

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
Фізико-механічний інститут ім. Г. В. Карпенка

На правах рукопису

**ЖУК**  
**Петро Антонович**

**ВЕРБАЛЬНО-СЕМАНТИЧНА  
ІНФОРМАЦІЙНО-АНАЛІТИЧНА СИСТЕМА ДЛЯ  
РОЗВ'ЯЗУВАННЯ ЗАДАЧ МАТЕМАТИЧНОГО  
МОДЕЛЮВАННЯ**

Спеціальність:

05.13.02 - математичне моделювання в наукових дослідженнях

05.13.04 - автоматизовані системи управління і системи  
обробки інформації

Автореферат  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

ЛЬВІВ - 1996

01:51  
04

ЛННБ України ім.В.Стефаника



00425919 (U)

Дисертацією є рукопис

Робота виконана в Центрі з інформаційних проблем територій Інституту прикладних проблем механіки і математики ім. Я.С.Підстригача Національної академії наук України

Науковий керівник: Член-коресп. НАН України,  
докт.техн.наук, проф. ГРИЦИК Володимир Володимирович

Офіційні опоненти: професор, докт. фіз.-мат. наук  
ВАЛЬКОВСЬКИЙ В.О., канд. техн. наук, ст. н. с., ДЕРКАЧ Б.Т.

Провідна установа: Інститут кібернетики  
ім. В.М.Глушкова НАН України

Захист відбудеться "12" березня 1996 р. о 14 год.  
на засіданні спеціалізованої ради Д 04.01.02 при  
Фізико-механічному інституті НАН України  
(290601, Львів, вул. Наукова, 5).

З дисертацією можна ознайомитись в бібліотеці Інституту  
(вул. Наукова, 5).

Автореферат розісланий "10" лютого 1996 р.

Вчений секретар  
спеціалізованої вченої ради,  
канд.техн.наук, ст.наук.сп.

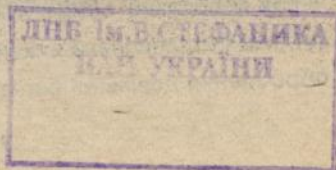
Р.А.Бунь

Д.В. 71.034

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність роботи та стан проблеми. З швидким розвитком потужностей комп'ютерної техніки і застосуванням інформаційних систем у все більш різноманітних сферах людської діяльності суттєво зростає значення математичного моделювання та створення універсальних інформаційно-аналітичних систем і баз знань. Для цих систем характерними особливостями є багатотематичність, багатоцільовість, відкритість, можливість сприйняття та семантичної інтерпретації будь-якої інформації. Оскільки універсальним способом представлення інформації, з якою оперує людина, є природна мова, необхідною властивістю повинна бути можливість семантичної інтерпретації природної мови. Можливість представлення в інформаційних системах алгоритмів та довільних зв'язків між об'єктами засобами природної мови робить такі системи надзвичайно потужним інструментом для розв'язування задач математичного моделювання, зокрема, в економіці та гуманітарній сфері.

Суттєвий внесок у розвитку наукового напрямку, пов'язаного з інтерпретацією природної мови та семантичним представленням знань в інформаційних системах, внесли українські й зарубіжні вчені В.М.Глушков, Ю.В.Капітонова, А.О.Стогній, Т.К.Вінцюк, Н.Хомський, М.Мінський, Т.Віноград, П.Вінстоу, Дж.Слейґал, А.С.Наріньяні, А.Е.Кібрик, В.Квайн та інші. Переважна кількість робіт у цьому напрямку присвячена теоретичному аналізу проблем, які виникають у зв'язку з семантичною інтерпретацією природної мови, або побудові спеціалізованих інформаційних систем, що інтерпретують досить вузькі й обмежені підмножини природної мови, як



правило, в окремих галузях знань. У даній роботі вперше розроблено і досліджено універсальне семантичне представлення української мови, що реалізоване як базове представлення знань довільної тематики в інформаційно-аналітичній системі.

Як показує досвід розвитку сучасної науки, дуже часто важливі нові результати виникають на стику різних галузей досліджень. При цьому необхідний одночасний розгляд і аналіз великого обсягу різноматематичних знань, що без використання потужних інформаційних систем у багатьох випадках є практично неможливим. Такі системи повинні забезпечувати збереження і сумісну обробку інформації, що стосується різних галузей, її ефективне тематичне структурування, можливості розширення та поглиблення тематики. Повинна бути забезпечена можливість використання однієї й тієї ж інформації для розв'язування великої кількості цілком різних за типом і складністю задач: від довідок до реалізації потужних математичних моделей і керування складними технічними системами. Тому все частіше інформаційні системи створюються не для розв'язування конкретних актуальних на даний час задач, а для комплексного збору й обробки інформації про об'єкт у цілому. Представлення знань засобами природної мови найкраще забезпечує їх повне та багатоцільове використання, оскільки саме природна мова відображає модель світу, систему позначень і домовленостей, яка об'єднує різні теми та галузі.

