

УДК 004.75.002.71:004.42:021.7



ТОВМАЧЕНКО
Василь Миколайович,
пров. наук. співроб.,
Державна наукова сільськогосподарська
бібліотека УАН
(м. Київ)

СИСТЕМИ ОБРОБКИ ІНФОРМАЦІЇ ДЛЯ ІНТЕРНЕТ-ТЕХНОЛОГІЙ БІБЛІОТЕЧНИХ МЕРЕЖ : ІСТОРИКО-НАУКОВИЙ АНАЛІЗ

Розглянуто досягнення кібернетики в таких напрямках як «інтелектуальні системи», «штучний інтелект», інформаційно-комунікаційні мережі, в т.ч. глобальна мережа, системи розпізнавання образів, експертні системи, комп'ютерна лінгвістика, тобто такі, що визначають нові технології обробки інформації для потреб інформаційного суспільства.

Рассмотрены достижения кибернетики в таких направлениях как «интеллектуальные системы», «искусственный интеллект», информационно-коммуникационные сети, в т.ч. глобальная сеть, системы распознавания образов, экспертные системы, компьютерная лингвистика, то есть такие, которые определяют новые технологии обработки информации для потребностей информационного общества.

Achievement of cybernetics is considered in such directions as «intellectual systems», «artificial intelligence», of informatively communication networks, including global, systems of recognition of patterns, consulting models, computer linguistics, that such which determine new technologies of treatment of information for the necessities of informative society.

Процес інформатизації нашого суспільства потребує переведення бібліотек на нові бібліотечні технології та формування принципово нових за змістом електронних ресурсів. Упровадження інтернет-мережі та цифрових технологій відкрило нову еру інформаційного суспільства, створило необхідні умови для переходу нашого суспільства до суспільства знань [1].

Розподілена глобальна комп'ютерна мережа Інтернет містить колосальний масив інформації, накопичений світовою спільнотою. Спостерігається лавиноподібне зростання інформації (обсяг створюваної інформації подвоюється кожні 8–10 років), що зумовлює необхідність

автоматизації технологій її обробки (пошук, стиснення, передача, аналіз). Розробка нових прогресивних методів обробки інформації є прерогативою кібернетики. Наразі слушно говорити про інформаційно-комунікаційні технології, комп'ютеризацію інформаційної сфери, що зокрема стимулювало розвиток нового напрямку прикладної лінгвістики – комп'ютерної, до компетенції якої входить лінгвістичне забезпечення функціонування інформації.

Прогресивні інформаційні технології визначають розвиток суспільства і значною мірою зумовлюють його статус у сучасному світі.

Найважливішою ланкою науково-інформаційного забезпечення є бібліотечна сфера. Бібліотечний технологічний процес, що складається з таких підпроцесів, як комплектування, каталогізування, багатоаспектний пошук, видача документів, МБА, підтримка технологій роботи зі штрихкодами [2], наразі здебільшого забезпечує доступ до електронного каталога в *мережі Інтернет*. Великі бібліотечні центри мають також розвинуті підрозділи реферування текстової інформації.

Суперечність між невпинно зростаючим обсягом знань і обмеженими можливостями існуючої системи документних комунікацій, що базується на паперових носіях інформації, обумовлює необхідність кардинальної активізації робіт по введенню в цю систему електронних бібліотечно-інформаційних ресурсів.

Створення бібліографічної продукції і її доведення до споживачів засобами сучасних комп'ютерних технологій значно зменшили гостроту суперечності між обсягом документних ресурсів і можливістю їх використання суспільством в системі документних комунікацій [3].

Трансформація функцій книгозбірень відзначається і в професійній термінології. З'явилися нові поняття «віртуальна бібліотека», «цифрова бібліотека» та «електронна бібліотека», що були запозичені з інформатики. Перший з них – «віртуальна бібліотека» (Virtual library) — використовується для визначення загалу інформаційних джерел, які є доступними через глобальні

комп'ютерні мережі, що в сукупності утворюють Internet [2, 3]. Ресурси "віртуальної бібліотеки" розподілені по всьому світі. Можна вважати, що "віртуальна" бібліотека — це інтегровані комп'ютерними мережами "цифрові" та "електронні" бібліотеки.

