

УДК 004.9+811.93

[https://doi.org/10.52058/2786-6025-2023-3\(17\)-537-549](https://doi.org/10.52058/2786-6025-2023-3(17)-537-549)

Брусенцев Віталій Олександрович кандидат технічних наук, доцент, доцент кафедри інформаційних технологій, Харківська державна академія культури, Бурсацький узвіз, 4, м. Харків, тел.: (057) 731-32-82, <https://orcid.org/0000-0002-0020-4025>

Філіпова Людмила Яківна доктор педагогічних наук, професор, професор кафедри інформаційних технологій, Харківська державна академія культури, Бурсацький узвіз, 4, м. Харків, тел.: (057) 731-32-82, <https://orcid.org/0000-0003-0273-7922>

Шелестова Анна Миколаївна кандидат наук із соціальних комунікацій, доцент, доцент кафедри інформаційних технологій, Харківська державна академія культури, Бурсацький узвіз, 4, м. Харків, тел.: (057) 731-32-82, <https://orcid.org/0000-0003-4866-1767>

ЗАСОБИ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ В ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМАХ

Анотація. У статті розглянуто та проаналізовано головні засоби комп'ютерної лінгвістики, призначені для ефективної роботи інформаційно-пошукових систем (ІПС).

Досліджено роль комп'ютерної лінгвістики в розвитку ІПС та надано докладний опис компонентів, необхідних для функціонування ІПС. Наголошується на важливості розвитку інформаційно-пошукових мов (ІПМ), які існують на когнітивному рівні та використовуються для здійснення обробки та пошуку інформації в ІПС. В статті розглядаються декларативні та процедурні засоби комп'ютерної лінгвістики, які необхідні для аналізу та опрацювання текстової інформації в ІПС. Дослідження, які були проведені авторами, доводять, що питання теорії та практики інформаційного пошуку тісно пов'язані з комп'ютерною лінгвістикою. Це пояснюється тим, що основний обсяг інформації та тексти документів представлені природними мовами, а для отримання інформації необхідні знання автоматичної обробки мов, зокрема правила комп'ютерної морфології, семантики тощо.

Проаналізовано засоби накопичення та опрацювання знань, представлені у концептуальній схемі ІПС, яку доповнено аналізом алгоритмів пошуку інформації в сучасних ІПС мережі Інтернет, розробками моделі пошуку та відповідним понятійним апаратом. Докладно охарактеризовано

взаємозв'язок ІПС з системами штучного інтелекту, де знання представляються спеціальними моделями на штучних мовах лінгвістичного забезпечення цієї системи.

Відзначено, що всі інформаційні мови в автоматизованих ІПС використовуються через програмні засоби комп'ютерної лінгвістики, що надають очікуваний результат - релевантну інформацію у відповідь на запит користувача. Розглянуто та охарактеризовано штучні мови подання знань та семантики (МПЗ), серед яких – логічні, мережеві, фреймові, які відзначаються тенденціями до інтелектуалізації роботи ІПС та персоналізації пошуку.

Підкреслено, що традиційні лінгвістичні засоби, які використовуються в інформаційно-пошуковій мові (ІПМ), спрямовані на обробку текстової інформації, зокрема на кодування, аналіз, згортання та розпізнавання тексту. Проте, інтелектуальні ІПМ є перспективними інструментами в ІПС, особливо для пошукових систем багатьох електронних бібліотек та архівів. Вони використовують комп'ютерні програми для автоматичної обробки та синтезу різних видів інформації, таких як числова, невербальна, графічна, аудіальна, візуальна та відеоінформація.

Ключові слова: комп'ютерна лінгвістика, інформаційно-пошукова система, інформаційний пошук, інформаційно-пошукова мова, штучна мова.

