_2, val; \dots, \mu_n : \bar{x}_n, val \rangle, \quad (8)$$

де i- ім'я концепту, для якого задаються приватні приклади,  $\mu_n$  - морфізм, що визначає тип зв'язку (відносин) концепту зі списком понять (ознак)  $\bar{x}_n, val$  - оцінка присутності відповідних ознак з множини {Pr, Ab, Un, Cn}.

Проведений аналіз існуючих моделей мов показав, що для побудови моделі внутрішньої мови найбільш доцільно використовувати апарат контекстно-свободних (КС) – граматик, розширивши їх правилами, що дозволяють враховувати семантику предметної галузі  $G(L) = (V_T, V_N, P, SP, A)$ , де  $V_N$  - словник нетермінальних символів,  $V_T$  - словник термінальних символів, P- множина правил, що дозволяють будувати синтаксично правильні конструкції мови, A- множина апріорно-істинних конструкцій, що називаються аксіомами, SP-семантичні продукційні правила вигляду  $S_k(sem_k), P_z \rightarrow S_j(sem_j), P_m$ .

Дане правило означає, що якщо  $S_k$  з відповідною семантичною ознакою присутній у правилі переписування з номером z, то в правилі з номером m повинна бути присутня словоформа із семантичною ознакою  $sem_j$ .

Єдиною аксіомою граматики є "формула"

$$A = \{\text{Формула}(\Phi)\} \quad (9)$$

Вона формується у відповідності з наступними продукціями:

П1.  $\Phi \rightarrow \langle \text{ТОб}; \text{СтТ} \rangle$ .

П2.  $\text{ТОб} \rightarrow \text{Роль: ИО, КОЗ, ТО, СИ}$ ,

де ИО - унікальне ім'я, по якому можна розпізнати об'єкт або його теорію, КОЗ- код області знань, що визначає "можливий світ" даного об'єкта (теорії структур об'єктів, їхніх прогнозованих станів, обмежень і т.д.); ТО - тип конкретного об'єкта; СИ - службова інформація про теорії об'єкта.

П3.  $\text{Роль} \rightarrow \text{імя ролі (конст1, ИОП)}$ , де константа1(конст1) характеризує причину зміни оцінки присутності ролі і може приймати одне з наступних значень.

П4.  $\text{Конст1} \rightarrow \text{СТАТ} \mid \text{ВЫВОД} \mid \text{ТЕОРИЯ}^\wedge \text{ строка}$ ,

де СТАТ описує структуру в статистиці, ВЫВОД характеризує зміну оцінки присутності в результаті логічного виведення, ТЕОРИЯ<sup>^</sup> строка - теорію, у результаті якої відбулася зміна оцінки присутності.

П5.  $\text{ИОП} \rightarrow \text{Конст2-Конст2}$ .

П6. Конст2 → ПР | ОТС | НЕОПР | ПРОТ | НЕИЗВ, де ПР означає присутність, ОТС - відсутність, НЕОПР - невизначено, ПРОТ - протиріччя, НЕИЗВ - невідомо (повна апріорна невизначеність).

П7. Имя роли → ПодКл<sup>Λ</sup> строка | ПодЦ<sup>Λ</sup> строка | ПодЗад<sup>Λ</sup> строка | Пр<sup>Λ</sup> строка | Огр<sup>Λ</sup> строка | строка | ВхД<sup>Λ</sup> строка | ВыхД<sup>Λ</sup> строка | Предш | Дейст | Подч, де ПодКл, ПодЦ, ПодЗад, Пр, Огр, ВхД, ВыхД, Предш, Дейст, Подч позначають відповідно бути підкласом, підціллю, підзадачею, ознакою, обмеженням, вхідними і вихідними даними, попередньої ЦУ, ЦУ в результаті виконання дії, підлеглої ЦУ.

П8. Имя объекта → ИП | ЗП | ИПОб | ИР | ХарВр | ХарПрв | ПрКач<sup>Λ</sup> Число | ПрКач<sup>Λ</sup> ХарВр | ПрКач<sup>Λ</sup> ХарПрв | строка, де ИП - ім'я ознаки, ЗП - значення ознаки, ИПОб - ім'я підоб'єкту, ИР- виконавець ролі, ХарВр- характеристика часу, ХарПрв - характеристика простору.

П9. ПрКач → больше | меньше | не больше | не меньше.

П10. КОЗ → КОД: область.

П11. область → структура | прогноз | ограничения | ресурсы | оценки | область<sup>Λ</sup> строка | строка | <sup>Λ</sup>.