Мета роботи і задачі дослідження. Метою роботи є розроблення вербально-семантичної інформаційно-аналітичної системи, яка би дозволяла виконувати семантичну інтерпретацію та оброблення довільної багатотематичної вербальної інформації

для цілей математичного моделювання.

Для досягнення мети виконувалися наступні завдання:

-формулювання й обґрунтування концептуальних положень, які лежать в основі інформаційної технології, що створюється;

-формулювання вимог до мови семантичного представлення вербальної інформації, побудова формальної граматики, що породжує мову, яка задовольняє цим умовам, дослідження властивостей такої граматики;

-побудова та дослідження властивостей формальної семантичної моделі, яка дає спосіб інтерпретації вербальних знань в інформаційній системі;

-створення та дослідження системи базових функцій, яка забезпечує семантичну інтерпретацію та оброблення інформації в інформаційній системі;

-створення алгоритмів вербально-семантичної інформаційно-аналітичної системи та їх комп'ютерна реалізація;

-розв'язування засобами створеної системи конкретних комплексних задач наукових досліджень у суттєво різних галузях: когнітологічні дослідження територій України та дослідження вітроенергетичного потенціалу в регіонах України.

Методи досліджень. При виконанні роботи використовувалися методи та апарат теорії множин, теорії формальних граматики, теорії дедуктивного виведення, загальної теорії систем, теорії алгоритмів, математичної лінгвістики, теорії графі, математичної логіки, семантики, математичної статистики, програмування.

Теоретичний аналіз поєднується з розробленням алгоритмів і створенням програмного забезпечення, яке використовується для розв'язування прикладних задач.

Наукова новизна. Запропонований підхід до семантичної інтерпретації природної мови дозволив вперше реалізувати комп'ютерну інформаційну систему, в якій може бути представлена та семантично інтерпретована інформація, виражена засобами природної української мови. Запропоновано та досліджено безконтекстну граматику семантичного представлення вербальної інформації та семантичну модель предметної сфери інформаційної системи у формі узагальненого гіперграфа. Доведено інтерпретовність семантичного представлення, тобто однозначну відповідність будь-якого речення, породженого граматикою семантичного представлення, підграфу узагальненого гіперграфа, який реалізує семантичну модель. Сформульовано критерії якості семантичного представлення вербальної інформації. Розроблено алгоритми семантичної інтерпретації вербальної інформації, генерації нових знань шляхом дедуктивного виведення на узагальненому гіперграфі. Запропоновано технологію реалізації математичних моделей на основі вербального представлення активних і пасивних знань в інформаційній системі.

На захист виносяться наступні основні положення:

- концепція побудови універсальної інформаційно-аналітичної системи з базовим вербально-семантичним представленням інформації;
- безконтекстна граматика семантичного представлення знань, виражених природною українською мовою;
- семантична інформаційна модель, яка забезпечує інтерпретацію речень вербально-семантичного представлення в інформаційній системі;
- формулювання та доведення тверджень, які виражають основні властивості граматики семантичного представлення

та семантичної інформаційної моделі, зокрема, семантичну інтерпретовність;

- базова система функцій на семантичній інформаційній моделі, яка реалізує представлення, інтерпретацію і повну множину операцій над вербально-семантичною інформацією;

- алгоритми та діюча програмна реалізація вербально-семантичної інформаційно-аналітичної системи "Поступ", яка реалізує семантичну інтерпретацію й оброблення довільної багатотематичної вербальної інформації для цілей математичного моделювання;

- наукові дослідження в двох різних галузях знань на основі математичних моделей, реалізованих засобами системи "Поступ".

Достовірність основних наукових положень і отриманих результатів забезпечується застосуванням математичних методів і викладок, безпосередньою перевіркою придатності запропонованих формальних структур для опису відповідних класів об'єктів, співпадінням результатів у відомих часткових випадках. Отримані результати використані для розв'язування конкретних задач математичного моделювання. Достовірність отриманих даних підтверджено практикою.

Практична цінність. Одержані результати закладають основу для повної семантичної інтерпретації знань, виражених природною українською мовою, що дозволить перейти до створення комп'ютерних інформаційних систем якісно нового рівня, які характеризуються універсальністю, простотою введення інформації, зокрема алгоритмів, повнотою використання інформації, широкими можливостями її аналізу та одержання нових знань. На основі проведених досліджень розроблено структуру та алгоритми вербально-семантичної інформаційно-аналітичної

системи "Поступ", яка програмно реалізована і апробована. Система "Поступ" є потужним інструментом для реалізації математичних моделей і наукових досліджень у різних галузях. Особливо ефективним виявилось застосування системи для розв'язування задач в економічних і гуманітарних науках, оскільки вона забезпечує семантичне представлення й ефективне використання довільної вербальної інформації. Доцільно застосовувати систему для досліджень на стику різних наук, зокрема для комплексних досліджень територіальних об'єктів, а також для побудови математичних моделей на основі великої кількості емпіричних і теоретичних закономірностей, різнопланових і різних за достовірністю вхідних даних.

