

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

КОЧУЄВА ЗОЯ АНАТОЛІЇВНА

УДК 519.7:007.52

**МЕТОДИ І МОДЕЛІ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ
ІНФОРМАЦІЙНИХ ОБ'ЄКТІВ У СУЧАСНИХ БІБЛІОТЕЧНИХ
СИСТЕМАХ**

Спеціальність 05.13.06 – інформаційні технології

Автореферат дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2014

Дисертацією є рукопис.

Робота виконана на кафедрі інтелектуальних комп'ютерних систем Національного технічного університету «Харківський політехнічний інститут» Міністерства освіти і науки України.

Науковий керівник доктор технічних наук, професор
Шаронова Наталія Валеріївна,
Національний технічний університет
«Харківський політехнічний інститут»,
завідувач кафедри інтелектуальних
комп'ютерних систем

Офіційні опоненти: доктор технічних наук, професор
Шабанов-Кушнарєнко Сергій Юрійович,
Харківський національний університет
радіоелектроніки, провідний науковий співробітник
кафедри програмної інженерії

доктор технічних наук, доцент
Шостак Ігор Володимирович,
Національний аерокосмічний університет
ім. М.Є. Жуковського «Харківський
авіаційний інститут», професор кафедри інженерії
програмного забезпечення

Захист відбудеться _____ 2014 р. о 14.30 годині на засіданні спеціалізованої вченої ради Д 64.050.07 в 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» за адресою: 61002, м. Харків, вул. Фрунзе, 21.

Автореферат розісланий « ____ » 2014 р.

Вчений секретар
спеціалізованої вченої ради

В. П. Северин

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. З впровадженням інформаційних технологій, з розвитком мережі електронних бібліотек все більшої актуальності набувають задачі переробки інформації, яка міститься у бібліотеках, розташованої як на традиційних носіях у вигляді книг, газет, журналів, так і представленої в електронному вигляді на сучасних носіях інформації. Сучасна бібліотека – це знання-орієнтована інформаційна система, яка оперує природно-мовними інформаційними об'єктами, що зберігаються у текстовому вигляді, але недостатня кількість перевірених на практиці теоретичних розробок та ефективних програмно-апаратних систем комп'ютерної обробки таких об'єктів складає певну проблему. Аналіз досліджень та публікацій показав, що за останні десять років досягнуто значних успіхів у розробці спеціалізованих бібліотечних інформаційних систем, але вони не повністю задовольняють зростаючі потреби користувачів, оскільки не забезпечують певної відповідності читацьких запитів та виданої інформації. Перспективним завданням сьогодення є створення інформаційних технологій для розробки моделей, методів та інструментальних засобів автоматизації бібліотек та архівів у вигляді електронних каталогів, автоматизованих робочих місць, систем автоматизованої переробки документів тощо.

Основою створення автоматизованих інформаційних бібліотечних систем (АІБС) є розробка математичного і лінгвістичного забезпечення, а саме: моделей, алгоритмів і методів, які охоплюють процеси предметизації, каталогізації, анотування та реферування, при цьому проблеми, що розглядаються, можна класифікувати таким чином: ідентифікація, представлення та обробка знань у АІБС; розробка природномовного інтерфейсу АІБС із залученням методів та засобів комп'ютерної лінгвістики; імітація процесу «розуміння» при аналітико-синтетичній обробці документу, для чого необхідним є залучення знань з моделювання інтелектуальних функцій людини.

У межах окресленого напрямку важливими є наукові завдання розробки моделей, методів, алгоритмів та програм, які здійснюють моделювання процесів інтелектуальної обробки інформаційних об'єктів з метою визначення їх основних характеристик для побудови інформаційного, математичного, лінгвістичного і програмного забезпечення АІБС. У вирішенні задачі ідентифікації знань істотний внесок внесли вчені В. М. Глушков, А. К. Жолковський, Ю. М. Марчук, О. В. Палагін, Д. О. Поспелов, Р. Ш. Рубашкін, Ч. Дж. Філлмор, Р. Шенк, Я. Л. Шрайберг та інші.

Теоретичні аспекти комп'ютерної лінгвістики, моделювання природної мови та інтелектуальної обробки текстів знайшли відображення у працях Ю. Д. Апресяна, М. Ф. Бондаренка, Ф. С. Воройського, І. В. Замаруєвої, О. В. Зубова, Н. Ф. Хайрової, Ю. П. Шабанова-Кушнаренка, Н. В. Шаронової, В. А. Широкова та інших.

Усе вищезазначене обумовлює актуальність розвитку моделей та методів інтелектуальної обробки даних і застосування цих методів для автоматизованої обробки інформаційних об'єктів у сучасних бібліотечних системах, що складає напрямок дисертаційної роботи.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана на кафедрі інтелектуальних комп'ютерних систем НТУ «ХПІ» у межах держбюджетних тем МОН України «Розробка математичних моделей та методів розв'язання задач інтелектуальної обробки інформації» (ДР № 0108U003926), «Розробка моделей та методів для інформаційно-пошукових, лексикографічних інтелектуальних систем» (ДР № 0111U002258), у яких здобувач брала участь як виконавець.

Мета і задачі дослідження. Метою дисертаційної роботи є удосконалення ефективності роботи сучасної бібліотеки за рахунок використання інтелектуальних методів і моделей обробки інформаційних об'єктів. Відповідно до зазначеної мети поставлено такі задачі:

1) виконати аналіз методів і моделей представлення і обробки знань, ідентифікації інформаційних об'єктів при автоматизації інформаційно-бібліотечних систем і сформулювати основні вимоги до розробки відповідного математичного та лінгвістичного забезпечення;

2) розробити математичні та лінгвістичні засоби для розв'язання задач обробки текстових документів на основі моделювання лінгвістичної діяльності людини та інтелектуального аналізу даних методом компараторної ідентифікації;

3) розробити засоби моделювання процедур аналізу і синтезу, стиску та узагальнення текстової інформації для застосування у задачах анотування та реферування повнотекстових документів;

4) удосконалити моделі процесів предметного групування, індексації, систематизації, предметизації, рубрикації повнотекстових документів у бібліотечних системах;

5) розробити модель бази знань інформаційної бібліотечної системи на основі побудови семантичної мережі понять зі сталими зв'язками між ними;

б) виконати практичну реалізацію запропонованих методів і математичних моделей, впровадити результати дисертаційної роботи у практику створення інформаційних бібліотечних систем.

Об'єктом дослідження є інформаційні об'єкти і процеси в автоматизованих інформаційних бібліотечних системах.

Предметом дослідження є методи і моделі інтелектуальної обробки інформаційних об'єктів і процесів.

Методи дослідження засновані на комплексному використанні теорії інтелекту, алгебри скінчених предикатів та предикатних операцій, методу компараторної ідентифікації, методів штучного інтелекту для розробки алгебрологічних моделей обробки інформаційних об'єктів, які представлено текстовою формою. Алгебра предикатів та предикатних операцій використовується для формалізації знань, опису природно-мовних відношень та моделювання процесів анотування та реферування повнотекстових документів. Метод компараторної ідентифікації використовується для опису інтелектуальних функцій людини при аналітико-синтетичній обробці документів.

Наукова новизна отриманих результатів визначається наступним:

– вперше запропоновано та програмно реалізовано метод компараторної ідентифікації інформаційних об'єктів сучасної бібліотеки, що дозволило сформувати

теоретичну базу для представлення й обробки текстових документів в АІБС;

– дістали подальшого розвитку логіко-алгебраїчні методи і моделі, які є основою математичного і лінгвістичного забезпечення АІБС та базуються на методі компараторної ідентифікації, що дозволяє переходити від опису суб'єктивних станів людини при обробці текстової інформації до об'єктивних математичних моделей;

– удосконалено моделі процесів угруповання, індексації, систематизації, предметизації та рубрикації повнотекстових документів за рахунок створення інформаційної технології обробки повнотекстових документів, що дозволяє підвищити точність, швидкість та релевантність пошуку документів.