П12. СИ → СИ, СИ.

П13. ТО → объект-класс | объект-действие | объект-состояние | объект-процесс | объект-задача | теория | частный пример | тип данных | ЦУ.

П14. Тип данных → числовой | символьный | логический

П15. СИ → Мод | Комментар: строка | Кв | ОцПр | Проц<sup>Λ</sup> строка | <sup>Λ</sup> | тип вершины.

П16. Мод → необходимо | возможно | обязательно | разрешено | запрещено | доказуемо | опровержимо.

П17. Кв → существует | все | некоторые | всегда | иногда | везде.

П18. ОцПр → ПР | ОТС | НЕОПР | ПРОТ.

П19. СтГ → ТОб.

П20. ТОб → (ТОб) | [ТОб] | /ТОб/ | {ТОб} | Не ТОб.

П21. ТОб → ТОб ; ТОб.

Правила П20, П21 дозволяють поєднувати вхідні в теорію об'єкта елементи: круглі дужки вказують на те, що всі елементи усередині цієї конструкції об'єднані операцією спільної присутності; квадратні дужки поєднують ТОб операцією невиключаючого "АБО" по присутності; похилі дужки – відповідають операції альтернативної присутності; фігурні дужки - операції композиції. Заповнення деяких елементів формул, таких як "имя роли", "имя объекта", "название теории" і ін., покладається цілком на користувачів. Це вимагає додаткових знань про семантику предметної галузі. Розроблена ВМ дозволяє формально описувати теорії різних об'єктів ПГ. Оскільки в процесі взаємодії з ЕС користувачі переслідують визначені цілі, то для їхнього обліку введені спеціальні прагматичні функції, що забезпечують маніпулювання знаннями.

**У третьому розділі** приведені результати досліджень, зв'язані з розробкою лінгвістичного процесора ЕСМД, моделі ОПМ й алгоритму обробки природно-мовних текстів у ЛП, що забезпечують перехід від текстів ОПМ у внутрішнє представлення системи на ВМ.

Для переходу від тексту ОПМ (L-мови) до внутрішнього представлення (С-мови) запропоновано використовувати модель мови, засновану на багаторівневому описі. Тексти ОПМ представляють поверхневий рівень мови. Поверхнево-синтаксичний рівень мови задається структурою ССГ. Глибинний - структурою СССГ, що відповідає розміченій структурі РССГ. При цьому зв'язки між словоформами або синтаксичними групами (СГ) установлюються не тільки з використанням знань про синтаксис мови, але і знань про проблемну галузь і семантику внутрішньої мови системи. Рівень змістів представлений вираженнями С-мови.

При цьому задача трансляції полягає в перетворенні виражень  $l \in L$  (де  $L$  - множина усіх виражень L-мови) у деякий текст  $c \in C$  (де  $C$  - множина усіх виражень С-мови), або на множині  $L$  визначається відображення вигляду

$$l \xrightarrow{\mu_1} \text{ССГ} \xrightarrow{\mu_2} \text{СССГ} \xrightarrow{\mu_3} c, \quad (10)$$

де  $l$ ,  $c$  - відповідно вираження ОПМ і внутрішньої мови системи,  $\mu_1$ ,  $\mu_2$ ,  $\mu_3$  - процедури, що забезпечують відображення текстів ОПМ в ССГ, СССГ і внутрішнє представлення системи і дозволяють по заданому представленню  $l$  побудувати адекватний за змістом  $c$ .

У процесі створення моделі ОПМ і процедур її обробки природна мова користувачів була піддана нормалізації. Ступінь нормалізації визначалася з однієї сторони потребами користувачів, з іншого боку - можливостями формалізації мови і реалізації процедур обробки на ЕОМ. В результаті тексти ОПМ являли собою розповсюджені речення природної мови, які складаються з СГ (іменники з підпорядкованими ним словами) та групи управляючого слова (ГУС) (дієслово, віддієслівні іменники та ін.).

Сутність методу представлення тексту з використанням апарата ССГ полягає в наступному. Нехай  $x$  - довільний непорожній ланцюжок над словником, а  $X$  - деяка множина непорожніх підмножин  $x$ . Елементи  $X$  будемо називати синтаксичними групами. Граф  $\{X, \rightarrow\}$  називається системою синтаксичних груп на  $X$ .

Розмічена ССГ визначається як шістка  $\langle X, \rightarrow, W, Z, \varphi, \psi \rangle$ , де  $\{X, \rightarrow\}$  - ССГ,  $W$  і  $Z$  - кінцеві множини,  $\varphi$  - відображення  $X$  у множину усіх підмножин  $W$ ,  $\psi$  - відображення множини дуг графа  $\{X, \rightarrow\}$  у  $Z$ . Елементи  $W$  є мітками при СГ, елементи  $Z$  - мітками при стрілках.