Наразі розроблено значну кількість програмних систем для автоматизації бібліотек [2]: система **"МАРК"** (в/о «Информсистема», м. Москва, Росія), в т.ч. мережевий варіант даної системи, відома система ІРБІС, що набула широкого застосування, система *ALEP* (Іерусалим), система LIBER MEDIA (Франція) та ін.

Подальші досягнення в бібліотечній справі цілком залежать від стану розробок в таких наукових напрямках кібернетики, як *«інтелектуальні системи»*, *«штучний інтелект»*, *методи обробки (пошук, збереження, стиснення, захист) інформації* в інформаційно-комунікаційних мережах, *системи розпізнавання образів, експертні системи, математична (комп'ютерна) лінгвістика* і т. ін. Метою даного повідомлення є фрагментарне висвітлення сучасного стану досліджень в наукових напрямках кібернетики, що визначають технології обробки інформації.

Інтелектуальні системи.

Напрямок *інтелектуальні системи*, утворений на перетині математики, логіки, електроніки, фізики, психології та нейрофізіології має своїми задачами вивчення, моделювання та штучне відтворення інтелектуальних властивостей людського мозку з метою вирішення суспільнокорисних проблем [4]. Він включає дві наукові дисципліни – *штучний інтелект* та *штучні нейронні мережі*, що втілюють протилежні підходи до моделювання інтелектуальних властивостей мозку. Предметом вивчення для *штучного інтелекту* є розумові властивості, моделювання та практичне використання притаманних людському інтелекту методів рішення задач, сприйняття та аналізу інформації. В цьому сенсі штучний інтелект наближений до психології, відрізняючись від останньої метою та методами досліджень. Для наукового напрямку *штучні нейронні мережі* предметом вивчення є будова та функції нервової системи, методи її

моделювання. Метою досліджень є створення штучних нейронних мереж, здатних до навчання, асоціативного пошуку, інтерпретації сенсорної (зорової, слухової, смакової, тактильної) інформації. За предметом вивчення цей напрямок близький до нейрофізіології, результати якої становлять базу для моделювання штучних нейронних мереж.

Експертні системи.

Серед прикладних розробок штучного інтелекту мабуть найважливішими є *експертні системи* [4], тобто інформаційні системи, в яких знання експертів у певних предметних областях доповнюються формальними та евристичними методами пошуку рішень. В період 1970–80 рр. було створено кілька десятків потужних експертних систем, які і досі широко застосовуються у промисловості, медицині та в наукових дослідженнях. Найбільш відомою є створена в Стенфордському університеті в середині 70-х років система MYCIN, в якій знання експертів з медицини використовуються для діагностики та лікування менінгіту та заражень крові.

Поширення експертних систем сприяло появі та розвитку інженерії знань – напрямку штучного інтелекту, задачами якого є добуток знань (Data Mining), розробка методів їх представлення, створення баз знань та їх практичне використання.

Добуток та представлення знань пов'язані з *проблемою розуміння природної мови*. Розуміння мови є фундаментальною властивістю людського розуму, тому вирішення цієї проблеми з самого початку розглядалось як одне з ключових завдань штучного інтелекту.

Протягом останніх двох десятиріч з'явився новий напрямок штучного інтелекту – *інтелектуальні агенти* [4, 15]. Інтелектуальний агент – це програмний робот, здатний аналізувати інформаційне оточення та спілкуватися з іншими агентами для координації спільних дій, наприклад, редагування файлів або пошуку даних на серверах комп'ютерної мережі. Поява цього напрямку зумовлена поширенням технології розподілених систем, зокрема *ІНТЕРНЕТ*, утворенням інформаційного середовища для таких агентів.

Нейронні мережі.