Практичне значення отриманих результатів. На основі розроблених у дисертаційному дослідженні методів і моделей інтелектуальної обробки інформаційних об'єктів і процесів у сучасній бібліотеці запропоновано інформаційну технологію формалізації процедур систематизації та предметизації повнотекстових документів, анотування й реферування документів, сформульовано критерії ефективності інформаційного пошуку в електронній бібліотеці, запропоновано розв'язання задачі книгозабезпеченості. Математичні результати роботи можуть бути використані в системах автоматичної обробки природної мови, при розробці різних інформаційно-пошукових, експертних, аналітичних засобів інформаційних систем широкого призначення.

Результати дисертаційного дослідження знайшли практичне застосування при вирішенні задачі індексації повнотекстових документів ключовими словами, систематизації й рубрикації документів в науковій бібліотеці Харківського національного медичного університету (м. Харків), науковій бібліотеці Харківського національного університету радіоелектроніки (м. Харків), науковій бібліотеці Національного технічного університету «ХПІ» (м. Харків). Результати, отримані у роботі, використані при розробці автоматизованої системи підтримки інформаційного забезпечення. Застосування розроблених моделей і методів дозволило підвищити якість аналітико-синтетичної обробки інформаційних об'єктів у сучасній бібліотеці. Теоретичні результати дисертації використовуються в навчальному процесі на кафедрі інтелектуальних комп'ютерних систем НТУ «ХПІ» при викладанні спеціальних дисциплін «Математичні основи лінгвістики», «Інформаційно-ресурсне забезпечення лінгвістичної діяльності», «Штучний інтелект: лінгвістичні проблеми», «Автоматизована обробка природної мови» для спеціальності «Прикладна лінгвістика». Програмне забезпечення, розроблене у дисертації, використовується при виконанні курсових й дипломних робіт на кафедрі інтелектуальних комп'ютерних систем НТУ «ХПІ».

Особистий внесок здобувача. Усі основні результати дисертаційної роботи, що виносяться на захист, отримані здобувачем особисто, серед них: підхід до використання методу компараторної ідентифікації для обробки текстової інформації в АІБС; використання алгебри предикатів та предикатних операцій для представлення знань; математична модель аналітико-синтетичної обробки інформації у задачах анотування та реферування повнотекстових документів; алгоритм для автоматизованого індексування повнотекстових документів ключовими словами; удосконалення моделі процесу систематизації та предметиза-

ції повнотекстових документів у бібліотеці.

Апробація результатів дисертації. Результати дисертаційної роботи доповідались та обговорювались на: Міжнародних конференціях «Інформаційні технології в освіті і управлінні» (Нова Каховка, 2006, 2007, 2012), Міжнародному молодіжному форумі «Радіоелектроніка й молодь в 21 столітті» (Харків, 2008), Міжнародній науково-практичній конференції «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я» (Харків, 2012), Міжвузівських наукових конференціях «Інтелектуальні системи і комп'ютерна лінгвістика» (Харків, 2008–2011), Всеукраїнських науково-практичних конференціях «Інтелектуальні системи і прикладна лінгвістика» (Харків, 2012, 2013).

Публікації. Основні результати дисертації опубліковані у 18 наукових працях, серед яких 11 статей у фахових наукових виданнях України (1 – у наукометричному виданні), 7 публікацій у матеріалах конференцій.

Структура та обсяг дисертації. Дисертаційна робота складається зі списку використаних скорочень, вступу, п'ятьох розділів, висновків, додатків, списку використаних джерел. Повний обсяг дисертації складає 181 сторінку, з них 27 рисунків по тексту, 3 таблиці по тексту, 3 додатки на 15 сторінках, 163 найменування використаних науково-технічних джерел на 16 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми дисертації, зазначено зв'язок роботи з науковими темами, сформульовано мету і задачі дослідження, визначено об'єкт, предмет і методи дослідження, показано наукову новизну та практичне значення отриманих результатів, наведено інформацію про практичне використання, апробацію результатів та їх висвітлення у публікаціях.

У **першому розділі** на основі аналітичного аналізу наукових праць проведено аналіз сучасного стану інформатизації бібліотек, досліджено основні підходи до побудови електронних бібліотек, зроблено огляд досліджень у галузі моделювання змісту інформаційних об'єктів. Наведено дослідження стану сучасних інформаційно-бібліотечних технологій та математичного і лінгвістичного забезпечення АІБС; зроблено огляд існуючих задач у бібліотечній діяльності, пов'язаних з представленням знань і обробкою текстової інформації. Проаналізовано основні типи моделей представлення знань, розглянуто основні методи ідентифікації знань. Використання сучасних методів ідентифікації знань дозволяє ефективніше вилучати знання з бібліотечних електронних сховищ. Обґрунтовано необхідність розвитку математичного апарату моделювання інтелектуальних функцій людини для формалізації процедур оброблення повнотекстових документів в АІБС.

Більшість бібліотечних та бібліографічних інформаційних процесів характеризуються збереженням та обробкою масивів неформалізованої та слабо формалізованої інформації, при цьому невелика кількість існуючих систем адаптована до російської або української мов. Для такого роду інформації типовими є процеси автоматичної обробки документів: аналітико-синтетична обробка, класифікація, предметизація, на яких базується організація інформаційного пошуку користувачами бібліотечних систем. Метою такого пошуку може бути

знаходження факту, документу чи відомостей про його присутність, або відшукування елементів знання (змістовних елементів) з наступним їх поданням користувачеві. Ефективність інформаційного пошуку в АІБС значною мірою залежить від її лінгвістичного забезпечення.

Дослідження основних підходів до побудови електронних бібліотек дозволило дійти висновку, що однією з задач, яка постає перед дослідниками та розробниками електронних бібліотек, є відображення існуючої різноманітності електронних колекцій у їх однорідне представлення та забезпечення доступу до інтегрованої колекції неоднорідної інформації. Найбільш актуальними завданнями переробки повнотекстових баз даних залишаються завдання, які пов'язані з розумінням смислу текстового повідомлення: систематизація, реферування, анотування, корекція помилок й т. і. Інформаційний простір електронної бібліотеки може бути представлений у вигляді сукупності самостійних інформаційних об'єктів (електронних документів). Під інформаційним об'єктом у сучасній бібліотеці розуміється закінчений твір, який має автора та дозволяє однозначну ідентифікацію й не втрачає свого інформаційного значення поза конкретною інформаційною системою.

Відзначено, що практично всі задачі автоматизації бібліотечної діяльності пов'язані з переробкою текстів, поданих природною мовою, тому впровадження у практику математичних моделей оброблення природномовної інформації приводить до суттєвого підвищення ефективності роботи таких систем. Оброблювані бібліотечними системами інформаційні об'єкти є дискретними, скінченими та детермінованими, що дозволяє використовувати при обробці цих об'єктів засоби алгебри скінчених предикатів і метод компараторної ідентифікації. Аналіз лінгвістичного забезпечення АІБС показав, що лишаються актуальними завдання побудови моделей аналітико-синтетичної обробки текстової інформації.

На підставі критичного аналізу існуючих невирішених проблем та задач автоматизації бібліотечної діяльності сформульовано й обґрунтовано необхідність дослідження моделей та методів інтелектуальної обробки інформаційних об'єктів для впровадження їх в автоматизовані інформаційні бібліотечні системи.