У класичному трактуванні елементи  $Z$  є відносинами синтаксичного підпорядкування, обумовлені функцією тих або інших частин мови в реченні.

Для рівня внутрішньої мови елементи  $Z$  являють собою імена ролей, що виконують підоб'єкти в об'єкті.  $W$  - імена підоб'єктів (ознак).

Для виділення СГ враховувалися наступні знання про мову:  $gr_i$  - граматична інформація про словоформи (частина мови, рід, число, відмінок і т.д.);  $sint_i$  - синтаксична інформація про словоформи (з якими частинами мови і як сполучатися);  $sem_i$  - семантична інформація про словоформи ( поняття, що позначаються, значеннева сполучуваність з іншими словоформами).

Виходячи з викладеного, можна виділити СГ і установити зв'язки між словоформами (СГ) речення, використовуючи правила наступного вигляду:

$$H; S_i(gr_i, sint_i, sem_i); S_j(gr_j, sint_j, sem_j); D \rightarrow R_{i,j}, \quad (11)$$

де  $H$ - умова застосування правила (наприклад, погодженість  $gr$ ,  $sint$  і  $sem$  інформації);  $D$  - інформація про розташування аналізованих словоформ у фразі;  $R_{i,j}$  - установлений зв'язок між словоформами  $S_i$  та  $S_j$ .

Для встановлення зв'язку СГ з іншими одиницями речення використовувався опис ГУС за допомогою моделі управління - сукупності семантичної і граматичної інформації про словоформи, що грають у реченні основну роль. У загальному вигляді модель управління може бути представлена у вигляді

$$r_{k1}:CG(gr_{k1}, sem_{k1}), \dots, r_{kn}:CG(gr_{kn}, sem_{kn}), \quad (12)$$

де  $r_{ki}$  ( $i=1, \dots, n$ ) -  $i$ -а роль, виконувана СГ у  $k$ -ій моделі управління,  $gr_{ki}, sem_{ki}$  - сукупність граматичної та семантичної інформації про групу управляючого слова.

Відповідно до обраної моделі мови запропонована загальна структура ЛП, яка включає алгоритми, що реалізують відзначені етапи обробки текстів ОПМ і базу лінгвістичних знань (рис. 1). Алгоритм заснований на стратегії псевдопаралельного аналізу. Особливість цієї стратегії складається в послідовному переході від поверхневої структури тексту ОПМ до глибинного. Але при цьому вже на етапі синтаксичного аналізу використовується інформація про семантику предметної галузі, а на семантичному рівні використовуються знання про синтаксис мови.

При аналізі вираження ОПМ використовується наступна інформація, що зберігається у відповідних файлах:

1. Словник, що має дві зони. Одна зона – морфологічна, призначена для збереження коду словоформ і приписаної до них граматичної інформації. Інша зона використовується для збереження інформації про моделі управління словоформ.

2. Знання про сполучуваність словоформ. Ці знання містять дві групи правил. Перша група правил визначає можливість об'єднання словоформ у групи слів: СГ і ГУС. Друга група правил дозволяє будувати поверхнево-синтаксичні відносини між словоформами усередині кожної групи.

3. Знання, що установлюють відносини між словоформами.

4. Знання про внутрішню мову. Вони дозволяють будувати синтаксично правильні конструкції виражень  $S$ -мови.

5. Знання, отримані в процесі аналізу вхідного речення і використовувані на всіх етапах обробки вхідного тексту. Ці знання зберігаються у файлі формуляра речень (FR).

При попередній обробці тексту (A2) повідомлення, що надійшло, підготовляється для подальшої обробки. При цьому використовуються знання, що зберігаються в словнику, і формується формуляр речення.

Для побудови ССГ (A3) речення первісно розбивається на СГ і ГУС. Потім усередині кожної групи встановлюються семантико-синтаксичні відносини. Якщо такі відносини вдається установити, то здійснюється перехід до наступного алгоритму.

У протилежному випадку будується новий варіант розбивки повідомлення на групи.

У реченні, що надійшло, (A4) відшукується ім'я об'єкта і підоб'єкти, через які описується даний об'єкт, що відповідає СССГ.

На даному етапі використовуються знання, що утримуються в моделях управління і знання про формування відносин між словоформами.