Brusentsev Vitalii Oleksandrovyich PhD in Technical Sciences, Associate Professor, Associate Professor at the Information Technologies Department, Kharkiv State Academy of Culture, Bursatsky uzviz, 4, Kharkiv, tel.: (057) 731-32-82, <https://orcid.org/0000-0002-0020-4025>

Filipova Liudmyla Yakivna Doctor of Pedagogical Sciences, Professor, Professor at the Information Technologies Department, Kharkiv State Academy of Culture, Bursatsky uzviz, 4, Kharkiv, tel.: (057) 731-32-82, <https://orcid.org/0000-0003-0273-7922>

Shelestova Anna Mykolayivna PhD in Social Communication Studies, Associate Professor, Associate Professor at the Information Technologies Department, Kharkiv State Academy of Culture, Bursatsky uzviz, 4, Kharkiv, tel.: (057) 731-32-82, <https://orcid.org/0000-0003-4866-1767>

COMPUTATIONAL LINGUISTICS TOOLS IN INFORMATION RETRIEVAL SYSTEMS

Abstract. The article discusses and analyzes the main tools of computational linguistics designed for efficient operation of information retrieval systems (IRS). The role of computational linguistics in the development of IRS is explored, and a

detailed description of the components necessary for the functioning of IRS is provided. The importance of the development of information retrieval languages (IRL) is emphasized, which exist at a cognitive level and are used for processing and searching information in IRS. The article discusses declarative and procedural tools of computational linguistics necessary for analyzing and processing textual information in IRS. The studies conducted by the authors prove that the issues of the theory and practice of information retrieval are closely related to computational linguistics. This is explained by the fact that the main volume of information and document texts are represented in natural languages, and knowledge of automatic language processing, including the rules of computational morphology, semantics, etc., is necessary to obtain information.

The tools for accumulating and processing knowledge presented in the conceptual scheme of IRS are analyzed, which is supplemented by an analysis of information retrieval algorithms in modern Internet-based IRS, as well as the development of a search model and corresponding conceptual apparatus. The relationship between IRS and artificial intelligence systems is characterized in detail, where knowledge is represented by special models in artificial languages of linguistic support of this system.

It is noted that all information languages in automated IRS are used through computational linguistic software, which provides the expected result – relevant information in response to the user's request. Artificial languages of knowledge representation and semantics are discussed and characterized, among them are logical, network, and frame-based languages which have the trends towards the intellectualization of IRS work and personalization of search.

It is emphasized that traditional linguistic tools used in information retrieval language (IRL) are aimed at processing textual information, including encoding, analysis, compression, and recognition of text. However, intelligent IRLs are promising instruments in IR systems, especially for search engines of many electronic libraries and archives. They use computer programs for automatic processing and synthesis of various types of information, such as numeric, nonverbal, graphic, audio, visual, and video information.

Keywords: computational linguistics, information retrieval system, information search, information retrieval language, artificial language.

Постановка проблеми. Сучасні напрями комп'ютерної лінгвістики пов'язані, перш за все, з використанням комп'ютерів у мовознавстві та охоплюють багато процесів переробки інформації, серед яких: автоматичний аналіз текстів; автоматичний синтез текстів; створення автоматизованих інформаційно-пошукових систем; створення лінгвістичних баз даних та інше. Усі ці складники орієнтовані на інформацію, яка функціонує в межах усього інформаційного простору людства.

Серед інформаційних процесів важливим і майже головним вважається інформаційний пошук, який активно використовують в різних сферах діяльності. Для створення та функціонування автоматизованих інформаційно-пошукових систем (ІПС) мають бути задіяні багато інструментів комп'ютерної лінгвістики, що охоплюють внутрішні інформаційно-когнітивні процеси та програмні засоби обробки інформації.

Важливим напрямом вдосконалення функціонування сучасних ІПС має стати аналіз існуючих засобів комп'ютерної лінгвістики, які знаходяться на стику лінгвістики, математики, інформатики та штучного інтелекту. Серед них – штучні мови або інформаційно-пошукові мови (ІПМ), які дозволяють ефективно здійснювати пошук релевантної інформації. Аналіз сучасних ІПМ та їх різновидів є важливим для оптимізації пошуку не тільки текстової, але й будь-якої іншої інформації в межах будь-яких пошукових систем.

Аналіз останніх досліджень і публікацій. Найбільш значущими для дослідження засобів комп'ютерної лінгвістики, зокрема в інформаційно-пошукових системах, є праці українських вчених та дослідників-практиків, зокрема Бідюк Н. М., Комочкова О.О., Денисюк В. В., Цимбал Н. Т. В публікаціях цих науковців досліджуються докладно базові теоретичні питання прикладної лінгвістики, її напрями, зокрема сучасний стан та перспективи розвитку комп'ютерної лінгвістики. В наукових працях зарубіжних дослідників Крофта В., Гарпера Г., Гріффа В. розкриваються питання використання ймовірнісних моделей пошуку документів, теорії зважування термів в процесі аналізу даних тощо.