Другий розділ дисертації присвячено розробці математичних засобів для моделювання інтелектуальної діяльності людини при роботі з інформаційними об'єктами як основи побудування інформаційної технології. Визначено математичний інструментарій теорії інтелекту для ідентифікації інформаційних об'єктів, представлених природномовними текстами. Метод порівняння, або компараторної ідентифікації, знайшов застосування при розв'язанні багатьох завдань моделювання лінгвістичних закономірностей, а саме: показано застосування методу компарації при проведенні лінгвістичного експерименту на документах повнотекстової бази даних. Показано використання ситуаційно-текстового предикату при вирішенні задач аналізу й синтезу текстової інформації для анотування й реферування інформаційних об'єктів. Описано декомпозицію ситуаційно-текстового предикату. Досліджені функції сприйняття, розуміння й усвідомлення як основні частини ситуаційно-текстового предикату.

Можливість використання теорії інтелекту, виходячи виключно з фізично спостережуваних фактів, заснована на методі аксіоматичного опису розуму людини. Сутність методу полягає у тому, що випробуваний у спеціально поставлених

P_1, P_2, \dots, P_l позначені предикати від предикатів X_1, X_2, \dots, X_r . Предикат $X_j(x_1, x_2, x_n), j \in \{1, 2, \dots, r\}$ заданий на декартовому добутку $A_{1j} \times A_{2j} \times \dots \times A_{nj}$. Рішення $X_1 = P_1, X_2 = P_2, \dots, X_r = P_r$ задовольняє системі рівнянь (1).

Будь-якому відношенню Q у взаємно однозначну відповідність ставиться предикат, який на мові алгебри предикатів записується у формульному вигляді з використанням базисних предикатів 0, 1 та предикату упізнання предмету x_i^a , а також базисних операцій кон'юнкції \wedge та диз'юнкції \vee . Предикат упізнання предмета a , який знаходиться на місці x_i , визначається формулою:

$$x_i^a = \begin{cases} 1, & x_i = a, \\ 0, & x_i \neq a. \end{cases}$$

При розв'язанні задачі опису інформаційних об'єктів важливу роль відіграє процедура введення ознак та їх значень, які є предметними змінними при побудові моделей. Для перевірки правильності введення ознак використовуються наступні вирази. Для скінченного універсуму $U = \{a_1, a_2, \dots, a_k\}$ виконуються:

– закон істинності

$$x_i^{a_1} \vee x_i^{a_2} \vee \dots \vee x_i^{a_k} = 1, i = \overline{1, n};$$

– закон заперечення

$$\overline{x_i^{a_j}} = x_i^{a_1} \vee x_i^{a_2} \vee \dots \vee x_i^{a_{j-1}} \vee x_i^{a_{j+1}} \vee \dots \vee x_i^{a_k}, i = \overline{1, n}, j = \overline{1, k};$$

– закон хибності

$$x_i^{a_j} x_i^{a_l} = 0, i = \overline{1, n}.$$

Кожному відношенню P на U^n відповідає свій предикат P на U , і навпаки. Наявність такої взаємно однозначної відповідності між відношеннями і предикатами дає змогу математично записувати будь-яке відношення у вигляді деякої формули алгебри предикатів, за якою обчислюються значення предикату, відповідного цьому відношенню. Доведено, що будь-який t -арний предикат T від предикатів P_1, P_2, \dots, P_t можна виразити формулою

$$T(P_1, P_2, \dots, P_t) = \vee P_1^{A_1} P_2^{A_2} \dots P_t^{A_t}.$$

З використанням наведеного інструментарію моделюються природномовні відношення, які спостерігаються при описі інформаційних об'єктів. Оскільки об'єкти, які оброблюються бібліотечними системами, є дискретними, скінченими і детермінованими, це дозволяє використовувати при їхній обробці метод компараторної ідентифікації. Сутність методу представлено схемою на рис. 1.

На вхід системи подається множина сигналів x_1, x_2, \dots, x_n , під якими розуміються умовні знаки, що служать для передачі інформації (тексти документів, ключові поняття тощо). Вхідні сигнали беруться зі скінчених множин X_1, X_2, \dots, X_n , при цьому $x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n$. У результаті роботи системи, яка здійснює обробку текстової інформації, на вихід поступає певна множина

елементів y_1, y_2, \dots, y_n . В нашому випадку y_1, y_2, \dots, y_n означають ключові поняття, дескриптори, символи УДК, ББК і т.д. При цьому $y_1 \in Y_1, y_2 \in Y_2, \dots, y_n \in Y_n$ – множина вихідних алфавітів.

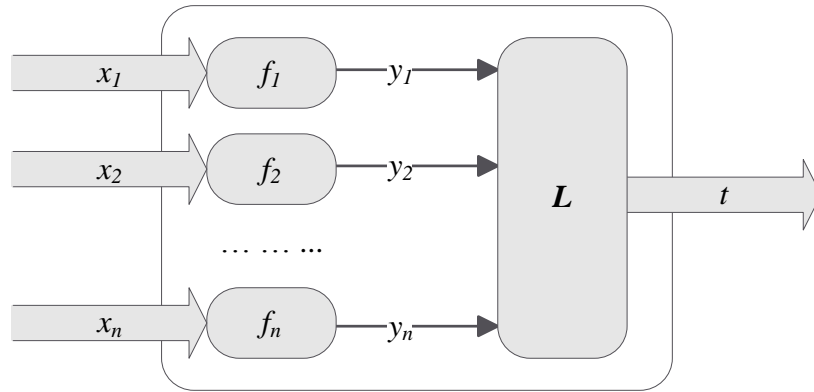


Рисунок 1. – Сутність методу компараторної ідентифікації

Предикат $P(x_1, x_2, \dots, x_n) = Q(f_1(x_1), f_2(x_2), \dots, f_n(x_n))$ характеризує роботу системи, що здійснює інтелектуальну бібліографічну обробку документальної інформації, яка на сигнали x_1, x_2, \dots, x_n реагує відповіддю $q = P(x_1, x_2, \dots, x_n)$. Моделювання будь-якої із задач аналітико-синтетичної обробки інформації полягає у тому, щоб із властивостей предиката P , який здійснює компараторну ідентифікацію інформаційних об'єктів, вилучити внутрішню структуру сигналів x_1, x_2, \dots, x_n , елементів y_1, y_2, \dots, y_n , вид функцій f_1, f_2, \dots, f_n і вид предиката Q .

Наведено застосування методу компараторної ідентифікації для обробки документів у повнотекстовій базі даних, зокрема, для розв'язання задачі предметного групування текстів повнотекстової бази даних.

Показано використання ситуаційно-текстового предиката при вирішенні задач аналізу і синтезу текстової інформації для анотування і реферування інформаційних об'єктів. Розглянуто декомпозицію ситуаційно-текстового предиката на функції сприйняття, розуміння й усвідомлення, що дозволяє здійснювати лінгвістичний експеримент і вирішувати задачі суто фізичним методом, без залучення суб'єктивних даних. Метод компараторної ідентифікації дозволяє здійснити автоматичне розбиття на класи еквівалентності текстів документів повнотекстових баз даних, тотожних по відношенню до певної підтеми, і розбиття ключових термінів, що відносяться до даної предметної області, на класи еквівалентності, які представляють певну підрубрику. Застосування даного методу в АІБС дозволяє вести як вузько тематичний, так і багатоаспектний пошук документів у повнотекстових базах даних, автоматизувати процес рубрикації ключових термінів, який є складним семантичним процесом, що потребує спеціальних знань у даній предметній області.

У **третьому розділі** розглядається задача інтелектуальних інформаційних систем – проблема представлення знань. Запропоновано модель бази знань АІБС на основі побудови семантичної мережі понять, які записано мовою алгебри скінчених предикатів. Модель предметної області містить концепти понять, об'єктів і відношень з області комп'ютерних технологій і мережі Інтернет. Розглянуто задачу автоматизованого індексування повнотекстових документів ключовими словами та запропоновано підхід до знаходження ключових слів у доку-

ментах, заснований на морфологічному, синтаксичному й статистичному аналізі тексту. Запропоновано розв'язання задачі формалізації процедур систематизації й предметизації повнотекстових документів. Для поєднання в групи тематично однакових документів й розподілу нових надходжень застосовується метод компарації. Уведення дескрипторно-текстового предиката дозволяє моделювати процедури систематизації й класифікації інформаційних об'єктів.