Отримана глибинна структура повідомлення, усі дані про яку записані в FP, є вихідною для побудови вираження С-мови. Послідовне застосування правил переписування П1-П21, праві частини яких збігаються з даними в FP, дозволяє побудувати вираження внутрішньої мови системи і перевірити його коректність.

У четвертому розділі приведені результати реалізації розроблених методів і моделей. Розроблено моделі знань задачі діагностики захворювань на прикладі гострої кишкової непрохідності (ГКН) мовою БЧП у вигляді формул аксіоматичних теорій

$$AT_{ЗАБ}: \Delta_{PP} \mu_C(C_{1.1}^{KLM}) \cap \Delta_{PP} \mu_P(P_{1.1}^M) \cap \Delta_{PP} \mu_T(T_{1.1}^{BP}) \cap \dots \cap \Delta_{PP} \mu_C(C_{i.j}^{KLM}) \cap \Delta_{PP} \mu_P(P_{i.j}^M) \cap \Delta_{PP} \mu_T(T_{i.j}^{BP}) \Rightarrow \Delta_{QP} \mu_X(X^{ЗАБ}), \quad (13)$$

де використані наступні типи морфізмів:

$$\mu_C: C_{i.j}^{KLM} \rightarrow X_k^{ЗАБ} - \text{симптом } C_{i.j}^{KLM} \text{ характеризує захворювання } X_k^{ЗАБ};$$

$$\mu_P: P_{i.j}^M \rightarrow X_k^{ЗАБ} - \text{місце } P_{i.j}^M \text{ фіксації (ігродація) симптому } C_{i.j}^{KLM} \text{ захворювання } X_k^{ЗАБ};$$

$$\mu_T: T_{i.j}^{BP} \rightarrow X_k^{ЗАБ} - \text{час } T_{i.j}^{BP} \text{ прояву симптому } C_{i.j}^{KLM} \text{ захворювання } X_k^{ЗАБ};$$

$$\mu_X: X_k^{ЗАБ} \rightarrow Y_n^{БОЛ} - \text{захворювання } X_k^{ЗАБ} \text{ виявлене у хворого } Y_n^{БОЛ}.$$

На ОПМ і формальній VM модель знань для МД класів ГКН з урахуванням (9) може бути представлена у вигляді:

ОКН обуславлюють подклассы: СПКН, ПКН, ОБКН, СТКН, ККН

$$\langle \text{ПодКл ОКН (ВЫВОД, НЕИЗВ-ПР): } AT_{ОКН,ЗАБ}: AT_{ОЗЖ}, \text{ Об-Кл } \mu_X, X^{ЗАБ} \rightarrow Y^{БОЛ};$$

[ПодКл СПКН (ВЫВОД, ПР-ПР):  $AT_{СПКН}$ , Об-Кл  $\mu_X$ , обуславлювает;

ПодКл ПКН (ВЫВОД, ПР-ПР):  $AT_{ПКН}$ , Об-Кл  $\mu_X$ , обуславлювает; (14)

ПодКл ОБКН (ВЫВОД, ПР-ПР):  $AT_{ОБКН}$ , Об-Кл  $\mu_X$ , обуславлювает;

ПодКл СТКН (ВЫВОД, ПР-ПР):  $AT_{СТКН}$ , Об-Кл  $\mu_X$ , обуславлювает;

ПодКл ККН (ВЫВОД, ПР-ПР):  $AT_{ККН}$ , Об-Кл  $\mu_X$ , обуславлювает]>.

На мові подання знань БЧП аксіоматична теорія МД класів ГКН з урахуванням (4) і (13) може бути представлена у вигляді

$$AT_{ОКН}: \Delta_{PP} \mu_X(X^{СПКН}) \cup \Delta_{PP} \mu_X(X^{ПКН}) \cup \Delta_{PP} \mu_X(X^{ОБКН}) \cup \Delta_{PP} \mu_X(X^{СТКН}) \cup \Delta_{PP} \mu_X(X^{ККН}) \Rightarrow \Delta_{QP} \mu_X(X^{ОКН}). \quad (15)$$

Показано, що формули (14) і (15) реалізують ті самі моделі знань на VM і БЧП. Це свідчить про те, що розроблені моделі мов ЛП забезпечують синтез знань у базу знань ЕСМД на основі ОПМ.

Схема граматичного розбору формули (14) представлена на рис.2, де показані АТ – аксіоматичні теорії класів захворювань ГКН. Метою граматичного розбору є не тільки одержання відповіді на питання про приналежність даного мовного об'єкту до внутрішньої мови, але і побудова структури вхідної послідовності, що дозволяє надалі працювати з деревом виведення

(шукати підоб'єкти, поєднувати їх між собою, утворювати топологічний простір теорії об'єкта і т.д.).