Науковий напрям «штучні нейронні мережі», виник у 50-х роках минулого сторіччя, практично одночасно з напрямом штучного інтелекту. Предметом досліджень для нього є будова та функції нервової системи, а метою – створення штучних нейронних мереж, здатних відтворювати поведінку елементів нервової системи мозку. Виникнення цього напрямку пов'язують зі здатністю мережі з формальних нейронів виконувати логічні обчислення.

Першу модель штучної нейронної мережі- перцептрон «Марк-1», було створено в 1957 р. Ф. Розенблаттом [5, 7]. На ній було вперше продемонстровано здатність нейронної мережі навчатись розпізнавати прості зображення (вертикальні, або горизонтальні лінії). Це досягнення мало значний резонанс в науковому суспільстві і невдовзі в різних лабораторіях світу з'явилися аналоги та модифікації перцептрона. В Україні перший аналог перцептрона було створено в 1961р. О.Г. Івахненко (система Альфа) [5]. В 1961-62 рр. академік В.М. Глушков запропонував оригінальну теоретичну модель перцептрона Розенблатта, та дослідив параметри режимів його навчання [7]. В 1970 р. в Україні було створено перший в Європі цифровий перцептрон АДАМ [5]. Серед українських вчених, що протягом 60-х років внесли внесок в розвиток штучних нейронних мереж можна згадати також Ю.П. Зайченко, М.М. Шлезінгера, О.М. Різника, А.О. Бондара, А.В. Анісімова та багатьох інших.

Перспективи та проблеми інтелектуальних технологій.

Перспективи галузі інтелектуальних систем пов'язані з проблемами інтелектуалізації засобів ІНТЕРНЕТ [2]. Ця система являє собою майже ідеальне середовище для впровадження новітніх інформаційних технологій. Розробляються численні проекти використання методів штучного інтелекту та штучних нейронних мереж для поліпшення методів зберігання, пошуку та захисту інформації в розподілених базах даних системи, інтелектуалізації інтерфейсу користувача, створення принципово нових інформаційних сервісів. Для штучного інтелекту перспективними напрямками є дослідження

колективної поведінки інтелектуальних роботів, створення гнучких автономних систем ситуативного управління, здатних пристосовуватись до змін оточення.

Природний людський інтелект завжди побудований на принципі єдності функціонування образного і символічного; а сучасний рівень штучного інтелекту оперує лише (і то в обмеженому вигляді) символічними структурами. Тому із загальнофілософського погляду, тенденція подальшого розвитку штучних інтелектуальних систем (сучасної комп'ютерної техніки зокрема) та їх перспектива лежать в напрямі досягнення аналогу (ідентичності) – із своїм першоджерелом та першоосною – з природним людським інтелектом. А це можливо, якщо сучасна комп'ютерна техніка розвиватиметься в напрямі, передбаченому ще десять років назад Б. Гейтсом. В одній із своїх доповідей знаменитий засновник і керівник фірми «Microsoft» зазначив, що майбутнє – за інтелектуальними інформаційними технологіями та системами, які сприймають і розуміють людську мову, друковані та рукописні тексти, обробляють і розуміють різноманітні зображення, креслення тощо, сприймають просторові та звукові сцени. Саме розробивши ці технології, Б.Гейтс збирається заволодіти світовим інформаційним простором та отримати надприбутки [10].

Образний комп'ютер [8, 9].

Про сучасний стан вітчизняних досліджень в галузі нових інформаційно-комунікаційних технологій та штучного інтелекту дає уявлення Державна науково-технічна програма «Образний комп'ютер», затверджена Постановою Кабінету Міністрів ще в 2000 р. терміном на 10 років [9]. Хоча з виконанням програми не все добре, вона все ж свідчить про високий світовий рівень кібернетичних досліджень вітчизняних вчених.

В обґрунтуванні даної програми її творці зазначають, що «сьогодні науково-технічний прогрес визначається значними проривами екстенсивного характеру в галузі розроблення та впровадження інформаційних систем, засобів телекомунікації та комп'ютеризації.

Досягнені фантастичні швидкодії та обсяги пам'яті у комп'ютерів, глобалізується Інтернет, надається широкий спектр інформаційних послуг обчислювального та пошукового характеру.