Інформаційній бібліотечній системі, яка розв'язує задачу фактографічного та бібліографічного (документального) пошуку, необхідна база знань з розвинутою системою парадигматичних зв'язків, що відображають основні співвідношення понять в області знань. Відомості про такі відношення використовуються при логіко-семантичних перетвореннях інформаційного запису, необхідних для побудови відповідей на запити, які звернені до даної інформаційної системи.

Модель предметної області містить концепти понять, об'єктів і відношень з області комп'ютерних технологій та мережі Інтернет. Область знань, що розглядається, достатньо добре структурована, тому вона адекватно відображається ієрархічною семантичною мережею (ІСМ). Модель, заснована на використанні ІСМ, спирається на принципи організації людської пам'яті. Під семантичною мережею розуміють систему знань, що представлена у вигляді цілісного образу мережі, вузли якої відповідають поняттям і об'єктам, а дуги – відношенням між об'єктами. ІСМ використано як засіб представлення семантичних відношень між концептами. На першому кроці для створення ієрархічної системи знань виділяються основні об'єкти предметної області, їх чисельність у розробленій моделі 34. Після цього групи близьких об'єктів композуються у об'єкти другого рівня абстракції, а об'єкти, які можна розбити на більш дрібні, – декомпозуються. У результаті таких дій у цій моделі залишилося 497 об'єктів.

Поняття у даній моделі реалізують відношення трьох типів: «бути частиною» (PART-OF); «є» (IS-A), «має» (HAS). Крім цих трьох основних типів відношень у семантичних мережах використовуються й похідні цих відношень. Для відображення ієрархічних відношень між точками зіткнення концептів, а також для встановлення зв'язків між вузлами, які показують концепти та їх екземпляри, використовуються відношення IS-A. Відношення IS-A передає наслідування атрибутів між рівнями ієрархії, тобто відношення IS-A є відношенням включення або збігу понять. Спосіб включення можна назвати поняттям верхнього рівня, а спосіб видалення – поняттям нижнього рівня. Екземпляр нижнього рівня містить в основному всі атрибути, які має екземпляр поняття верхнього рівня. Прикладом відношення даного типу є відношення між поняттями *A195 – мультискануючий монітор* і *A191 – монітор*:

*A195 мультискануючий монітор IS-A A191 – монітор,
«мультискануючий монітор – це різновид монітору».*

Такий спосіб представлення семантичної мережі дозволяє здійснювати вивід завдяки ієрархії наслідування. При цьому проявляється головна перевага представлення знань за допомогою ІСМ – зменшення кількості інформації, необхідної для збереження у системі. Факт, який оголошено для вершин на верхньому рівні ієрархічної структури, на основі припущення про справедливості його для вузлів нижнього рівня, показує можливість виводу множини фактів за

допомогою відношення IS-A. Вивід такого типу називається *наслідуванням властивості*, а гілка IS-A називається *гілкою наслідування властивості*. Ця можливість проявляється за рахунок представлення атрибутів об'єкта.

Універсум U розглянутої задачі представлено множиною всіх понять, які входять у базу знань системи. При переході від графічного зображення до скінчених предикатів введемо два бінарні предикати: предикат $P(x_1, x_2)$ описує відношення виду «частина – ціле» (PART-OF): він дорівнюється 1, якщо x_1, x_2 знаходяться у відношенні «частина – ціле», і дорівнюється 0 у протилежному випадку. Предикат $Q(x_1, x_2)$ описує відношення виду «є» (IS-A): він дорівнює 1, якщо x_1, x_2 знаходяться у відношенні «рід – вид», і 0 у протилежному випадку. Тут x_1, x_2 – набори літерних змінних, які задані на множині $A = \{a_1, a_2, \dots, a_{497}\}$.

Кожна пара вершин мережі упорядкована одним із відношень, яке відповідає застосованим у моделі предикатам, кожному предикату відповідає дуга графа, що з'єднує вершини, у яких містяться його змінні (або їх значення). У предикатному запису $P(x_1, x_2)$ і $Q(x_1, x_2)$ два елементи x_1, x_2 знаходяться у деякому відношенні залежності, коли один є основним (x_1), другий – підлеглим (x_2). Приклад графічного представлення подібних відношень показаний на рис. 2.

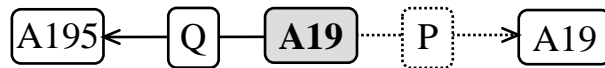


Рисунок 2. – Графічне представлення відношень

Побудована семантична мережа являє собою орієнтований граф, який не має петель, тобто не може бути одночасно $P(x_1, x_2) = 1$ і $Q(x_1, x_2) = 1$, де x_1 і x_2 – елементи мережі; і також не має кратних дуг, тобто не може бути одночасно вірно $P(x_1, x_2) = 1$ і $P(x_2, x_1) = 1$. Така модель ПО дозволяє отримати семантичне представлення тексту, який відноситься до даної предметної області (рис. 3).

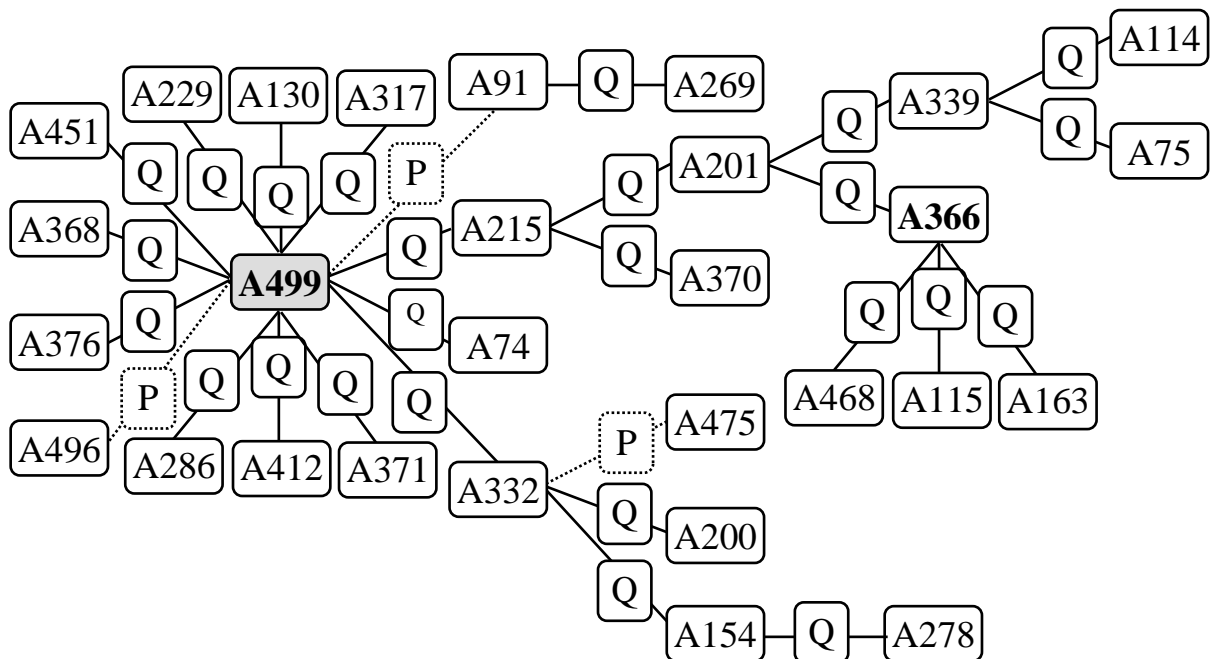


Рисунок 3. – Графічне представлення відношень між поняттями предметної області запам'ятовуючих пристроїв

Для процесів систематизації та предметизації є формування пошукового образу документів у вигляді набору простих і складних індексів та предметних рубрик. Сукупність алгоритмів процедур співставлення документів із рубриками являє собою математичне та програмне забезпечення процесів предметизації і систематизації бібліотечної системи. Автоматизація цих процесів дозволяє поєднати продуктивність з глибиною і стійкістю (послідовністю) класифікації. Задачею систематизації та предметизації є визначення рубрик класифікаційної схеми або рубрикатора, які відповідають змісту конкретного тексту.