Для перевірки викладених принципів обробки текстів і визначення якості розробленого лінгвістичного забезпечення в ході досліджень були програмно реалізовані алгоритми ЛП А2, А3, А4, А5 (рис. 1) при створенні ПМІ діючої ЕСМД.

У процесі взаємодії в систему вводилися тексти на обмеженій природній мові. Довжина речень складала до 30 слів, що відповідало запитам, що використовуються при діагностиці захворювань. Результати експерименту показали, що час реакції системи на запит ОПР не перевищує 1 с, що задовольняє психофізіологічним вимогам, що пред'являються до інтерактивних систем при веденні зв'язного діалогу.

Під якістю ЛЗ розуміється можливість мовних засобів задовольняти потреби користувачів по опису закономірностей предметної галузі. Якість ЛЗ  $k_{лз}$  оцінювалася з використанням наступного показника

$$k_{лз} = 1 - \sum_{j=1}^n \beta_j \sum_{i \in q_j} a_i, \quad (16)$$

де  $q_j$  - множина факторів, узятих зі шкали значимості, що враховуються в моделі мови при використанні  $j$ -го способу врахування;  $a_i$  - відносна важливість  $i$ -го фактору;  $\beta_j$  - відносний коефіцієнт обліку фактора в моделі мови.

Для ЛЗ оцінювалися наступні фактори: природність взаємодії; глибина проникнення в зміст (ключові слова, просторово-часовий опис ситуації, облік причинно-наслідкових зв'язків); можливість обліку неоднозначності і невизначеності; можливість опису понять (конкретних, абстрактних, метапонять); здатність опису евристичних знань користувачів. Розрахунки показують, що запропонована мова взаємодії в 1,2-2,4 рази перевершує по своїм можливостям існуючі мови (рис. 3). Підвищення якості лінгвістичного забезпечення ЕСМД отримано за рахунок природності МВ, можливості обліку невизначеності й опису метапонять.

Експериментальні дослідження показали, що розроблені методи і моделі можуть також використовуватися при створенні природно-мовних інтерфейсів для інтелектуальних систем різного призначення й інших предметних галузей.

## ВИСНОВКИ

У дисертації приведено теоретичне узагальнення і нове рішення актуальної науково-практичної задачі розробки комплексу методів і моделей створення природно-мовного інтерфейсу для ЕСМД на основі логіко-категорного підходу, що дозволяють підвищити ефективність процесу взаємодії користувачів із системою. При цьому отримані наступні наукові і практичні результати:

1. Проведений аналіз існуючих методів, моделей і технологій побудови підсистеми взаємодії користувачів з ЕСМД показав, що вони не дозволяють враховувати такі особливості предметної галузі, як неповнота і суперечливість знань, динамічність і невизначеність об'єктів. В зв'язку з цим обґрунтована необхідність використання логіко-категорного підходу для створення ПМІ.

2. Розроблено метод структуризації і подання знань у ЕСМД, у вигляді логічного методу, теоретичною основою якого є БЧП і теорія категорій, що дозволяє будувати ієрархічні формалізовані АТ об'єктів предметної галузі, враховувати невизначеність і суперечливість знань.

3. Запропоновано метод семантичних формальних граматик для опису структур внутрішньої мови в лінгвістичному процесорі ЕСМД, відмінність від відомих полягає в розширенні апарата формальних граматик семантичними продукційними правилами, що дозволяють враховувати події і динаміку процесів.

4. Для спілкування користувачів з ЕСМД було розроблено модель мови взаємодії. У моделі кожен текст мови взаємодії представляється у вигляді двох підмножин. Підмножини ОПМ, що представляє в плані вираження поверхневу структуру тексту, і підмножини внутрішньої мови системи, що описує в плані змісту семантико-прагматичний (значеннєвий) рівень мови взаємодії.

5. Розроблено модель внутрішньої мови ЕСМД, яка дозволяє адекватно представляти процеси предметної галузі і відрізняється від відомих використанням логіко-категорного підходу до подання знань на основі БЧП. Це забезпечує декларативний спосіб подання знань, що робить систему універсальною і дозволяє без істотної зміни алгоритмів використовувати її в інших предметних галузях.

6. Створена багаторівнева модель ОПМ взаємодії користувачів з ЕСМД, відмінність від існуючих полягає у використанні апарата ССГ і розміченої СССГ для опису поверхневої і глибинної структури мови, використанні єдиного словника синтаксичних і семантичних знань про мову і предметну галузь на різних рівнях, у можливості природним образом перейти від поверхневого представлення текстів взаємодії до структур ВМ ЕС.