Тож наявні напрацювання дослідників дозволяють визначити засоби комп'ютерної лінгвістики як напряму прикладної лінгвістики в різних галузях, проте залишаються недостатньо вивченими лінгвістичні засоби в межах розвитку сучасних автоматизованих інформаційно-пошукових систем, зокрема в умовах систем штучного інтелекту.

Метою статті є аналіз та характеристика засобів комп'ютерної лінгвістики, які використовуються в умовах автоматизованих інформаційно-пошукових систем.

Виклад основного матеріалу. Визначення комп'ютерної лінгвістики (computational linguistics) тлумачиться, на думку багатьох вчених та фахівців, як прикладна лінгвістична дисципліна; як широка область використання комп'ютерних інструментів – програм, комп'ютерних технологій організації та обробки даних – для моделювання функціонування мови в тих чи інших умовах, ситуаціях, проблемних областях, а також сфера застосування комп'ютерних моделей мови не тільки в лінгвістиці, а й у суміжних із нею дисциплінах [1].

Про це свідчать численні міжнародні конференції та симпозиуми, наукові журнали з комп'ютерної лінгвістики (в США та інших країнах). В той

же час, на практиці сфера комп'ютерної лінгвістики охоплює практично все, що пов'язано з використанням комп'ютерів у мовознавстві. Так, до напрямів комп'ютерної лінгвістики відносять: автоматичний аналіз текстів; автоматичний синтез текстів; створення та підтримка автоматичних словників; створення автоматизованих інформаційно-пошукових систем (ІПС); машинний переклад; створення автоматичних систем навчання мові; автоматична атрибуція та дешифрування анонімних текстів; створення лінгвістичних баз даних; розробка програмних інструментів для вирішення задач теоретичної та прикладної лінгвістики тощо [1].

Процес інформаційного пошуку, який закладений в основі створення та функціонування автоматизованих ІПС, охоплює багато компонентів. Його активно використовують зокрема соціокомунікативні інституції, перш за все – електронні бібліотеки та електронні архіви в умовах інтернет-комунікаційного середовища. Так, в умовах бібліотек залежно від мети розрізняють інформаційний пошук адресний та семантичний. Адресний пошук визначається як процес пошуку інформаційних повідомлень за суто формальними ознаками, зазначеними у запиті користувача (адресами можуть виступати елементи бібліографічного опису документів, їх авторські знаки, інвентарні номери та ін.). Семантичний пошук розглядається як процес пошуку інформаційних повідомлень за їх змістом. Однією з умов здійснення семантичного пошуку є переклад змісту інформаційних повідомлень (і запитів) з природної мови на інформаційно-пошукову мову (ІПМ), тобто йдеться про кодування інформації.

У процесі кодування співвідношення слова (символу) та його значення відносять до понять семантики, тобто правила, що виражають загальні синтаксичні властивості слів та груп слів; а також – граматики, як правила, що дозволяють виробляти та/або описувати правильні речення мови.

Відомо, що інформація може кодуватись вербально або невербально. Різниця між способами кодування інформації (аудитивний, тактильний, візуальний і т.д.) обумовлює безліч способів її подання: тексти; малюнки, креслення, фотографії; світлові чи звукові сигнали; електричні та нервові імпульси; жести та міміка; запахи та смакові відчуття; хромосоми і т.д. А власне комп'ютерні програми можуть обробляти всі відомі види інформації, включаючи: числову, літерну (вербальну), графічну, звукову, відеоінформацію.

У сучасному світі, з постійно зростаючими обсягами інформації, для людини, що використовує цю інформацію з метою її перетворення у знання, постає проблема орієнтації, тобто – пошуку. Тому актуальними стають техніки відбору, фільтрації та оцінки інформації. Традиційними способами фільтрації та відбору інформації людиною є: пошук «зверху» (за змістом); пошук «знизу» (за допомогою різних показників); пошук за допомогою гіпертекстових

зв'язків (перехресних посилань); повнотекстовий пошук шляхом перегляду всього тексту. Останній вид пошуку є найбільш точним, але й найбільш трудомістким, оскільки вимагає найбільше часу та зусиль [4].