Реалізація результатів роботи. На базі системи "Поступ", яка є основним практичним результатом даної роботи, реалізовано комп'ютерну інформаційно-аналітичну систему Львівської області, інформаційну систему для цілей землекористування Івано-Франківської області, на замовлення Ради з вивчення продуктивних сил створено комп'ютерну інформаційно-картографічну систему Житомирської області, орієнтовану на інформацію про динаміку зміни стану радіоактивного забруднення території внаслідок аварії на ЧАЕС. Здійснено розроблення інформаційно-аналітичної системи Західного регіону України для потреб альтернативної енергетики. Найповніше використані можливості системи "Поступ" при створенні багатотематичної комп'ютерної інформаційної системи об'єктно-орієнтованого типу з українознавства для наукових досліджень, прийняття рішень та забезпечення навчального процесу в рамках Державної науково-технічної програми 8.3.1. На основі цієї системи проведено ряд комплексних соціологіч-

них досліджень і прогнозів, виконано аналітичні дослідження соціально-політичної, демографічної й екологічної ситуації в територіях України, розроблено та передано в Адміністрацію Президента проект нового адміністративного поділу України. Результати досліджень, виконаних на основі інформаційної системи з українознавства, використовувались Адміністрацією Президента України, Верховною Радою, Національним інститутом стратегічних досліджень.

#### Апробація роботи.

Основні результати були викладені та обговорені:

- на Третій міжнародній науково-практичній конференції "Проблеми українізації комп'ютерів", Львів, 1993р.,
- на міжнародній науково-практичній конференції "Карпати - український міст в Європу: проблеми і перспективи", Львів, 1993р.,
- на Другій українській конференції з автоматичного керування, Львів, 1995р.,
- на семінарах у Державному комітеті з питань науки і технологій України, Національному інституті стратегічних досліджень, Інституті психології НАНУ, ІППММ НАНУ, ФМІ НАНУ, Секції інформатики ЗНЦ НАНУ, 1991-1995рр.

Публікації. За темою дисертаційної роботи опубліковано дев'ять наукових робіт.

Структура й об'єм роботи. Дисертація складається з вступу, чотирьох розділів, висновків і викладена на 148 сторінках включно з переліком літератури та додатків.

### ЗМІСТ РОБОТИ

У вступі окреслюється коло питань, які є предметом дослідження, показується актуальність напрямку роботи, вводяться

основні поняття і коротко описується зміст дисертації.

Завданням першого розділу є опис концепції системи "Поступ", формальне представлення основних структур, що задають семантику інформації і дослідження зв'язку між ними.

Концепція системи "Поступ" задає наступні основні її властивості: об'єктність, природну ієрархічність інформаційних структур, суб'єкту розподіленість, паспортизованість інформації, вербальність, семантичну інтерпретованість інформації, вербальне представлення активних знань.

Базовим концептуальним постулатом, на якому ґрунтується система, є **принцип первинності слова, природної мови, як основи форм і змісту інформації**. Модель світу та інформаційні структури системи будуються, виходячи з їх відповідності моделі та структурі знань, відображеній у природній мові. Звичайно, це не виключає можливості спеціалізованих форм введення та відображення інформації (таблиці, графіки тощо), які широко використовуються в математичному моделюванні, однак ці форми є похідними від вербального представлення.

Вербальне представлення виконує в системі три основні функції: *сприйняття інформації, внутрішнє представлення для збереження й оперування інформацією, відображення інформації для користувача*. Кожна з цих функцій вимагає від вербального представлення певних специфічних властивостей.

Отже, в роботі інформаційної системи з природною мовою виділяються 3 блоки. *Герменевтичний блок* - перетворення вхідної інформації у внутрішнє (семантичне) представлення, відділяючи зміст інформації від особливостей стану суб'єктів та діалогу. *Семантичний блок* - зберігання, доповнення, пошук, групування, перетворення семантичної інформації, генерування нової семантичної інформації засобами дедук-

тивного виведення та математичного моделювання. *Риторичний блок* - перетворення вербальної інформації з семантичного представлення в форму, яка найкраще відповідає конкретним потребам користувача, його особливостям і стану.

Підхід, запропонований в даній роботі, передбачає першочергову побудову та розвиток семантичного блоку. У цьому варіанті інформаційна система створюється, функціонує й апробується для розв'язування задач математичного моделювання й аналізу складних систем, використовуючи семантичне представлення і як внутрішнє, і як вхідне, і як вихідне.

Отже, першочерговим завданням є побудова семантичного представлення вербальної інформації для природної української мови, яке повинно задовольняти наступним умовам.

**А.1.** Будь-який інформативний текст, написаний українською мовою, може бути без втрати інформації про об'єкти, імена яких є в тексті, представлений у формі послідовності речень семантичного представлення.

**А.2.** Кожне речення семантичного представлення повинно бути семантично інтерпретовним, тобто задавати певне співвідношення чи стан об'єктів, якому можуть бути однозначно приписані певні властивості, яке може бути за наявності відповідних засобів змодельовано чи перевірено його істинність.