7. Розроблені моделі знань ЕСМД на БЧП і внутрішній мові для опису задач диференціальної діагностики захворювань, що підвищують якість (вірогідність) процесу прийняття рішень по встановленню діагнозу.

8. Запропоновані структура й алгоритми лінгвістичного процесора, що реалізовані у вигляді комп'ютерних програм ПМІ діючої ЕСМД і здійснюють перетворення текстів мови взаємодії з одного рівня мови на інший і забезпечують синтез формалізованих АТ логіко-категорних моделей знань про предметну галузь внутрішньою мовою ЕСМД, що підтверджує практичну значимість роботи.

9. Проведено оцінку якості лінгвістичного забезпечення ЕСМД. Розрахунки показали, що застосування розробленої мови взаємодії в 1,2-2,4 рази перевершує по своїм можливостям існуючі мови, дозволяє адекватно описувати процеси предметної галузі і забезпечує користувачів доступними і зручними мовними засобами взаємодії з ЕСМД.

10. Запропоновані методи і моделі створення ПМІ реалізовані у Харківському інституті загальної і невідкладної хірургії у складі ЕСМД і у навчальному процесі Національного технічного університету «Харківський політехнічний інститут».

## СПИСОК ОПУБЛІКОВАНИХ РОБІТ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Судаков Б.Н., Омар А.Х. Авадала, Фоменко А.А. Подход к синтезу естественно-языковых сообщений по формализованному представлению базы знаний // Системы обработки информации.- Харьков: ХВУ.- 2001.- Вип. 6(16).-С.116-119.

Здобувачем був запропонований підхід до структуризації знань експертної системи.

2. Судаков Б.Н., Омар А.Х. Авадала. Принципы трансляции текстов диалогового взаимодействия пользователей с экспертной системой // Вісник Національного технічного університету «Харківський політехнічний інститут».-Харків: НТУ "ХПІ".- 2002.-№18.- С. 123-126.

Здобувачем були розроблені моделі обмеженої природної і внутрішньої мови для взаємодії користувачів з ЕС, правила виведення, що були використані в процесі трансляції.

3. Судаков Б.Н., Омар А.Х. Авадала. Лингвистический процессор в системе взаимодействия пользователя с экспертной системой // Системы обработки информации.- Харьков: ХВУ.- 2003.- Вип.1.- С.155-160.

Здобувачем було розроблено структуру і алгоритми лінгвістичного процесора, що реалізує взаємодію користувачів з експертною системою.

4. Омар А.Х. Авадала. Общение на естественном языке как процесс преобразования «Текст-Смысл» // Системы обработки информации.- Харьков: ХВУ.- 2003.- Вип. 2.-С.139-144.

5. Судаков Б.Н., Омар А.Х. Авадала. Синтез естественно-языковых сообщений при общении пользователя с экспертной системой // Праці II Міжнар. наук.-техн. конф. «Електронні інформаційні ресурси: проблеми формування, збереження, обробки, розповсюдження, захисту і використання». -К.: УкрІНТЕІ.- 2001.- С.14-15.

Здобувачем був запропонований підхід до синтезу природно-мовних текстів.

6. Омар А.Х. Авадала, Першин А.В., Судаков Б.Н. О формализованной модели языка для взаимодействия пользователей с экспертной системой // Матеріали Міжнар. наук.-практ. конф. «Динаміка наукових досліджень».- Дніпропетровськ: Наука і освіта.- 2002.- Т.1.-С.38-41.

Здобувачем була запропонована модель мови взаємодії користувачів з експертною системою.

7. Судаков Б.Н., Омар А.Х. Авадала. Естественно-языковой интерфейс интеллектуальной экспертной системы // Матеріали II Міжнар. наук.-техн. конф. «Проблеми інформатики і моделювання». – Харьков: НТУ "ХПІ".- 2002.- С.18.

Здобувачем були обґрунтовані структура інтерфейсу експертної системи, а також методи і моделі для його створення.

8. Судаков Б.Н., Омар А.Х. Авадала, Першин А.В. Подход к разработке лингвистического обеспечения экспертной системы // Матеріали 13 Міжнар. наук.-практ. конф. «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я».-Харків: Курсор.- 2005.- С.53.



Здобувачем були проаналізовані існуючі методи і моделі для створення ПМІ і запропонована методика розробки лінгвістичного забезпечення ЕС.