Відзначимо, що одна із назв – інформаційно-пошукові системи (ІПС) отримана як упорядкована сукупність документів та інформаційних технологій, призначених для зберігання і пошуку інформації, представленої у вигляді текстів чи їх частин (фактів) [1]. Призначена ІПС переважно для пошуку текстів (документів, їх частин, фактографічних записів) в інформаційних сховищах (базах даних) за формальними характеристиками. В роботі ІПС виділяють етапи збору і зберігання інформації та етап пошуку і видачі інформації користувачам.

Такі поняття щодо ІПС можна вважати узагальненими, оскільки вони охоплюють процеси пошукової системи, які можна здійснювати як у автоматизованому режимі, так і в не автоматизованому (наприклад, бібліотечні або архівні паперові каталоги та картотеки).

Звернемося до сучасних автоматизованих (або комп'ютерних) ІПС, які для здійснення інформаційного пошуку використовують такі засоби комп'ютерної лінгвістики, як декларативні та процедурні. До декларативних відносять словники одиниць мови й мовлення, тексти і різноманітні граматичні таблиці, до процедурних – засоби маніпулювання одиницями мови і мовлення, текстами і граматичними таблицями. Комп'ютерний інтерфейс стосується процедурного компонента комп'ютерної лінгвістики.

Аналіз процедурних засобів комп'ютерної лінгвістики (комп'ютерного інтерфейсу та програм) базується на твердженні про те, що комп'ютерна лінгвістика утворилася на стику лінгвістики, математики, інформатики та штучного інтелекту. Завдяки цьому з'явилися такі відомі технології як пошукові системи, голосові помічники, перекладачі текстів та багато інших. Перед комп'ютерною лінгвістикою стоять задачі лінгвістичного забезпечення процесів збирання, накопичення, обробки та пошуку інформації. Найважливіші з них: розпізнавання та синтез мови, генерування тексту, автоматичний переклад тексту, створення та використання електронних корпусів текстів, вилучення інформації з тексту, автореферування, автоматичне виявлення та виправлення помилок під час введення тексту на комп'ютері, створення систем «питання-відповідь», створення електронних словників і тезаурусів.

У основі кожної ІПС покладено процеси вдосконалення засобів накопичення та опрацювання знань. Концептуальна схема ІПС має вигляд, представлений на рисунку 1.

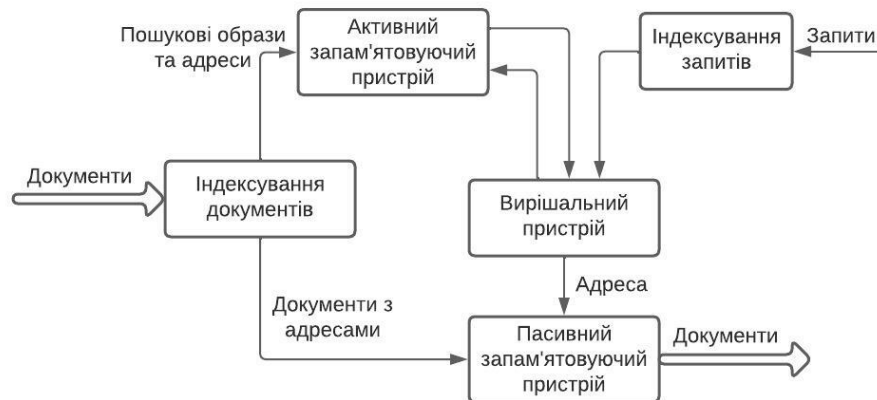


Рис. 1. Концептуальна схема ІПС

Аналіз алгоритмів пошуку інформації в сучасних ІПС мережі Інтернет показав, що пошукові Інтернет-провайдери вивчають зміст сторінок, щоб розрахувати їх змістовну вагу і релевантність при відображенні в результатах пошуку. Одним з механізмів вимірювання змістовної ваги сторінки серед інших сторінок є алгоритм PageRank, розроблений компанією Google, відповідно до якого чим більше сторінок посилаються на створену сторінку, тим вище її змістовна вага, особливо, якщо сторінки, що посилаються, в свою чергу також мають високу змістовну вагу.