Для генерації мови семантичного представлення пропонується безконтекстна формальна граматики  $G$ .

$$G = \langle P, V_N, V_T, \langle \text{речення } 0 \rangle \rangle,$$

де  $V_N$  - множина нетермінальних символів,  $V_T$  - множина термінальних символів (слів),  $P$  - множина правил підстановки.

$$V_N = V_{N0} \cup \bar{Z},$$

$V_{N0} = \{ \langle \text{речення } 0 \rangle, \langle \text{речення} \rangle, \langle \text{уточнення} \rangle, \langle \text{питання речення} \rangle \} \cup \{ \langle \text{речення присудка } \beta \mid \beta \in B \rangle \} \cup \{ \langle \text{група}$

{<питальне речення> → " <питальне речення>" <група присудка  $\beta$  > |  $\beta \in V'$  >

{<реалізація абстракта  $\kappa$  > →  $\kappa$  , <питальне речення> , |  $\kappa \in K$  }  
{<питальне речення> → <питальне речення> <уточнення>

*Примітка.* У описі  $P$  всі нетермінальні та термінальні символи в правих частинах продукцій розділені пробілами. { , } , → , ... , | , вирази після | і позначення змінних є символами метамови опису  $P$ .

У описі  $P$  використані наступні позначення:  $z_j(\beta)$  - зміст  $j$ -го члена речення присудка  $\beta$  ,  $v(\beta)$  - кількість фіксованих членів речень присудка  $\beta$  ,  $\mu(z)$  - множина слів, що можуть бути значеннями змісту, який відображається нетермінальним символом  $z$  ,  $\bar{Z}'$  - множина змістів, що можуть бути відображені цілими реченнями,  $\alpha(z)$  - множина абстрактів, що можуть мати зміст, який задається  $z$ ;  $\alpha(z) \subset K$  ,  $C(\gamma)$  - зміст уточнення сенсу  $\gamma$  (уточнення певного сенсу задають умови, для яких формулюється основний зміст речення, й інші необов'язкові для основного змісту особливості),  $V'$  - множина присудків, підметами в реченнях яких можуть бути цілі речення.

Розглядається питання відповідності мови, згенерованої граматикою  $G$ , умовам А.1 і А.2. Семантична інтерпретовність розглядається як відповідність між реченнями семантичного представлення й елементами певної семантичної моделі світу, що описує відношення та властивості об'єктів, які становлять зміст відповідних речень. Така семантична модель представлена в формі узагальненого гіперграфа  $H$ .

$H = \langle V, N, R \rangle$ ,

де  $V$  - множина вершин,  $N$  - базова множина нав'язок (узагальнених кольорів) на дуги гіперграфа,  $R$  - множина дуг (орієнтованих) гіперграфа, позначених нав'язками. Вершинами

присудка  $\beta \mid \beta \in B \cup \{ \langle \text{реалізація абстракта } k \rangle \mid k \in K \}$ , де  $B$  - множина всіх присудків,  $K$  - множина всіх абстрактів (загальних назв об'єктів, які можуть мати різні конкретні реалізації),  $\bar{Z}$  - нетермінальна множина змістів окремих членів речень.

$$V_T = V_{T0} \cup V'_T,$$

де  $V_{T0}$  - множина всіх слів природної мови, які інтерпретуються системою, тобто простих імен об'єктів і чисел,

$V'_T = \{ \dots, \text{який}, " \}$  - множина службових (системних) термінальних символів (тут одним з символів є кома (,), яка в цій формулі є ще й розділювачем метамови).

Імена об'єктів (елементи  $V_{T0}$ ) у семантичному представленні вживаються в основних (інфінітивних) граматичних формах, чи в формах, які відображають специфіку їх використання та розуміння тексту.

$$P: \langle \text{речення } 0 \rangle \rightarrow \langle \text{речення} \rangle .$$

$$\langle \text{речення} \rangle \rightarrow \langle \text{речення} \rangle \langle \text{уточнення} \rangle$$

$$\{ \langle \text{речення} \rangle \rightarrow \langle \text{речення присудка } \beta \rangle \mid \beta \in B \}$$

$$\{ \langle \text{речення присудка } \beta \rangle \rightarrow z_1(\beta) \langle \text{група присудка } \beta \rangle \mid \beta \in B \}$$

$$\{ \langle \text{група присудка } \beta \rangle \rightarrow \beta \ z_2(\beta) \dots z_{v(\beta)}(\beta) \mid \beta \in B \}$$

$$\{ z \rightarrow \omega \mid z \in \bar{Z}, \omega \in \mu(z) \}$$

$$\{ z \rightarrow " \langle \text{речення} \rangle " \mid z \in \bar{Z}' \}$$

$$\{ z \rightarrow \langle \text{реалізація абстракта } k \rangle \mid z \in \bar{Z}, k \in \alpha(z) \}$$

$$\{ \langle \text{уточнення} \rangle \rightarrow \gamma \ c(\gamma) \mid \gamma \in \Gamma \}$$

$$\{ \langle \text{реалізація абстракта } k \rangle \rightarrow \langle \text{реалізація абстракта } k \rangle , \langle \text{питальне речення} \rangle \mid k \in K \}$$

$$\{ \langle \text{питальне речення} \rangle \rightarrow \text{який} \langle \text{група присудка } \beta \rangle \mid \beta \in B \}$$