## АНОТАЦІЯ

**Омар А.Х. Авадала. Методи і моделі створення природно-мовного інтерфейсу експертної системи медичної діагностики.** – Рукопис.

Дисертація на здобуття вченого ступеня кандидата технічних наук за спеціальністю 05.13.06 – “Автоматизовані системи управління і прогресивні інформаційні технології”. – Національний технічний університет “Харківський політехнічний інститут”, Харків, 2005.

У дисертаційній роботі розроблено комплекс знання-орієнтованих методів, моделей, алгоритмів та інструментальних засобів природно-мовного інтерфейсу для підвищення ефективності взаємодії користувачів з експертною системою медичної діагностики (ЕСМД).

Новими науковими результатами є: 1) метод структуризації та подання знань, що оснований на багатозначному численні присутності та теорії категорій і забезпечує найбільш адекватний опис ієрархічних об'єктів медичної діагностики; 2) метод семантичних формальних граматик для опису структур внутрішньої мови, що дозволяє враховувати події і динаміку процесів; 3) багаторівнева модель мови взаємодії для спілкування користувачів з ЕСМД, яка представляє поверхневу структуру тексту у вигляді обмеженої природної мови (ОПМ) і глибинну внутрішню мову системи, що описує семантико-прагматичний (значенневий) рівень мови взаємодії; 4) модель внутрішньої мови ЕСМД, що дозволяє більш адекватно представляти процеси предметної галузі у виді аксіоматичних теорій; 5) моделі знань ЕСМД для опису задач диференціальної діагностики захворювань; 6) структура і алгоритми лінгвістичного процесора, що реалізовані у виді комп'ютерних програм природно-мовного інтерфейсу діючої ЕСМД.

Розроблений комплекс методів, моделей і засобів забезпечує підвищення якості й вірогідності результатів медичної діагностики.

**Ключові слова:** інформаційно-пошукова система, природно-мовний інтерфейс, система синтаксичних груп, структура семантико-синтаксичних груп, експертна система медичної діагностики, логіко-категорний підхід, багатозначне числення присутності, лінгвістичний процесор, лінгвістичне забезпечення.

## АННОТАЦИЯ

**Омар А.Х. Авадала. Методы и модели создания естественно-языкового интерфейса экспертной системы медицинской диагностики.** – Рукопись.

Диссертация на получение ученой степени кандидата технических наук по специальности 05.13.06 – “Автоматизированные системы управления и прогрессивные информационные технологии”. – Национальный технический университет «Харьковский политехнический институт», Харьков, 2005.

Автоматизация процессов медицинской диагностики (МД) играет значительную роль в повышении надежности и точности (достоверности) диагностики заболеваний. Сравнительный

анализ существующих методов и моделей показал, что они не позволяют решать задачи, требующие сложных логических умозаключений в условиях высокой степени неопределенности, неполноты и противоречивости исходных данных.

Возникает научная задача разработки новых методов и моделей создания естественно-языкового интерфейса (ЕЯИ) для экспертной системы медицинской диагностики (ЭСМД) заболеваний на основе логико-категорного подхода, которые позволяют повысить эффективность процесса взаимодействия пользователей с системой.

В диссертационной работе разработан комплекс знаниеориентированных методов, моделей, алгоритмов и инструментальных средств для обеспечения процесса взаимодействия пользователей с ЭСМД. При этом получил дальнейшее развитие метод структуризации и представления знаний в ЭСМД, теоретической основой которого являются многозначное исчисление присутствия (МИП) и теория категорий, позволяющий строить иерархические формализованные аксиоматические теории (АТ) объектов предметной области, учитывать неопределенность и противоречивость знаний.

Проведена структуризация знаний ЭСМД на основе теории категорий по всем уровням представления с разработкой формального описания структуры категорной модели знаний.

Первым научным результатом диссертации является метод семантических формальных грамматик для описания структур внутреннего языка в лингвистическом процессоре ЭСМД, отличие от известных заключается в расширении аппарата формальных грамматик семантическими продукционными правилами, позволяющими учитывать события и динамику процессов.

Вторым научным результатом является модель языка взаимодействия для общения пользователей с ЭСМД. В модели каждый текст языка взаимодействия представляется в виде двух подмножеств. Подмножества ограниченного естественного языка (ОЕЯ), представляющего в плане выражения поверхностную структуру текста, и подмножества внутреннего языка системы, описывающего в плане содержания семантико-прагматический (смысловой) уровень языка взаимодействия.