Для побудови моделі пошуку зазвичай використовують наступний понятійний апарат:

- множина документів, на якій виконується пошук $\{d_1, d_2, \dots, d_n\}$;
- множина термів $\{T(d_i)\} = U_{i=1}^n T(d_i)$, де $T(d_i)$ – набір слів (термінів), що входять в документ d_i , за якими виконується пошук.

На практиці для ранжування документів слід використовувати векторно-просторову модель [7], яка представляє запит та документи, в яких відбувається пошук, у вигляді n -мірного вектору у евклідовому просторі термів. При цьому модулем вектору є деяка величина, що визначає його вагу (значимість) [8]. У якості ваги w_{ij} терму $T_i(d_j)$ часто використовують нормовану частоту його використання $freq_{ij}$: $w_{ij} = \frac{freq_{ij}}{\max(freq_{ij})}$. Ця модель полегшує вирішення питання тематичної ідентичності документів.

Іншим способом визначення ваги сторінки є розрахунок значень набору атрибутів – наприклад, значення частоти появи терміну з пошукового запиту в URL-адресі сторінки, у заголовку та мета-даних сторінки в текстах посилань. Для складних запитів функції ваги враховують кілька різних термінів, які відображаються поруч на сторінках. Типові алгоритми зважування веб-сторінок враховують сотні або тисячі атрибутів.

Крім аналізу інформації в мережі Інтернет, пошукові системи також аналізують інформацію про пошукові запити – терміни, за якими виконується

пошук – для визначення рекламних оголошень, які можуть зацікавити користувача і забезпечити перехід на сайт рекламодавця. Переходи на сайти також враховуються для поліпшення подальших результатів пошуку.

Сучасні ІПС характеризуються певною простотою пошуку для користувача: користувач вводить ключові слова та отримує у відповідь документи на задану тему. Однак простота ця тільки здається, бо за всім цим стоять складні спеціальні методи та алгоритми пошуку.

Узагальнюючи вище перелічене, слід акцентувати увагу саме на когнітивному інструментарії комп'ютерної лінгвістики, тобто за напрямками використання комп'ютерних засобів обробки мовних даних. Оскільки комп'ютерні програми, що моделюють ті чи інші аспекти функціонування мови, можуть використовувати різні засоби програмування, а про загальну метамову не говорять. Однак існують загальні принципи комп'ютерного моделювання мислення, які так чи інакше реалізуються в будь-якій комп'ютерній моделі. В основі цієї мови лежить теорія знань, розроблена в штучному інтелекті і утворює важливий розділ когнітивної науки.

Система «штучного інтелекту» має розуміти питання людини, вирішувати інтелектуальні задачі та вести діалог із людиною на основі закладених у ній процедурних та декларативних знань. Прикладом інтелектуальної штучної системи є експертна система, якість якої визначається насамперед тим, наскільки природним є спілкування з нею людини при вирішенні задач. У штучному інтелекті важливим є поняття «знання», а при побудові інтелектуальних систем – мови подання знань.

Знання, як правило, включають 3 складові – досвід, навички та вміння – і бувають декларативними («знаю ЩО це») і процедурними («знаю ЯК це робити»). Знання в інтелектуальній системі представляють спеціальними моделями на штучних мовах лінгвістичного забезпечення цієї системи (див рис. 2).

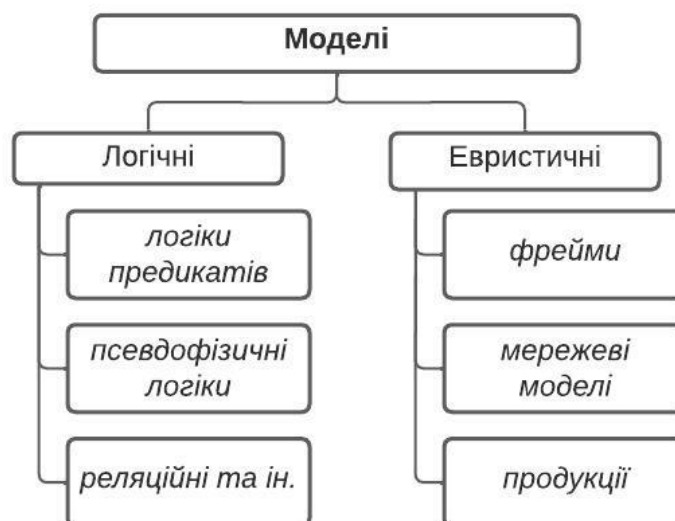


Рис. 2 Моделі подання знань

Для прикладної лінгвістики характерні мови подання знань та семантики (МПЗ). МПЗ – це штучні мови, побудовані за законами штучних мов. Найбільш популярні МПЗ – це логічні, мережеві, фреймові [2].