$$\{ \langle \text{питальне речення} \rangle \rightarrow \langle \text{речення} \rangle \ \gamma \ \text{який} \mid \gamma \in \Gamma \}$$

$$\{ \langle \text{питальне речення} \rangle \rightarrow Z_1(\beta) \ \beta \ \text{який} \ Z_3(\beta) \dots Z_{v(\beta)}(\beta) \mid (\beta \in B)$$

$$\& \ v(\beta) \geq 2 \}$$

гіперграфа є всі предметні члени речень, і ті речення, які є частинами складних речень. Дугам гіперграфа відповідають речення, а нав'язки відображають інтерпретацій-ний сенс речень, тобто суть актуальної властивості чи зв'язку, який задається реченням.

Таким чином,

$$V = V_{T0} \cup Q \cup R_T,$$

де  $V_{T0}$  - множина об'єктів, імена яких є неслужбовими термінальними символами граматики  $G$ ,  $Q$  - множина об'єктів із складними іменами (імена яких задаються реченнями чи означеннями),  $R_T$  - множина дуг, що можуть трактуватися, як вершини; їм відповідають речення граматики  $G$ , що виступають як члени складних речень,  $N$  задає множини слів, які використовуються при побудові нав'язок для конкретних речень.

$$N = \langle B, G, A, M \rangle,$$

де  $B$  - множина можливих присудків з  $G$ ;  $G$  - множина сенсів уточнень з  $G$ ;  $A$  - множина об'єктів з атрибутивним використанням, тобто таких, які можуть описувати певні властивості інших об'єктів, при цьому деякі речення, в які входять імена цих об'єктів, описують властивості чи зв'язки не цих об'єктів, а інших предметних членів речення, - в цьому випадку дуга, яка відображає речення, не інцидентна вершині, яка відповідає атрибутивному об'єкту;  $M$  - множина чисел, яка використовується для відображення кількісних властивостей об'єктів.

У роботі показано достатність  $N$  для семантичної інтерпретованості будь-якої множини речень, породжених  $G$ . Вказується спосіб побудови кожного речення певного підграфа  $N$ , що однозначно відображає його зміст і не залежить від інших речень.

Другий розділ містить дослідження властивостей семантичної моделі  $N$ , зокрема, зв'язаних із трактуванням її елементів у

системі понять природної мови, а також реалізацією в обчислювальній системі. Представлено трактування основних концептуальних положень інформаційної системи "Поступ" у термінах семантичної моделі світу  $H$ . Так, будується тематично-множинна ерархія вершин  $H$ . Ця ерархія задається множиною тем  $\Theta \subset V$ , множиною типів об'єктів  $T \subset V$ , сукупністю множин  $I$ , цілком й однозначно заданими на  $V$  функціями  $\vartheta(\omega)$  і  $\tau(\omega)$  та частковою на  $\Theta \times T$  функцією  $\mu_\tau(\vartheta, \tau)$ . При цьому  $\vartheta(\omega)$  - тема об'єкта  $\omega \in V$ ;  $\vartheta(\omega) \in \Theta$ ,  $\tau(\omega)$  - тип об'єкта  $\omega \in V$ ;  $\tau(\omega) \in T$ ,  $\mu_\tau(\vartheta, \tau)$  - множина з  $I$ , що містить об'єкти типу  $\tau$  теми  $\vartheta$ . Множина  $\Theta$  частково впорядкована відношенням "надтема"  $< \tau$ .

Для такої ерархії повинні виконуватися наступні умови:

$$\forall \omega \in V \exists \mu \in I (\omega \in \mu)$$

$$\forall \tau \in T \forall \vartheta_1, \vartheta_2 \in \Theta (\mu_\tau(\vartheta_1, \tau) \subset \mu_\tau(\vartheta_2, \tau) \Rightarrow \vartheta_1 < \tau \vartheta_2)$$

$$\forall \mu \in I (\forall \omega \in \mu (\mu_\tau(\vartheta(\omega), \tau(\omega)) = \mu))$$

$$\forall \mu \in I \exists \vartheta \in \Theta \exists \tau \in T (\mu = \mu_\tau(\vartheta, \tau))$$

Розглядаються зв'язки між моделями світу, які реалізуються в різних локальних базах. Дослідження зв'язностей різних підграфів  $H$  можуть дати висновки про повноту представлення інформації в темах і дозволяють формулювати запити на цілкове введення інформації, якої бракує.

Для деяких присудків наявні певні логічні умови, що зв'язують коректність речень з цими присудками з коректністю певних інших речень. При приєднанні до  $H$  дуг, що відповідають таким реченням, виникає задача несуперечливості інформації в  $H$ , яка вимагає ефективного алгоритму виявлення інформаційних конфліктів і формулювання правил зміни модальностей речень у цих випадках. Такі алгоритми пропонуються в узагальненому вигляді як алгоритми для розв'язування задач

дедуктивного виведення нових знань на семантичній моделі. Доводяться теореми про коректність запропонованих алгоритмів.