На першому етапі розв'язання даної задачі складається пошуковий образ документа, який полягає у виділенні основного змісту документа. Другий етап систематизації і предметизації зводиться до прийняття рішення про рівнозначність смислів стислого документа і рубрики. При цьому рубрики будуть відігравати роль запитів, які записані на пошуковій мові дескрипторного типу. Співставлення документів і рубрик, які записані на інформаційно-пошуковій мові дескрипторного типу за допомогою алгоритму пошуку, дозволяє моделювати другий етап систематизації або предметизації. На третьому етапі здійснюється автоматичне присвоєння документам рубрик або індексів у відповідності з прийнятим критерієм видачі документів, яким може бути коефіцієнт збіжності. Даний метод дозволяє алгоритмізувати і, таким чином, автоматизувати процес традиційної систематизації і предметизації. Принцип роботи рубрикатора показаний на рис. 4.

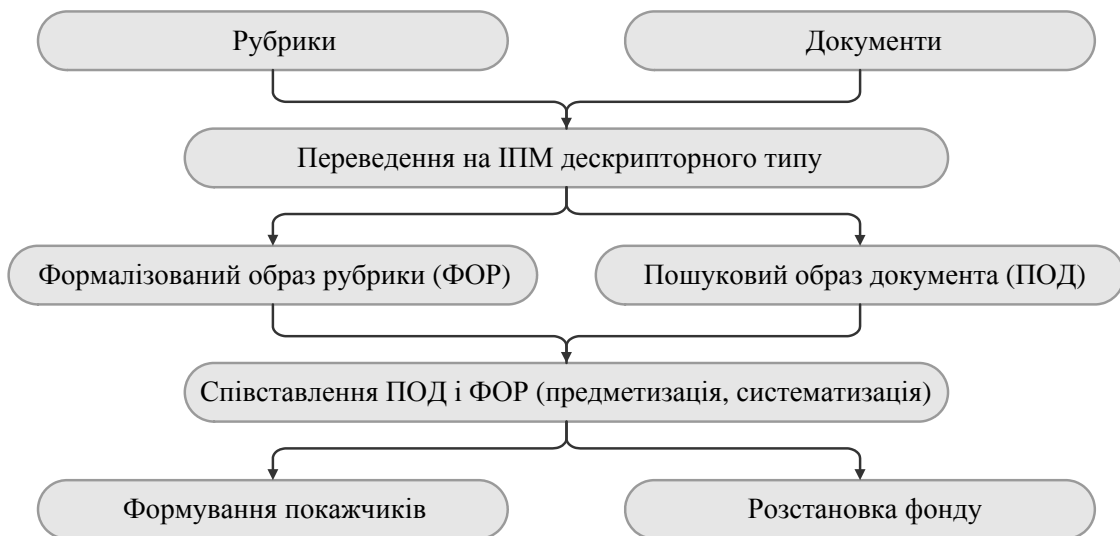


Рисунок 4. – Принципова схема роботи рубрикатора

Для складання формалізованих образів рубрик проводиться дослідження фонду документів з метою визначення множини R ключових слів фонду. Далі визначені ключові слова у нормалізованому вигляді розподіляються по підмножинам R_1, R_2, \dots, R_m , де m – кількість тематичних рубрик або предметів рубрикатора. Ключові слова класифікуються за групами згідно з тематичними аспектами або зв'язками між ними. Для об'єднання у групи тематично однакових документів і розподілу поточних бібліотечних постачань за тими класами, які є у даний момент часу, з урахуванням термінів, які містяться у них, використовується метод компараторної ідентифікації. Ця модель вдосконалює не тільки

процес систематизації, а й процес інформаційного пошуку у бібліотечно-пошуковій системі.

Четвертий розділ присвячено розв'язанню основних завдань інтелектуальної обробки інформації у сучасних АІБС. Розглянуто основні принципи аналітико-синтетичної обробки інформації, такі як стиснення й узагальнення, при розв'язанні задач анотування й реферування повнотекстових документів. Наведено програмну реалізацію задачі автоматичного реферування.

В основі існуючих систем автоматичного реферування лежать здебільшого статистичні й позиційні методи, які зводяться до автоматичного екстрагування окремих речень, які тематично і логічно не пов'язані між собою. Складовими процесу автоматизованого реферування є аналіз, узагальнення й синтез. Процес узагальнення є найбільш складним, оскільки передбачає згортання змісту шляхом пошуку найбільш ємних засобів і форм представлення інформації. Зміст реферату має залишатися семантично адекватним і еквівалентним первісному документу. Основне завдання реферування полягає у переході від розгорнутої смислової структури тексту до узагальненої смислової структури реферату в результаті виділення нової й найважливішої інформації. На рис. 5 наведено загальну схему процедури, яка здійснюється людиною у ході реферування.

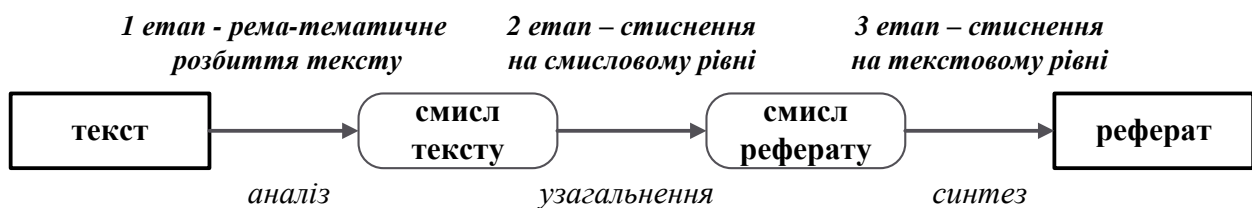


Рисунок 5. – Основні етапи реферування

Моделювання процесу реферування зводиться до побудови моделі реферату як набору типових для індикативних рефератів синтактико-семантичних конструкцій (структур) із загальним значенням і з наступним змістовним конструюванням реферату: наповнення синтаксем, які входять до моделі, поняттями певних класів. Описано процес автоматичного реферування тексту, запропоновано алгоритм реферування, створено його експериментальну програмну реалізацію «Автореферат» (рис. 6). Алгоритм автореферування містить три основні етапи: знаходження слова з найбільшою частотою вживання; знаходження ланцюжка зі слів в оточенні слова з найбільшою частотою вживання; формування тексту результату.

Досліджено проблеми ранжування і релевантності повнотекстових документів у інформаційному пошуку. Сформульовано основні завдання інформаційного пошуку у сучасній бібліотеці, обґрунтовано вибір основних критеріїв оцінки релевантності тексту. Досліджено теоретико-множинні, ймовірнісні та алгебраїчні моделі ранжування, наведено їх порівняльні характеристики.

П'ятий розділ присвячено практичній реалізації отриманих результатів. Наведено практичне впровадження результатів дослідження у діяльність Наукової бібліотеки Харківського національного медичного університету (ХНМУ). Розроблено математичні моделі семантичної структури хімічних сполучень, які застосовано для обробки текстів медичної та хімічної тематики. Проведено оці-

нку ефективності запропонованих семантичних моделей і методів, проаналізована особливість рубрикації медичних повнотекстових документів. Показано перспективи отриманих моделей для обробки повнотекстової інформації.