Коротко охарактеризуємо кожну з таких мов.

Логічні МПЗ представляють знання у вигляді синтаксично правильних формул будь-якої формальної логічної системи. Логічних систем дуже багато. Розроблено навіть псевдофізичні логіки, тобто системи правил, що описують відношення реального світу, наприклад, логіки часу, причини та наслідки, логіки просторових відношень. Часто використовують логіку предикатів та її мову логічних формул. Вирішення задач у логіці на логічній МПЗ – це логічний (дедуктивний) висновок за правилами цієї логічної системи.

Мережеві МПЗ як моделі використовують семантичні мережі, де вузли мережі – це будь-які інформаційні одиниці – поняття, факти, процеси, імена, а дуги – відношення між ними. Відношення можуть бути будь-якими (тимчасові, причинно-наслідкові, більш-менш тощо). Мережеві МПЗ часто використовують для явного опису відношень у тій чи іншій ситуації чи опису семантики структур. Розв'язання задач на мережевих моделях зводиться до пошуку фрагмента мережі, що збігається з даним зразком та до організації логічного виведення на семантичній мережі. Зв'язки елементів семантичної мережі можна представити у вигляді формул – записів на метамові семантичних мереж, тому що графічно можна представити наочно лише прості структури.

Фреймові МПЗ базуються на понятті «фрейму», яке у класичному розумінні пов'язане з мінімальним описом факту чи явища, у якого не можна видалити жодної частини опису без того, щоб не втратити повноту цього опису. Фрейм може бути рольовий чи структурний. Структурний фрейм більшою мірою відображає декларативні знання, тобто описує структуру будь-якого поняття, об'єкта чи документа. Рольовий кадр, на відміну від структурного, представляє процедурні знання, наприклад, скрипти чи сценарії будь-яких типових процедур чи робіт. Скрипти широко використовуються у програмуванні бізнес-додатків та стандартних автоматів, наприклад, при оплаті послуг у банкоматах або для дистанційних операцій. Одна з найскладніших областей використання рольових кадрів – це робототехніка.

Основні характеристики цих МПЗ – це компактність та вкладеність рівнів.

Тож для здійснення інформаційного пошуку в автоматизованих ІПС вище розглядалися процедурні засоби комп'ютерної лінгвістики – в контексті систем штучного інтелекту, зокрема штучні мови або мови подання знань. Їх важливість обумовлена сучасними тенденціями до інтелектуалізації ІПС та персоналізації пошуку. Саме таким шляхом йдуть сьогодні великі компанії на ринку пошукових систем, які активно впроваджують різні сервіси та

інструменти персоналізації та інтелектуалізації пошуку. Зазначені механізми пошуку спрямовані на усунення основних недоліків традиційного контекстного пошуку за ключовими словами.

На шляху створення персоналізованих інтелектуальних ІПС стоїть багато невирішених проблем, але розробка подібних систем дуже актуальна, оскільки вони мають низку переваг, але найголовніше в тому, що такі системи прямують до вирішення проблеми вибору релевантної інформації. Необхідно наголосити, що для підвищення ефективності ІПС інтелектуальні та персоналізовані механізми пошуку повинні застосовуватися комплексно разом із традиційними пошуковими засобами [2].

Останнім часом пошук почав використовувати нейронні мережі для того, щоб знаходити документи не за словами, які були вказані в запиті чи в самому документі, а за смисловою значущістю запиту та заголовку. Хоча останні десятиріччя дослідники працювали над проблемою семантичного пошуку, в якому документи ранжуються, виходячи з смислової відповідності запиту.