Ряд задач виникає у зв'язку з тематичною структурою інформації. Кожна тема утворює в  $H$  певний підграф. Такий підграф може бути видалений з  $H$  (знищення в базі інформації, яка стосується певної теми), приєднаний до іншого семантичного гіперграфа (імпорт тематичної інформації іншою локальною базою). Теми можуть також задавати обмеження на пошук чи виведення. Оскільки об'єкти в темі зв'язані з іншими темами (зокрема, частини нав'язок - відношення, уточнення, - часто мають значно ширшу сферу використання, ніж тема), при розділенні та злитті тематичних підграфів виникає ряд проблем зі зовнішніми зв'язками. Об'єкти, які використовуються в різних темах, доцільно групувати в теми, які є перетинами тем, і мають вищий пріоритет збереження та єрархічний рівень. Таким чином, виникає задача порівняння природної єрархії тем з єрархічною структурою, яка виходить з будови  $H$ .

Вводиться ряд інформаційних структур, що необхідні для програмної реалізації системи оперування зі сукупністю знань, представлених у вигляді  $H$ . Вся інформація, яка представлена в гіперграфі  $H$  зберігається з об'єктами, що відповідають його вершинам, у структурах, які названі фреймами. При цьому обов'язковою і регламентованою є лише невелика частина інформації, переважно, для об'єктів спеціальних типів. Інша частина інформації довільна і для неї у фреймах нема спеціально виділеного місця, що відрізняє даний спосіб представлення інформації в фреймах від класичного. Речення в фреймах представлені у вигляді послідовностей логічних зносок їх членів і впорядковані лексографічно відповідно до порядку логзносок. При цьому, присудок, як структурно і семантично визначаль-

ний член речення, ставиться на перше місце. Речення, яке відповідає певній дузі  $H$ , зберігається в фреймах усіх об'єктів, які відповідають інцидентним до цієї дуги вершинам. З кожним реченням у фреймах зберігається деяка службова інформація, зокрема верифікаційний показник, який вказує на походження речення, та його модальність.

Формулюється ряд задач над введеними структурами, які зв'язані з оптимальною програмною реалізацією системи, і пропонуються шляхи їх розв'язку.

У третьому розділі описані засоби обробки інформації і математичного моделювання, які реалізовані в системі "Поступ".

Завданням інформаційної системи є реалізація на семантичній моделі світу  $H$  наступних основних функцій: доповнення  $H$  (вершинами, дугами, темами-підграфами заданої структури); вивід на засоби відображення інформації частин гіперграфа  $H$ , представлених у різних формах, зокрема, у вигляді семантичного тексту; пошук об'єкта (вершини  $H$ ) з заданими властивостями; пошук речення заданого змісту; видалення частин  $H$ ; виділення підграфа  $H$ , що відображає сукупність інформації певного тематичного напрямку; дедуктивне виведення речення заданого змісту; перевірка істинності твердження, заданого певним реченням; виконання дії, заданої реченням (зокрема, кроку алгоритма, який реалізує математичну модель). Виконуючи основні функції, можна створювати прикладні процедури оброблення та відображення інформації в різних тематичних галузях. З іншого боку, ці функції самі є достатньо складними, можуть бути реалізовані в різних варіантах, їх безпосередня програмна реалізація містила б багато груп операторів, які повторюються в різних функціях. Тому виділяється множина  $F_0$  базових елементарних функцій, які реалізуються на

базовій універсальній мові програмування, а всі інші можуть бути побудовані з них засобами системи. Множина  $F_0$  включає наступні функції: створення вершини (об'єкта); створення дуги (речення); текстове відображення імені об'єкта; текстове відображення речення; перебір (цикл) об'єктів множини; пошук речення, заданого його частиною; знищення дуги; знищення вершини; зміна імені об'єкта; підстановка в реченні (зміна члена речення); виділення частини складного речення; арифметичні, символічні та логічні операції.

Показано, що всі основні функції реалізуються через сукупність  $F_0$  базових функцій (повнота  $F_0$ ), і що жодна з базових функцій не може бути реалізована через сукупність інших (незвідність  $F_0$ ).

Функції реалізуються наказовими реченнями. Послідовності наказових речень, тобто певні алгоритми, становлять активні знання, які в системі "Поступ" зберігаються, як і пасивні, у вербально-семантичному представленні. При цьому, однак, крім вербально-імперативного способу представлення (наказові речення в граматиці  $G$ ), використовується також звичний для запису алгоритмів операторний спосіб представлення активних знань, в якому виклик функції представляється її іменем з взятими в дужки і розділеними комами параметрами. Вводяться додаткові множини термінальних символів  $V_T$  і нетермінальних символів  $V_N$  та правил підстановки  $P$ , які доповнюють граматику  $G$  до граматики  $G_2$ , що включає операторне представлення. При цьому

$$V_N = \{ \langle \text{оператор} \rangle, \{ \langle \text{функція типу } \xi \rangle \mid \xi \in \Xi \} \} \cup \bar{Z}_o,$$

де  $\Xi$  - множина різних типів функцій,  $\bar{Z}_o$  - нетермінальна множина операторних змістів параметрів,

$$V_T = \{ \text{виконати}, \{ . \}, \{ . \} \} \cup \bar{\sigma} \cup \bar{\phi},$$

де  $\bar{\sigma}$  - множина імен операторів,  $\bar{\varphi}$  - множина імен функцій,  $\lfloor$ ,  $\rfloor$  - функційно-операторні розділювачі (дужки),  $\{$ ,  $\}$  - коментарні дужки.