Спостерігаються й чіткі негативні тенденції, пов'язані зі світовою монополізацією розроблення та виробництва комп'ютерів та засобів комунікації, зі штучним стримуванням розвитку передових інформаційних технологій та систем, в яких має нагальну потребу людство.

А людство потребує створення принципово нових комп'ютерів, які здатні сприймати та розуміти звуки, зображення, людську мову, рукописні тексти, креслення, просторові та звукові сцени, інші образи, описувати та озвучувати зображення, перекладати з однієї мови на іншу тощо.

Розроблення таких комп'ютерів, які виконують не тільки обчислення, але й моделюють образне сприйняття світу та образне прийняття рішень відносять до проривних напрямів у науково-технологічному поступі...

Образний комп'ютер — це така кібернетична система, в якій виконується функційне моделювання інтелектуальної, головно підсвідомої, діяльності людини та всього живого, що пов'язана зі сприйняттям зорових, слухових та інших образів, аналізом сцен та складних ситуацій, плануванням дій та рухів, узагальненням спостережень, встановленням закономірностей, прогнозуванням, прийняттям рішень тощо...»

Ось деякі властивості образного комп'ютера (ОК):

«...Сприймає людську мову, автоматично розпізнає та розуміє її, озвучує довільні тексти. Автоматично редагує та друкує тексти під диктування.

Сприймає та розуміє надруковані та написані від руки тексти. Вміє перекладати з однієї мови на іншу. Робить усний переклад з однієї мови на іншу. Аналізує та розуміє зображення (фото, написи, креслення, графіки тощо). Аналізує та розуміє сцени, обстежує об'єкти, вимірює відстань до об'єктів, вказує їх місцезонашування. Орієнтується в часі та просторі. Аналізує причини та наслідки. Верифікує та ідентифікує людину за її голосом. Сприймає запахи та тактильну інформацію...».

Зауважимо, що для бібліотечної справи дуже цінним є вміння ОК робити «...автоматичне реферування та пошук текстової та аудіо-відеоінформації за заданою темою. ОК «Авто-референт»».

Системи та технології обробки текстової інформації. Сучасний стан та тенденції розвитку.

Дослідження та розробки в області автоматичної обробки тексту у Європі та США притягують увагу великих приватних фірм та державних організацій самого високого рівня. Європейський союз вже декілька років координує різноманітні програми в області автоматичної обробки текстів. Наприклад, Human Language Technology Sector of the Information Society Technologies (IST) Programme 1998–2000 [12–16].

Один з найбільш цікавих проектів у рамках даної програми – SPARKLE (Shallow PARsing and Knowledge Extraction for Language Engeneering). Серед його учасників – Dimler-Benz, Xerox Research Centre in Europe та Cambridge University Computer Laboratory. Мета проекту – створення часткових синтаксичних аналізаторів для основних мов Європейського союзу. В США з 1991 до осені 1998 року існував проект TIPSTER, що був організований DARPA, Департаментом Оборони та ЦРУ сумісно з Національним Інститутом Стандартів та Технологій і Центром військово-повітряних ті військово-морських озброєнь (SPAWAR). У роботі консультативної ради програми брали участь також ФБР, Національний Науковий Фонд та деякі інші організації. Основною метою програми було порівняння та оцінка результатів роботи різноманітних пошукових систем та систем реферування.

Необхідно зауважити, що такі задачі як розпізнавання та генерація мови, створення пошукових систем аж до тепер розв'язуються із мінімальною участю лінгвістів. Це зумовлено використанням при розв'язанні вищезгаданих задач в основному статистичних методів.

Не дивлячись на це, за багато років чітко визначились області, в яких позиції професіональних лінгвістів є найбільш сильними. Це лексико-граматичний аналіз речення, синтаксичний аналіз речення, знаходження

іменників та власних імен у тексті та автоматичне реферування. Підходи до лексико-граматичного аналізу речення та методи реалізації синтаксичного аналізу, а також розгляд проблеми автоматичного реферування тексту, як одного з найпоширеніших видів автоматичної обробки текстів, розглянуто в роботах [12–14].