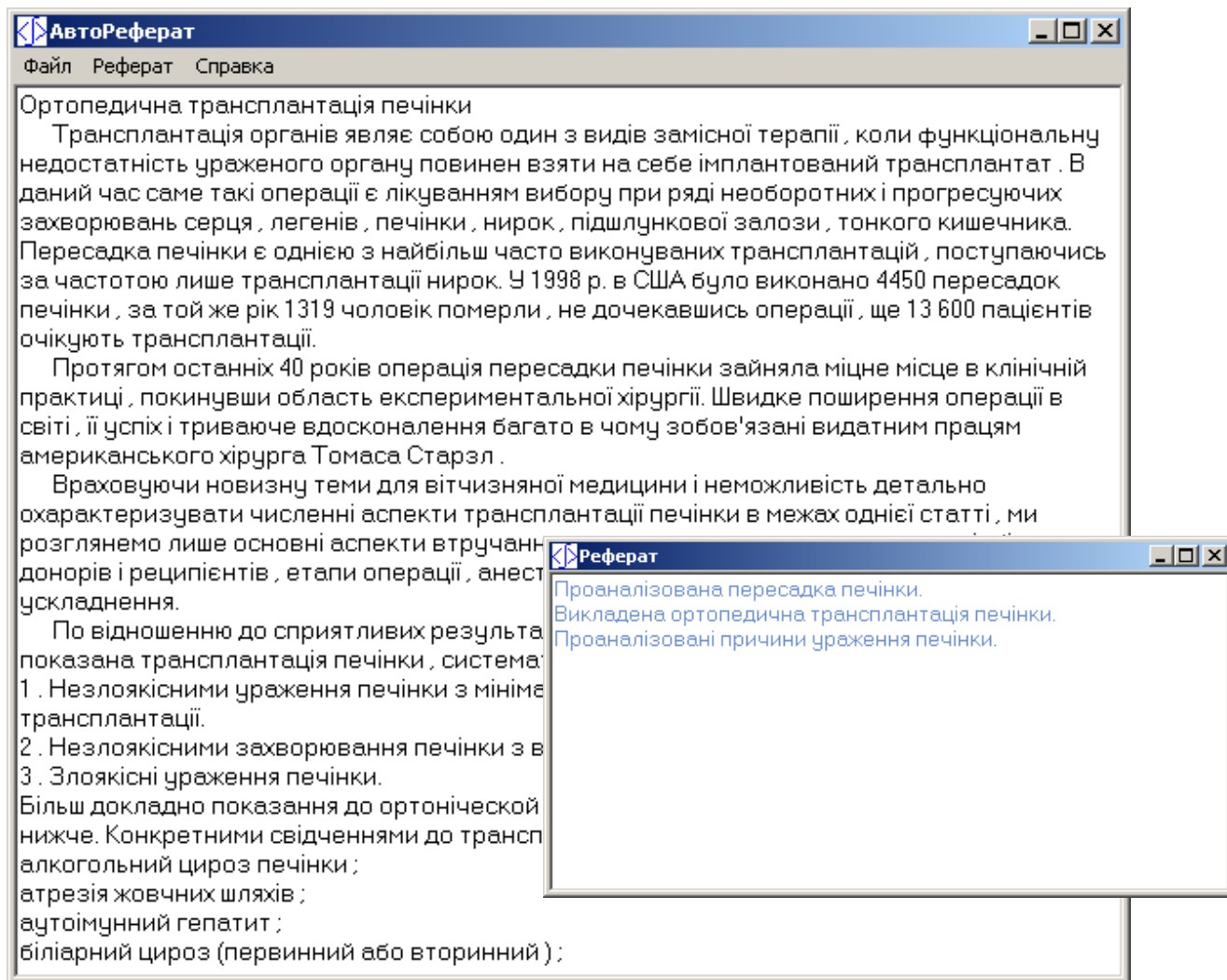


Рисунок 6. – Програмна реалізація алгоритму реферування

Розроблена програма для рубрикації медичних повнотекстових документів реалізована за допомогою методу компараторної ідентифікації. Текстовими даними була колекція україномовних та російськомовних повнотекстових документів: електронні ресурси наукової бібліотеки ХНМУ. Приклад інтерфейсу програми представлено на рис. 7.

Оцінка ефективності рубрикації документів у електронній бібліотеці виконана шляхом обчислення мір якості отриманих результатів класифікації (рубрикації). Запропоновано використання чотирьох показників якості інформаційних ресурсів по відношенню до включення у відповідь і релевантних рубриці документів:

- a – кількість документів, релевантних рубриці у пошуковій видачі;
- b – кількість документів, що не є релевантними рубриці у пошуковій видачі;
- c – кількість не виданих у пошукову видачу, але релевантних рубриці документів;
- d – кількість не виданих у пошукову видачу та нерелевантних рубриці

документів.

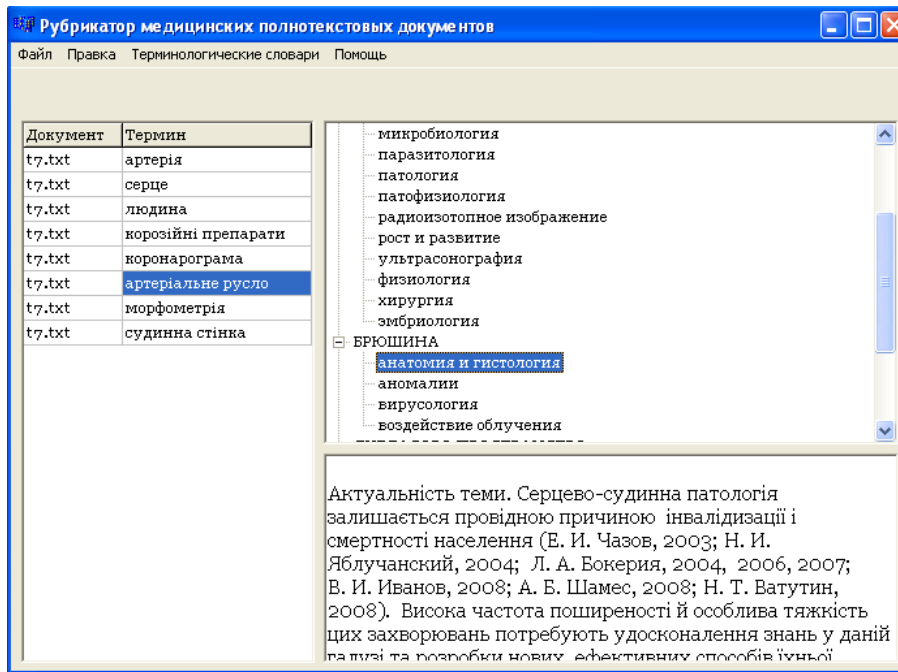


Рисунок 7. – Інтерфейс програми рубрикації повнотекстових документів

Проаналізовано близько 400 документів за 30 рубриками. Використано зовнішні міри для порівняння автоматичної розбивки з одержаною від експертів «еталонною» розбивкою цих же даних для розробленого програмного забезпечення. Для проаналізованих рубрик отримані наступні усереднені значення показників: міра повноти Recall = 0,873, міра точності Precision = 0,913, помилка автоматичної рубрикації Error = 0,014.

Проведені експерименти показали ефективність запропонованого в роботі методу компараторної ідентифікації для рубрикації повнотекстових документів. Програмна реалізація методу може бути використана для електронних бібліотек як елемент пошукових систем. У роботі також показані перспективи використання запропонованих моделей і методів ідентифікації знань: алгебри скінчених предикатів для автоматизації процедури переходу від формули хімічного сполучення до його текстового еквівалента, ідентифікації морфологічних характеристик нового слова при складанні електронних словників; алгебро-логічних моделей емоційно-насиченої лексики для оцінки емоційної насиченості текстів засобів масової інформації по відношенню до певного об'єкту.

ВИСНОВКИ

У дисертаційній роботі вирішена науково-практична задача підвищення ефективності роботи сучасної бібліотеки за рахунок використання інтелектуальних методів і моделей обробки інформаційних об'єктів.