Третьим научным результатом является модель внутреннего языка ЭСМД, позволяющая более адекватно представлять процессы предметной области и отличающаяся от известных использованием логико-категорного подхода к представлению знаний на основе МИП, обеспечивающего декларативный способ представления знаний, что делает систему универсальной и позволяет без существенного изменения алгоритмов использовать ее в других предметных областях.

Четвертым научным результатом является многоуровневая модель ограниченного естественного языка взаимодействия пользователей с ЭСМД. Отличие от существующих заключается в использовании аппарата системы синтаксических групп (ССГ) и размеченной структуры семантико-синтаксических групп (СССГ) для описания поверхностной и глубинной структуры языка, использовании единого словаря, синтаксических и семантических знаний о языке и предметной области на разных уровнях, что позволяет естественным образом перейти от

поверхностного представления текстов взаимодействия к структурам внутреннего языка экспертной системы.

Показано, что разработанные модели языков ЛП обеспечивают синтез знаний в базе знаний ЭСМД на основе ОЕЯ. Разработаны схемы грамматического разбора формул внутреннего языка системы с целью получения ответа на вопрос о принадлежности данного языкового объекта к внутреннему языку, а также построения структуры входной цепочки, которая позволяет в дальнейшем работать с деревом вывода (искать подобъекты, объединять их между собой, образовывать топологическое пространство теории объекта и т.д.).

Реализация разработанных методов и моделей осуществлялась при создании компьютерной знаниеориентированной программы ЕЯИ действующей ЭСМД. Разработаны новая структура и алгоритмы лингвистического процессора, которые осуществляют преобразование текстов языка взаимодействия с одного уровня языка на другой и обеспечивают синтез формализованных АТ логико-категорных моделей знаний о предметной области на внутреннем языке ЭСМД. Впервые разработаны модели знаний ЭСМД на МИП и на внутреннем языке для описания задач дифференциальной диагностики заболеваний, которые повышают качество (достоверность) процесса принятия решений по установлению диагноза.

Разработанный комплекс методов, моделей и средств обеспечивает повышение качества и достоверности результатов, что подтверждается экспериментальными исследованиями на примере формализации задачи МД классов заболеваний острой кишечной непроходимости.

Проведена оценка качества лингвистического обеспечения ЭСМД в целом. Расчеты показали, что применение разработанного языка взаимодействия в 1,2-2,4 раза превосходит по своим возможностям существующие языки, позволяет адекватно описывать процессы предметной области и обеспечивает пользователей доступными и удобными языковыми средствами взаимодействия с ЭСМД.

Экспериментальные исследования показали, что разработанные методы и модели могут также использоваться при создании ЕЯИ для интеллектуальных систем различного назначения и других предметных областей.

**Ключевые слова:** информационно-поисковая система, естественно-языковой интерфейс, система синтаксических групп, структура семантико-синтаксических групп, экспертная система медицинской диагностики, логико-категорный подход, многозначное исчисление присутствия, лингвистический процессор, лингвистическое обеспечение.

## ABSTRACT

**Omar A.H. Awadala. Methods and models for creation of natural-language interface of medical diagnostic expert system.** – Manuscript.

Thesis for scientific degree of Candidate of Technical Sciences by speciality 05.13.06 – “Automated control systems and advanced information technologies”. – National technical university “Kharkiv Polytechnical Institute”, Kharkiv, 2005.

The complex of knowledge-oriented methods, models, algorithms and computer tools for natural-

language interface was developed in the dissertation work. This complex is intended to increase the effectiveness of interaction between the users and medical diagnostic expert system (MDES).

The new scientific results are: 1) method of knowledge structurization and representation based on multivalued presence calculation and theory of categories that provides the most adequate description of hierarchical objects of medical diagnostics; 2) method of semantic formal grammars for description of natural language structures, that makes possible to take into account the events and dynamics of the processes; 3) multilevel model of interaction language intended for the dialogue between users and MDES. The model represents the superficial text structure in the form of limited natural language (LNL) and the depth internal system language that describes semantic-pragmatic (signified) level of interaction language; 4) model of MDES internal language, that makes possible to represent the processes of domain more adequately in the form of axiomatic theories; 5) knowledge models of MDES to describe the tasks of differential diagnostics of illnesses; 6) structure and algorithms of linguistic process realized in the form of computer programs of natural-language interface for the existing MDES.

The developed complex of methods, models and tools provides the improvement of quality and reliability of the medical diagnostics results.

**Key words:** information-search system, natural-language interface, system of syntactic groups, structure of semantic-syntactic groups, medical diagnostic expert system, logical-categorical approach, multivalued presence calculation, linguistic processor, linguistic software.