Звертаючись до власне інформаційного пошуку (Information Retrieval), слід підкреслити, що у якості засобів зберігання та пошуку інформації виступають інформаційно-пошукові системи (IRS/Information Retrieval Systems), елементами яких є структурований масив документів (база даних/індекс), що виступають як об'єкт пошуку, використовуючи різноманітні технічні та програмні засоби, як наприклад, програми-роботи, а також інформаційно-пошукова мова, що задає правила індексування документів, правила пошуку. При введенні документа в базу даних ІПС його індексують; процеси індексування переважно пов'язані з визначенням і вибіркою ключових слів оброблюваних документів і відображенні їх формально у вигляді пошукового образу [3]. Так база даних ІПС складається з множини індексних пошукових образів. Особливо важливим є поняття «релевантність» як фундаментальне поняття теорії інформаційного пошуку, яка вказує на повну відповідність смислової близькості між документом та інформаційним запитом.

Тобто у традиційному пошуку йдеться про інформаційно-пошукову мову (ІПМ), яка розглядається як спеціалізована штучна мова, призначена для: опису інформаційних запитів до інформаційно-пошукових систем (мова запитів); опису формальних показників документів як пошукового образу, що зберігається в основі системи (мова індексування). Необхідність впровадження штучних мов викликана необхідністю усунення надмірності природної мови для інформаційного пошуку, а також ліквідації мовної синонімії, омонімії та неоднозначностей різного роду. Інформаційно-пошукова мова, як і будь-яка мова, складається з фіксованих одиниць, наприклад, має свій словник та синтаксис, і є штучною мовою, тобто

обмеженою за своєю формою і структурою відносно задач, що стоять перед ним для пошуку. Слід також зазначити, що ІПМ слугують для запису інформації, її накопичення, упорядкування, аналізу і видачі за запитом користувача. Це різні алфавітно-предметні покажчики і каталоги, патентні класифікації, системи індексування (того чи іншого корпусу документів), хронологічні таблиці, системи зберігання фактів (що стосуються будь-якої предметної області), бази даних (тобто організовані сукупності фактів, понять і суджень, що допускають застосування нових даних та їх поповнення) та інші інформаційні системи.

Отже, виявлено, що будь-яка ІПМ на внутрішньому змістовному рівні базується на лінгвістичних складових (алфавіт, лексика, граматики), які у складі інформаційних одиниць пошуку через різні засоби – кодуються, обробляються, декодуються. Штучні мови для автоматизованих ІПС, які безумовно враховують змістовну складову, стають більш ускладненими завдяки забезпеченню різними комп'ютерними програмами, які включають відповідні засоби комп'ютерної лінгвістики, та розвиваються на інтелектуально-логічному рівні. (див. табл.1).

Таблиця 1.

Інформаційно-пошукові мови та засоби комп'ютерної лінгвістики інформаційно-пошукових систем

Традиційні ІПМ, (за ознаками: лексики, парадигматики та синтагматики)	Засоби комп'ютерної лінгвістики, що використовуються в ІПС	Інтелектуальні інформаційно- логічні мови (ІПЛ) – (мови подання знань та семантики на основі штучних мов, з використанням комп'ютерних програм)
Дескрипторні ІПМ	Автоматичний аналіз та синтез тексту Автоматичне розпізнавання тексту Автоматичне індексування та класифікація тексту Автоматичне реферування тексту	Логічні ІПЛ
Контрольовані та неконтрольовані ІПМ		Мережеві ІПЛ
Некоординовані та координовані ІПМ		Фреймові ІПЛ
Ієрархічні ІПМ (УДК)		Процедурні ІПЛ
Неієрархічні ІПМ		та інші ІПЛ
Фасетні ІПМ		

Сучасне людство використовує безліч інформаційно-пошукових мов. Серед них – інформаційно-логічні мови, розроблені задля суто логічного, стисненого та універсального (наднаціонального) представлення науково-технічної інформації. Вони створювалися як логічні обчислення розумових висновків і використовували логіко-математичну (міжнародну) символіку. На практиці, під час розробки автоматизованих (людино-машинних) систем

управління, корисними виявилися конкретні системи, придатні для символічного представлення обмеженого фрагмента реальності. Проблематиці інформаційно-логічних мов близькі задачі з розробки процедур (у т.ч. автоматичних) згортання і розгортання наукової інформації у галузях документно-графічного, фактографічного і концептографічного інформаційного обслуговування (інформативне і метаінформативне згортання тощо).