У процесі виконання дисертаційної роботи отримані наступні результати.

1. Проаналізовані сучасні інформаційні бібліотечні системи та сформульовані основні вимоги до їх математичного та лінгвістичного забезпечення. Виявлено особливості тексту як об'єкта моделювання, обробки та представлен-

ня знань, які містяться у інформаційних об'єктах.

2. Досліджено інструментарій теорії інтелекту, проведено аналіз властивостей та можливостей алгебри предикатів та предикатних операцій для створення моделей інформаційних об'єктів. Розроблено математичні та лінгвістичні засоби для розв'язання задач обробки текстових документів на основі моделювання лінгвістичної діяльності людини та інтелектуального аналізу даних методом компараторної ідентифікації, що дозволило застосувати методи теорії інтелекту для складних задач семантичної обробки текстової інформації.

3. Запропоновано використання ситуаційно-текстового предикату при вирішенні задач аналізу і синтезу, стиску та узагальнення текстової інформації для анотування та реферування інформаційних об'єктів. Розглянуто декомпозицію ситуаційно-текстового предиката на функції сприйняття, розуміння та усвідомлення, що дозволило вирішувати задачі методом проведення лінгвістичного експерименту, суто фізичним спостереженням, без залучення суб'єктивних даних.

4. Показано застосування методу компараторної ідентифікації на документах повнотекстової бази даних. Представлено розв'язання задач предметного групування, систематизації та рубрикації повнотекстових документів у бібліотечних системах з використанням методу компараторної ідентифікації, що дозволило вести як вузько тематичний, так і багатоаспектний пошук документів у бібліотечній пошуковій системі.

5. Запропоновано модель бази знань автоматизованої інформаційно-бібліотечної системи на основі побудови семантичної мережі понять, які записані на мові алгебри скінчених предикатів. Такий підхід дозволяє розглядати не сталі зв'язки мовних одиниць, а сталі зв'язки між поняттями, що надає можливість створювати модель предметної області, яка не залежить від вхідної та вихідної мов.

6. Виконано практичну реалізацію запропонованих методів і математичних моделей у вигляді прикладної інформаційної технології для обробки інформаційних об'єктів у автоматизованих інформаційних бібліотечних системах. Проведено оцінку ефективності запропонованих семантичних моделей та методів обробки текстової інформації для реальних автоматизованих інформаційно-бібліотечних систем. Виділено і проаналізовано кількісні та якісні показники підрахунку ефекту від упровадження моделей, методів та алгоритмів.

7. Результати роботи впроваджено при розробці інформаційного, математичного, алгоритмічного і програмного забезпечення у наукову бібліотеку Харківського національного медичного університету, наукову бібліотеку Харківського національного університету радіоелектроніки у вигляді інформаційно-логічних моделей, які були використані при розробці автоматизованої інформаційної бібліотечної системи, при вирішенні задачі оптимізації книгозабезпечення, а також для тематичної систематизації та рубрикації повнотекстових документів у науковій бібліотеці Національного технічного університету «Харківський політехнічний інститут» та використані у навчальному процесі на кафедрі інтелектуальних комп'ютерних систем НТУ «ХПІ».

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Алисейко З. А. (Кочуева З. А.) Программная поддержка автоматического реферирования текста / З. А. Алисейко, А. Б. Кованев, Н. В. Шаронова // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2005. – № 1(21). – С. 331-334.

Здобувач запропонував алгоритм автоматичного реферування повнотекстових документів на основі пошуку ключових одиниць.

2. Алисейко З. А. (Кочуева З. А.) Использование алгебры предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний / З. А. Алисейко, В. И. Булкин, О. В. Канищева, Н. В. Шаронова // Біоніка інтелекту. – Харків : ХНУРЕ, 2006. – № 1(64). – С. 59-63.

Здобувачем розроблено математичну модель представлення декларативних знань з використанням алгебри предикатів та предикатних операцій.

3. Алисейко З. А. (Кочуева З. А.) Исследование проблем ранжирования и релевантности полнотекстовых документов в информационном поиске / З. А. Алисейко, Н. В. Шаронова // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2006. – № 1(24). – С. 232-236.

Здобувач здійснив дослідження проблем ранжування та оцінки релевантності запитів та досліджуваних документів.

4. Алисейко З. А. (Кочуева З. А.) Автоматизированное индексирование полнотекстовых документов ключевыми словами / З. А. Алисейко, О. В. Канищева // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2007. – № 4(27). – С. 269-272.

Здобувачем розроблено алгоритм для автоматизованого індексування повнотекстових документів ключовими словами.

5. Алисейко З. А. (Кочуева З. А.) Моделирование систематизации и предметизации полнотекстовых документов / З. А. Алисейко, О. В. Канищева, Н. В. Шаронова // Проблеми інформаційних технологій. – Херсон : ХНТУ, 2007. – № 1. – С. 140-144.

Здобувач запропонував використання методу компараторної ідентифікації в процедурах систематизації та предметизації повнотекстових документів.

6. Кочуева З. А. Идентификация информационных объектов в современной библиотеке с использованием алгоритма реферирования / О. В. Канищева, З. А. Кочуева, Н. В. Шаронова // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2008. – № 1(30). – С. 126-130.

Здобувачем запропоновано підхід до ідентифікації інформаційних об'єктів у сучасній бібліотеці, на основі використання алгоритму реферування.

7. Кочуева З. А. Технологии Text Mining: обзор методов и задач обработки смысловой информации / Е. А. Орбинская, З. А. Кочуева // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2010. – № 2(38). – С. 348-353.

Здобувачем виконано огляд сучасного стану методів семантичної обробки інформаційних об'єктів.

8. Кочуева З. А. Моделирование процедур систематизации и классифика-

ции информационных объектов методом компараторной идентификации / Н. В. Борисова, З. А. Кочуева, Н. В. Шаронова, Н. Ф. Хайрова // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2012. – № 1(44). – С. 91-95.

Здобувач запропонував алгоритми процедур систематизації і рубрикації повнотекстових інформаційних об'єктів АІБС.

9. Кочуева З. А. Моделирование семантических отношений в онтологии / З. А. Кочуева, Н. В. Бабкова, О. В. Канищева // Біоніка інтелекту: науково-технічний журнал, 2013. – № 1(80). – С. 25-28

Здобувачем запропоновано використання алгебри скінченних предикатів у якості апарата моделювання семантичних відношень інформаційних об'єктів.

10. Кочуева З. А. Метод экстракции парадигматических отношений между понятиями толкового словаря / З. А. Кочуева, Н. Ф. Хайрова, С. В. Петрасова // Вісник Національного технічного університету «Харківський політехнічний інститут». – Харків : НТУ «ХПІ», 2013. – № 62 (1035). – С. 118-124.

Здобувач запропонував використання шаблонів лексичних послідовностей для ідентифікації семантичних відношень між концептами.

11. Кочуева З. А. Индексирование полнотекстовых документов для задачи интеллектуального поиска информации по ключевым словам / З. А. Кочуева, Н. В. Борисова // Східно-Європейський журнал передових технологій. – Харків : ПП «Технологічний Центр», 2014. – № 1/2 (67). – С. 4-8.

Здобувачем розроблена модель системи інтелектуального пошуку інформації за ключовими словами.

12. Алисейко З. А. (Кочуева З. А.) Информатизация образования и проблемы автоматизации библиотечных систем / З. А. Алисейко, Н. В. Шаронова // Теорія і практика управління соціальними системами. – Харків : НТУ «ХПІ», 2005. – № 2 – С. 68-75.

Здобувачем проведено аналіз сучасного стану інформатизації освіти та розглянуті питання автоматизації бібліотек.