$P'$ : <речення>  $\rightarrow$  виконати <оператор> .

<оператор>  $\rightarrow$  <оператор> <оператор>

{ <оператор>  $\rightarrow$   $\sigma$  ( $z_1(\sigma)$ , ...,  $z_{v(\sigma)}(\sigma)$ ) |  $\sigma \in \bar{\sigma}$ }

{  $z \rightarrow$  <функція типу  $\xi$ > |  $\xi \in \Xi$ ,  $z \in \bar{Z}_\varphi(\xi)$ }

{ <функція типу  $\xi$ >  $\rightarrow$   $\varphi$  ( $z_1(\varphi)$ , ...,  $z_{v(\varphi)}(\varphi)$ ) |  $\xi \in \Xi$ ,  $\varphi \in \bar{\varphi}(\xi)$  }

{  $z \rightarrow$  <оператор> |  $z \in \bar{Z}_\sigma$  }

{  $z \rightarrow \lfloor$  <оператор>  $\rfloor$  |  $z \in \bar{Z}_\varphi$  }

де  $z_j(\sigma)$  - зміст  $j$ -го параметра оператора,  $z_j(\sigma) \in \bar{Z}_\sigma \cup \bar{Z}_\varphi$ ,

$v(\sigma)$  - арність оператора,  $\bar{Z}_\varphi(\xi)$  - нетермінальна множина функційних змістів типу  $\xi$ ,

$\bar{Z}_\sigma$  - нетермінальна множина функційних змістів параметрів,  $z_j(\varphi)$  - зміст  $j$ -го параметра функції,

$v(\varphi)$  - арність функції,  $\bar{\varphi}(\xi)$  - множина функцій типу  $\xi$ .

Показується еквівалентність вербально-імперативного та операторного представлення активних знань.

Описуються особливості реалізації основних функцій у версії 3.1 системи "Поступ", зокрема представлення вивідної інформації в різних формах і способи організації діалогу системи з користувачем. В кінці третього розділу формулюються основні задачі герменевтичного та риторичного блоків, як передбачається реалізувати в наступних версіях системи.

Четвертий розділ містить опис застосування системи "Поступ" у двох суттєво різних галузях наукових досліджень. Когнітологічні дослідження територій України, проведені з допомогою системи, зв'язані з необхідністю оброблення великої кількості вербальної інформації, яка узагальнюється, групується і представляється в часовому та просторовому вимірах.

Дослідження вітроенергетичного потенціалу в регіонах України зв'язане зі значною базою активних знань, яка містить метеорологічні та географічні закономірності, емпіричні залежності, корекційні та інтерполяційні алгоритми. Часто ці знання є альтернативними між собою і їх застосування зв'язане з певними умовами. Тому в такому дослідженні важливою є можливість оперативного автоматизованого комбінування різних варіантів та швидкої апробації різних схем моделювання, що забезпечується системою. Для кожного з прикладів наводяться фрагменти моделей світу, ієрархічні структури тематичної інформації, зразки тематичних активних знань. Вказуються основні технічні характеристики системи при реалізації цих задач. Виділяються основні проблеми й альтернативні варіанти рішень, які виникли при апробації системи. Розв'язування задач з допомогою системи "Поступ" порівнюється з можливим розв'язуванням іншими методами.

Узагальненням досвіду використання системи "Поступ" є перелік основних переваг цієї системи порівняно з іншими засобами, а також основних напрямів її вдосконалення. Дається також характеристика класів задач, для яких використання системи найефективніше.

## ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

1. Сформульовано й обґрунтовано концепцію універсальної вербально-семантичної інформаційно-аналітичної системи для наукових досліджень з представленням знань природною мовою.

2. Запропоновано безконтекстну граматику, яка описує семантичне представлення будь-якої інформації, вираженої природною українською мовою, і досліджено її властивості.

3. Запропоновано спосіб представлення семантичної моделі довільних знань у формі узагальненого гіперграфа та досліджено властивості такого представлення.

4. Доведено семантичну інтерпретованість мови, породженої граматиною семантичного представлення, як однозначну відповідність між реченнями цієї мови і підграфами узагальненого гіперграфа.

5. Побудовано інформаційні структури, що необхідні для реалізації семантичного представлення природної мови й семантичної моделі знань в інформаційній системі, досліджено властивості цих структур, розв'язано ряд задач їх оптимізації.

6. Побудовано базову систему функцій вербально-семантичної інформаційно-аналітичної системи, досліджено її властивості, сформульовано ряд задач на семантичній моделі знань, що виникають при побудові та реалізації математичних моделей з допомогою такої системи, і запропоновано варіанти їх розв'язування.