Висновки. У результаті дослідження визначено та охарактеризовано засоби комп'ютерної лінгвістики, які використовуються в умовах автоматизованих інформаційно-пошукових систем (ІПС). Серед багатьох компонентів акцентується увага на головних для пошуку в ІПС – інформаційно-пошукових мовах (ІПМ). Підкреслено, що усі інформаційні мови в автоматизованих ІПС використовують через програмні засоби, комплекс засобів комп'ютерної лінгвістики (декларативні та процедурні), які надають очікуваний пошуковий результат – релевантну інформацію у відповідь на запит. Інформаційні комп'ютерні системи побудовані та працюють на базі штучних мов, а з визначення ІПС зрозуміло, що її основним лінгвістичним засобом є спеціалізована штучна мова, яка розвивається та вдосконалюється на когнітивному рівні.

ІПМ, умовно кажучи «традиційні», акумулюють в собі лінгвістичні засоби, які орієнтовані на текстову інформацію переважно: кодування, аналіз, згортання, розпізнавання тощо. Інтелектуальні ІПМ можна віднести до перспективних лінгвістичних засобів в ІПС, зокрема для пошукових систем багатьох електронних бібліотек або електронних архівів, бо вони використовують комп'ютерні програми автоматичної обробки і синтезу не тільки текстової, але й інших видів інформації: числової, невербальної, графічної, аудильної, візуальної, відеоінформації та інші.

Перспективними напрямками дослідження є аналіз лінгвістичних складових персоналізованих інтелектуальних ІПС, які орієнтовані на оптимізацію інформаційного пошуку в різних сферах діяльності.

Література:

1. Бідюк Н. М. Основи прикладної лінгвістики: вступ до спеціальності : навч. посіб. Хмельницький : ХНУ, 2017. 191 с.
2. Брусенцев В.О. Прикладна лінгвістика : консп. лекцій. Харків: ХДАК, 2021. 76 с.
3. Денисюк В. В. Інформаційні технології в лінгвістиці. Умань : ФОП Жовтий О. О., 2015. 135 с.
4. Прикладна лінгвістика : навчальний посібник; упорядник: Н. А. Цимбал. Умань: Візаві, 2019. 106 с.
5. Селіванова Л. Л. Сучасна лінгвістика: напрями та проблеми. Полтава: Довкілля. К, 2008. 712 с.
6. Семотюк О. Л. Сучасні технології лінгвістичних досліджень : навч. посібник . Львів : вид-во Львівської політехніки, 2011. 152 с.

7. Croft W.B., Harper D.J. Using probabilistic models of document retrieval without relevance information // *MCB UP Ltd, Journal of documentation*. 1979. 35(4). 285-295 pp.

8. Greiff W. R. A theory of term weighting based on exploratory data analysis // In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval / ACM*, 1998. 11-19 pp.

References:

1. Bidiuk, N. M. (2017). *Osnovy prykladnoi linhvistyky: vstup do spetsialnosti [Foundations of Applied Linguistics: Introduction to the Specialty]: navch. posib.* Khmelnytskyi: KhNU, 2017. 191 [In Ukrainian].

2. Brusentsev, V.O. *Prykladna linhvistyka [Applied Linguistics]: konsp. leksii.* Kharkiv: KhDAK, 2021. 76. [In Ukrainian].

3. Denysiuk, V. V. *Informatsiini tekhnolohii v linhvistytsi [Information Technology in Linguistics]*. Uman: FOP Zhovtyi O. O., 2015. 135 [In Ukrainian].

4. *Prykladna linhvistyka [Applied Linguistics]: navchalnyi posibnyk; uporiadnyk:* N. A. Tsymbal. Uman: Vizavi, 2019. 106 [In Ukrainian].

5. Selivanova, L. L. *Suchasna linhvistyka: napriamy ta problem [Modern Linguistics: Directions and Problems]*. Poltava: Dovkillia. K, 2008. 712 [In Ukrainian].

6. Semotiuk, O. L. *Suchasni tekhnolohii linhvistychnykh doslidzhen [Modern Technologies in Linguistic Research]: navch. posibnyk .* Lviv : vyd-vo Lvivskoi politekhniki, 2011. 152 [In Ukrainian].

7. Croft, W.B. & Harper, D.J. Using probabilistic models of document retrieval without relevance information // *MCB UP Ltd, Journal of documentation*. 1979. 35(4). 285-295 pp.

8. Greiff, W. R. A theory of term weighting based on exploratory data analysis // In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval / ACM*, 1998. 11-19 pp.