13. Алисейко З. А. (Кочуева З. А.) Автоматическое реферирование библиотечной информации для систем дистанционного обучения / А. Б. Кованев, Н. В. Шаронова, З. А. Алисейко // Образование и виртуальность: материалы 9-й Международной конференции Украинской ассоциации дистанционного образования «Образование и виртуальность-2005». – Харьков – Ялта : УАДО, 2005. – С. 220-226.

Здобувач запропонував алгоритм реферування повнотекстових документів у системах дистанційного навчання.

14. Кочуева З. А. Моделирование семантической структуры наименований химических соединений для обработки текстов медицинской и химической тематики / З. А. Кочуева // Радіоелектроніка і молодь в ХХІ ст. : тези доповідей 12-ого міжнародного молодіжного форуму. – Харків : ХНУРЕ, 2008. – С. 290.

15. Кочуева З. А. Опыт создания дистанционного курса по дисциплине «Основы программирования и алгоритмические языки» / О. В. Канищева, З. А. Кочуева, Н. В. Борисова // Экспертные оценки элементов учебного процесса: программа и материалы X межвузовской научно-практической конфере-

нции. – Харьков : ХГУ «НУА» – 2008. – С. 36-37.

Здобувачем запропоновано структуру інформаційних навчальних об'єктів у дистанційному курсі.

16. Кочуева З. А. Моделирование логико-семантических закономерностей перехода от химических формул к текстовой информации при создании отраслевых библиотечных ресурсов / З. А. Кочуева // Автоматика 2010 : материалы международной конференции. – Харьков : ХНУРЭ, 2010. – С. 114-115.

17. Кочуєва З. А. Методи і моделі обробки інформаційних об'єктів у сучасних бібліотечних системах / З. А. Кочуєва // Інтелектуальні системи і прикладна лінгвістика: Матеріали I Всеукраїнської науково-практичної конференції. – Харків: НТУ «ХП», 2012. – С. 20-21.

18. Кочуєва З.А. Моделювання інформаційних об'єктів у бібліотечних системах / Н.В. Борисова, З. А. Кочуєва // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: Тези доповідей XX Міжнародної науково-практичної конференції, Ч.1. – Харків: НТУ «ХП». – С. 5.

Здобувач запропонував використання алгебри скінчених предикатів у якості апарата моделювання інформаційних об'єктів.

АНОТАЦІЇ

Кочуєва З. А. Методи і моделі інтелектуальної обробки інформаційних об'єктів у сучасних бібліотечних системах. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Національний технічний університет «Харківський політехнічний інститут», Харків, 2014.

У дисертаційній роботі вирішена актуальна науково-практична задача підвищення ефективності роботи сучасної бібліотеки за рахунок використання інтелектуальних методів і моделей обробки інформаційних об'єктів. Проаналізовано наукові досягнення в області автоматизації бібліотечних процесів і сформульовано основні вимоги до розробки математичного та лінгвістичного забезпечення бібліотечних систем. Розроблено математичні та лінгвістичні засоби для розв'язання задач обробки текстових документів на основі моделювання лінгвістичної діяльності людини та інтелектуального аналізу даних методом компараторної ідентифікації. Представлено розв'язання задач предметного групування, систематизації та рубрикації повнотекстових документів у бібліотечних системах з використанням методу компараторної ідентифікації, що дозволяє вести як вузько тематичний, так і багатоаспектний пошук документів у бібліотечній пошуковій системі. Запропоновано модель бази знань автоматизованої інформаційно-бібліотечної системи (АІБС) на основі побудови семантичної мережі понять, які записані на мові алгебри скінчених предикатів. Виконано практичну реалізацію запропонованих методів і математичних моделей у вигляді прикладної інформаційної технології для обробки інформаційних об'єктів у АІБС. Проведено оцінку ефективності запропонованих семантичних моделей та методів обробки текстової інформації для реальних АІБС. Виділено і проаналізовано кількісні та якісні показники підрахунку ефекту від упровадження

моделей, методів та алгоритмів.

Ключові слова: інформаційний об'єкт, автоматизована переробка інформації, інтелектуальна обробка даних, автоматизовані інформаційні бібліотечні системи, алгебра скінчених предикатів, компараторна ідентифікація.

Кочуева З. А. Методы и модели интеллектуальной обработки информационных объектов в современных библиотечных системах. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. – Национальный технический университет «Харьковский политехнический институт», Харьков, 2014.

Диссертация посвящена решению актуальной научно-практической задачи повышения эффективности работы современной библиотеки за счет использования интеллектуальных методов и моделей обработки информационных объектов. Проанализированы достижения в области автоматизации библиотечных процессов и сформулированы основные требования к разработке математического и лингвистического обеспечения библиотечных систем. Разработаны математические и лингвистические средства для решения задач обработки текстовых документов на основе моделирования лингвистической деятельности человека и интеллектуального анализа данных методом компараторной идентификации, что позволяет применить методы теории интеллекта в сложных задачах семантической обработки текстовой информации.

Представлено решение задач предметного группирования, систематизации и рубрикации полнотекстовых документов в библиотечных системах с использованием метода компараторной идентификации, что позволяет вести как узкотематический, так и многоаспектный поиск документов в библиотечной поисковой системе. Предложена модель базы знаний автоматизированной информационной библиотечной системы (АИБС) на основе построения семантической сети понятий, записанных на языке алгебры конечных предикатов. Такой подход позволяет рассматривать устойчивые связи между понятиями, что дает возможность создавать модель предметной области, не зависящую от входного и выходного языков.

Выполнена практическая реализация предложенных методов и математических моделей в виде прикладной информационной технологии для обработки информационных объектов в АИБС. Проведена оценка эффективности предложенных семантических моделей и методов обработки текстовой информации в реальных АИБС. Выделены и проанализированы количественные и качественные показатели подсчета эффекта от внедрения моделей, методов и алгоритмов.

Результаты диссертационной работы внедрены в научной библиотеке Харьковского национального медицинского университета, научной библиотеке Харьковского национального университета радиоэлектроники, научной библиотеке Национального технического университета «Харьковский политехнический институт», а также в учебном процессе на кафедре интеллектуальных компьютерных систем НТУ «ХПИ» при преподавании специальных дисциплин «Информационно-ресурсное обеспечение лингвистической деятельности»,

«Автоматизированная обработка естественного языка», «Математические основы лингвистики», «Искусственный интеллект: лингвистические проблемы».

Ключевые слова: информационный объект, автоматизированная переработка информации, интеллектуальная обработка данных, автоматизированные информационные библиотечные системы, алгебра конечных предикатов, компараторная идентификация.

Kochueva Z. A. Methods and Models of Intellectual Processing of Information Objects in the Modern Library Systems. – On the rights of the Manuscript.

Thesis for a Candidate Degree in Technical Sciences, speciality 05.13.06 – Information technologies. – National Technical University «Kharkiv Polytechnic Institute», Kharkiv, 2014.

The dissertation is devoted to solving a relevant scientific and practical problem of increasing the efficiency of modern libraries using intelligent methods and models of information processing facilities. The paper analyzes the scientific advances in the automation of library processes and formulates the basic requirements for developing mathematical and linguistic support of library systems. The paper introduces the mathematical and linguistic tools for solving problems of processing of text documents based on modeling of human linguistic activity and data mining by means of the comparator identification method. The problems of subject grouping, ordering and categorization of full-text documents in the library systems using the method of comparator identification have been solved, which allows to perform both narrow thematic and multiaspect search for documents in the library search engine. There has been suggested a model of automated information library system (ALIS) knowledge base on the ground of semantic networks of concepts that are written in the language of algebra of finite predicates. The proposed methods and mathematical models have been practically implemented in the form of applied information technology to process information objects in ALIS. An evaluation of the effectiveness of the proposed semantic models and methods of processing text information in real ALIS has been made. Quantitative and qualitative indicators calculating the effect of the introduction of models, methods and algorithms have been determined and analyzed.

Key words: automated information processing, intellectual data processing, automated information library systems, algebra of predicates, algebra of predicate operations, comparator identification.