Комп'ютерна лінгвістика.

Як відомо в Національній бібліотеці України імені В.І. Вернадського розроблено засади побудови й організації функціонування Національної системи реферування з урахуванням світового досвіду й особливостей організації вітчизняних документальних комунікацій. [11]. Важливим елементом даної системи є автоматизація процесу аналітико-синтетичної обробки документальної інформації, тобто автоматичне реферування. Теоретичною базою побудови автоматичних систем реферування є комп'ютерна лінгвістика.

Одним з класичних завдань комп'ютерної лінгвістики є побудова систем автоматичного реферування наукових, технічних, політичних, ділових текстів. Актуальність цього завдання значно зростає у зв'язку з появою на інформаційному ринку повнотекстових баз даних. Ці бази містять у вигляді документів не реферати, а тексти наукових статей.

Лінгвістична обробка природномовних текстів є однією з центральних проблем інтелектуалізації інформаційних технологій. Цій проблемі приділяється значна увага в розвинутих країнах Європи та США, свідченням чого є виділення величезних коштів на розробку лінгвістичного програмного забезпечення. Велика кількість науково-дослідних програм спрямовані на розвиток лінгвістичних інформаційних систем. У зв'язку з бурхливим розвитком Інтернет, інших комп'ютерно-комунікаційних технологій ця проблема набуває ще більшої значимості.

Ще з середини 50-х років минулого століття значні зусилля науковців були спрямовані на розробку математичних алгоритмів та комп'ютерних програм обробки текстів природною мовою. Для автоматизації аналізу та

синтезу текстів створювалися різноманітні моделі процесів обробки тексту, а також відповідні алгоритми та структури представлення даних. Традиційно аналіз природномовних текстів представлявся як послідовність процесів – морфологічний аналіз, синтаксичний аналіз, семантичний аналіз. Для кожного з цих етапів були створені відповідні моделі та алгоритми.

Зауважимо, що значний вклад в розвиток комп'ютерної лінгвістики внесли київські вчені: докт. фіз.-мат. наук, професор А.В. Анісімов та його учні: М.М. Глибовець, О.О. Марченко та інші. Безперечна заслуга наукової школи проф. А.В. Анісімова – застосування методів паралельних обчислень для вирішення задач комп'ютерної лінгвістики [12–14]. Початок вивченню властивостей паралельних обчислень для різноманітних задач поклали роботи В.М. Глушкова, О.А. Летичевського, А.В. Анісімова та інших вчених. Саме результати математично-алгоритмічних досліджень паралельних розподілених процесів, отримані в 1980–90-ті роки професором А.В. Анісімовим, і знайшли своє ефективне застосування в задачах обробки природномовних текстів. Керуючі простори, що були запропоновані А.В. Анісімовим як універсальні засоби моделювання паралельних асинхронних розподілених процесів, були використані для привнесення в стандартні схеми обробки текстів елементів взаємодії між процесами при аналізі та синтезі природномовних текстів.

Поєднання послідовних та паралельних, формальних та евристичних підходів до вирішення описаних проблем дає можливість досягти значного підвищення ефективності програмної реалізації моделей обробки природномовних текстів.

Вивченню властивостей паралельних обчислень для різноманітних задач були присвячені роботи В.М. Глушкова, О.А. Летичевського, А.В. Анісімова та інших вчених. Саме результати математично-алгоритмічних досліджень паралельних розподілених процесів, отримані в 1980-90-і роки професором А.В. Анісімовим, і знайшли своє ефективне застосування в задачах обробки природномовних текстів. Керуючі простори, що були запропоновані Анісімовим як універсальні засоби моделювання паралельних асинхронних

розподілених процесів, були використані для привнесення в стандартні схеми обробки текстів елементів взаємодії між процесами при аналізі та синтезі природномовних текстів.