7. Розроблено алгоритми, програмно реалізовано і апробовано вербально-семантичну інформаційно-аналітичну систему "Поступ" і на прикладах конкретних задач показано її придатність як основи для створення багатотематичних семантичних інформаційно-аналітичних систем з широкими можливостями для розв'язування задач математичного моделювання в різних галузях знань.

Основні результати дисертаційної роботи викладено в наступних публікаціях:

1. Жук П.А., Комиссарчук А.А. Моделирование сложных гидрологических систем с использованием данных дистанционного зондирования/ Информационные системы исследова-

- ния природных ресурсов.- Киев:Ин-т кибернетики АН УССР, 1982.- С. 3-8.
2. Жук П.А. Метод виделения структуры для моделирования непрерывных систем/ Теоретико-системные методы и их использование в автоматизированных системах.- Киев:Ин-т кибернетики АН УССР,1983.- С. 65-76.
  3. Гера Б.В., Жук П.А., Починайко Р.С., Чапля Е.Я. Принципы построения и перспективы развития геоинформационных систем для мониторинга природных ресурсов/ Прикладные проблемы моделирования и оптимизации (тез. докладов).- Москва:ВНИИ Системных исследований,1991.- С. 14-15.
  4. Жук П.А., Кравець В.Є. Концепція створення географічної інформаційної системи Карпатського регіону/ Карпати - український міст в Європу: проблеми і перспективи (тези доп. міжн. наук.-прак. конф.).- Львів:Ін-т економіки АН України,1993.- С. 78-79.
  5. Жук П.А., Кравець В.Є. Дослідження екологічно чистих джерел енергії в Карпатському регіоні/ Карпати - український міст в Європу: проблеми і перспективи (тези доп. міжн. наук.-прак. конф.).- Львів:Ін-т економіки АН України,1993.- С. 177-178.
  6. Жук П.А. Вербально-семантична інформаційно-аналітична система "Поступ": концепція і застосування/ Друга українська конференція з автоматичного керування "Автоматика-95": праці.- Львів:ФМІ НАН України,1995.- Т.3.- С. 22.
  7. Жук П.А. Семантична інтерпретація природної української мови в інформаційній системі/ Друга українська конференція з автоматичного керування "Автоматика-95": праці.- Львів:ФМІ НАН України,1995.- Т.3.- С. 23-24.

8. Жук П.А. Використання вербально-семантичної інформаційно-аналітичної системи для психоінформаційної класифікації регіонів України/ Друга українська конференція з автоматичного керування "Автоматика-95": праці.- Львів: ФМІ НАН України, 1995.- Т.3.- С. 24-25.
9. Жук П.А. Картографування ефективності вітроенергетичних пристроїв з допомогою аналітично-інформаційної системи/ Перша міжнародна науково-технічна конференція математичного моделювання в електротехніці й електроенергетиці (тези доп.).- Львів: Державний Університет "Львівська Політехніка, 1995.- С. 72-73.

Особистий вклад. Всі результати, що складають основний зміст дисертаційної роботи, отримані автором самостійно. В публікаціях, написаних у співавторстві, дисертантові належать: в роботах [1,3] - концепція інформаційних систем, в роботах [4,5] - постановка задач, системна концепція та наукове керівництво.

Жук.П.А. Вербально-семантическая информационно-аналитическая система для решения задач математического моделирования.

Дисертация на соискания ученой степени кандидата технических наук по специальностям 05.13.02 - математическое моделирование в научных исследованиях, 05.13.04 - автоматизированные системы управления и системы обработки информации. Физико-механический институт им.Г.В.Карпенко НАН Украины, Львов, 1996.

В диссертационной работе разработана новая концепция вербально-семантической информационной системы, обеспе-

чиваючої семантичну інтерпретацію і обробку произвольної багатотематическої вербальної інформації для цілей математического моделювання. Предложена і досліджена граматики описання знань і метод її семантическої інтерпретації. Розробані алгоритми семантическої обробки знань в інформаційній системі, котрі програмно реалізовані і апробіровані для рішення задач математического моделювання.

Zhuk P.A. Verbally-semantic Informativ-Analytic System for Mathematical Modelling Problems Solutions.

Dissertation for obtaining of scientific degree of candidate of sciences (engineering) on the specialities 05.13.04 - the mathematical modelling in scientific research, 05.13.04 - the automated control systems and information processing systems. Physical and Mechanical Institute of National Academy of Sciences of Ukraine, Lviv, 1996.

New verbally-semantic informativ system's conception which has ability to support semantic interpretation and processing of multithematic verbal information for mathematical modelling purposes is performed in dissertation. Knowledges description grammar and its semantic interpretation method was proposed and investigated. Knowledges semantic processing algorithms for informativ system are performed corresponding software was used for solution of the mathematical modelling problems.

Ключові слова: математичне моделювання, інформаційна система, семантичний аналіз, програмне забезпечення, природна мова, база знань, формальна граматики, математична лінгвістика.

Ротапринт ЛьЦНТЕІ Замовлення 19 тираж 100



1920. H. A.

1920. H. A.

AB 71.094