Розроблені науковцями школи проф. А.В. Анісімова алгоритми онтологічного семантичного аналізу текстів природною мовою на основі обробки семантичного контексту та побудови асоціативно-семантичних шляхів в мережах онтологічних лінгвістичних систем є базою для створення математично-програмних моделей багатьох процесів смислової обробки текстів, таких як семантичне реферування текстів, індексування, смисловий переклад текстів з однієї мови на іншу, підтримка семантично зв'язного природномовного діалогу та багатьох інших.

Висновки.

Наразі в сфері науково-технічної інформації до використання на досить високому рівні адаптовані сучасні інформаційно-комунікаційні технології. Подальше впровадження в даній сфері досягнень в галузі передових новітніх технологій слід пов'язувати з розробками в таких наукових напрямках, як інтелектуальні системи (штучний інтелект, нейронні мережі), дослідженнями по програмі «Образний комп'ютер», новими розробками з комп'ютерної лінгвістики.

Список використаної література

1. *Онищенко О. С.* Українські бібліотеки у 90-х роках ХХ ст. : основні тенденції розвитку та напрями діяльності / О. С. Онищенко // Бібл. вісн. – 2009. – № 3. – С. 12–23.
2. *Сучасні інформаційні та новітні технології в бібліотечній справі* [Електронний ресурс]. – Режим доступу: <http://ukrref.com.ua/?id=MjE3Mw%3D%3D> – Заголовок з екрана.
3. *Павлуша І. А.* Електронні бібліотеки: системний підхід до формування фондів / І. А. Павлуша // Бібл. вісн. – 2000. – № 2. – С. 16–19.
4. *Інтелектуальні системи* [Електронний ресурс]. – Режим доступу: <http://is22.kpi.in.ua/Reznik/%CB%E5%EA%F6iAI> – Заголовок з екрана.
5. *Ивахненко А. Г.* Системы самообучающиеся с положительными обратными связями / А. Г. Ивахненко // Киев : Автоматика. – 1962. – № 3.
6. *Глушков В. М.* К вопросу о самообучении в перцептроне / В. М. Глушков // Журн. вычисл. математики и матем. физики. – 1962. – № 6. – С. 1102–1110.

7. *Минский М.* Перцептроны /М. Минский, С. Пейперт. – М. : Мир, 1971. Т. 2. – С. 759–766.
8. *Винцюк Т. К.* Анализ, распознавание и интерпретация речевых сигналов / Т. К. Винцюк. – К. : Наук. думка, 1987, – 264 с.
9. Державна науково-технічна програма «Образний комп'ютер» [Електронний ресурс]. – Режим доступу: <http://www.nas.gov.ua/ResearchActivities/NANU/Pages/01.aspx> – Заголовок з екрана.
10. *Яровий А. М.* Образне пізнання в контексті глобальних проблем сучасного інформаційного суспільства. Мультиверсум : філософський альманах / А. М. Яровий. – К. : Центр духовної культури, 2006. – № 58.
11. *Сорока М. Б.* Національна система реферування української наукової літератури / М. Б. Сорока ; НАН України, Нац. б-ка України імені В. І. Вернадського. — К. : НБУВ, 2002. – 209 с.
12. *Марченко О. О.* Система обробки тексту природною мовою / О. О. Марченко // Вісн. Київського ун-ту. Сер. фіз.-мат. науки. – 2001. – Вип. 5. – С. 125–130.
13. *Анісімов А. В.* Створення керуючого простору синтаксичних структур природної мови / А. В. Анісімов, О. О. Марченко, В. А. Нагорний // Вісн. Київського ун-ту. Сер. фіз.-мат. науки. – 2002. – Вип. 1. – С. 159–169.
14. *Анісімов А. В.* Система обработки текстов на естественном языке / А. В. Анісімов, О. О. Марченко // Искусственный интеллект / НАНУ и ИПШ. – К., 2002. – Вип. 4. – С. 157–164.
15. *Шевченко О. В.* Методи добування тематичної інформації з глобальної мережі на основі агентного підходу : дис... канд. фіз.-мат. наук : 01.05.03 / О. В. Шевченко ; Київський нац. ун-т ім. Тараса Шевченка. – К., 2007. – 123 арк. : рис. – Бібліогр. : арк. 115–122.