

**Омецинська Н.В.**

Таврійський національний університет імені В.І. Вернадського

**Боженко М.І.**

Таврійський національний університет імені В.І. Вернадського

## ПЕРСОНАЛІЗАЦІЯ ПОШУКУ НА ОСНОВІ АНАЛІЗУ ЗАПИТІВ КОРИСТУВАЧІВ

*У статті розглянуто підвищення ефективності пошуків сучасних інформаційно-пошуковою системою (ІПС). На сьогоднішній день (ІПС) є основним інструментом для пошуку інформації у Всесвітньому павутинні. Тому поняття релевантності має важливе значення для підвищення якості пошуку інформації в ІПС.*

*В статті визначено, що персоналізація пошуку та інтелектуалізація персоналізації неухильно зростають. З цією метою основні гравці сучасного ринку пошукових систем Інтернету активно впроваджують різноманітні сервіси та інструменти для персоналізації та інтелектуалізації пошуку. Ці пошукові механізми покликані усунути основні недоліки традиційного контекстного пошуку за ключовими словами.*

*З'ясовано, що в розробці персоналізованих інтелектуальних пошукових систем ІПС є ще багато невирішених питань, розвиток цих систем має багато переваг, механізми інтелектуального пошуку та персоналізації пошуку повинні використовуватися у поєднанні з традиційними засобами пошуку.*

*Зрозуміло, що в способах створення персоналізованих інтелектуальних ІПС існує безліч невирішених проблем. Розробка ІПС систем дуже важлива. Це пов'язано з тим, що така система має багато переваг, але найголовніше – вона наближає нас до вирішення проблеми відбору релевантної інформації.*

*Було встановлено, що такі системи наближають нас до вирішення проблеми відбору відповідної інформації; слід підкреслити, що для підвищення ефективності ІПС інтелектуальні та персоналізовані механізми пошуку повинні застосовуватися в інтегрованому вигляді спільно з традиційними засобами пошуку.*

**Ключові слова:** інформаційно-пошукова система (ІПС) штучний інтелект (ШІ), база даних (БД), бібліотека прецедентів (БП), пошукова система (ПС), інформаційний пошук (ІП, розподілений штучний інтелект (DAI), штучне життя (AL).

**Постановка проблеми.** Останніми роками неухильно зростає інтелектуалізація та персоналізація пошуку в ІПС Інтернету. Ці пошукові механізми покликані усунути основні недоліки традиційного контекстного пошуку за ключовими словами. Для забезпечення персоналізації та інтелектуалізації пошуку пропонується використовувати методи штучного інтелекту, зокрема мульти-агентні підходи, а також методи та інструменти обґрунтування правдоподібності на основі прецедентів (CBR-Case-Based Reasoning).

**Аналіз останніх досліджень і публікацій.** Інформаційно-пошукова система (ІПС) – це система пошуку та відбору необхідних даних зі спеціальної бази даних, що описує джерела інформації (індекси) на основі інформаційно-пошукової мови та відповідних правил пошуку [1, 2]. Пошук здійснюється відповідно до інформаційних потреб користувача, сформованими у вигляді запиту. Дуже важливо, щоб в результаті пошуку нічого не було втрачено, тобто, що в індексі були знайдені всі документи, відповідні запиту (повнота

пошуку), і щоб не було видано нічого зайвого (точність пошуку).

**Метою статті** є дослідження методу інформаційного пошуку та програмного засобу для ІПС на основі прецедентів. У статті розглядалися такі завдання:

- вивчення основних особливостей характеристик ІПС, а також різних методів пошуку інформації в ІПС;
- аналіз проблем побудови ІПС Інтернету та можливих шляхів їх вирішення.

**Виклад основного матеріалу.** Пошукові системи є найбільш поширеними доступним ресурсом для пошуку інформації в Інтернеті. І сьогодні всі сучасні ІПС представляють собою мульти-агентні (багатоагентні) системи (МАС).

МАС – це технологія, що сформувалася на стику штучного інтелекту (ШІ), загальної теорії систем, інформаційних технологій і комунікаційних технологій.

При поповненні бази даних пошукові системи використовують програмних агентів (мобільних

агентів) павуків, які представляють собою мережевих роботів. Павуки переміщуються від сайту до сайту в Інтернеті, які продовжують його індексувати.

Основними характеристиками пошукових систем є обсяг індексування, мова запитів користувача, подання вихідних та вхідних документів, час індексування та пошуку.

Як правило, пошукові системи надають інтерфейс типу «текстовий скринька-кнопка», який дозволяє користувачам створювати інформаційно-пошукові запити з використанням ключових словичних зв'язок. Більшість машин знаходять величезну кількість «пов'язаних» із запитом користувача сторінок, за допомогою різних методів, включаючи частоту зустрічності пошукових термінів на сторінці, «відстань» між пошуковими термінами, зміст META-тегів, просторовий і тимчасовий контекст документа, популярність ресурсу в оцінці та використання індексації цитування.

Типову конфігурацію пошукової системи можна побачити на прикладі машини WebCrawler, розробленої у Вашингтонському університеті (Сіетл, США).

Машина WebCrawler починає процес пошуку нових сайтів з відомих документів і переходить за посиланнями на інші сторінки. Він розглядає мережевий простір як спрямований граф, використовує алгоритм обходу графіка і працює наступного циклу:

- знайти нові документи;
- позначити витягнутий документ;
- декодувати посилання з цього документа;
- індексувати вміст документа.

Пошукова система працює у двох режимах: пошук документів у реальному часі і індексування документів. У режимі індексування система створює індекс інформації з витягнутих документів; в режимі пошуку-знаходить документи, що найбільш повно відповідають запиту користувача.

Агент системи WebCrawler відповідає за пошук документів з мережі. Для виконання цієї задачі пошукова система знаходить вільного агента і передає йому завдання на пошук. Агент починає роботу і повертає або вміст документа, або пояснення, чому документ не може бути доставлений. Агенти працюють як окремі процеси, тому основний процес системи ізольований від помилок і проблем з пам'яттю. Одночасно використовуються до 15 агентів.

У базі даних зберігаються метадані документів, посилання між ними повнотекстовий індекс. База даних оновлюється при кожному надходженні нового документа. Для відсікання семан-

тично неважливих слів є стоп-словник. Словом у документі присвоюється вага, що дорівнює частоті їх зустрічальності в даному тексті, поділеної на частоту їх зустрічальності в посиланнях на інші документи. Такий індекс дозволяє користувачеві швидко знаходити посилання на документи, що містять це слово.

Аналогічним чином структуровані та інші пошукові системи. Мережевим роботам стає все важче справлятися з постійним збільшенням ресурсів Інтернету, так як коні не може до оптуватися до уподобань користувачів і не мають достатніх засобів для аналізу інформації. Основним завданням пошукових систем є індексування ресурсів глобальної мережі. Дійсно, в базах даних пошукових систем зберігається інформація про те, що можна знайти в Інтернеті.

Основними напрямками розвитку MAC (рис. 1) є розподілений штучний інтелект (DAI) і штучне життя (AL). Ядром РІІ є вивчення взаємодії і координації невеликого числа інтелектуальних агентів, як, наприклад, в класичних інтелектуальних системах, що включають бази знань і вирішувачі; тут інтелектуальна поведінка групи формується на основі інтелектуальної поведінки індивідумів. Це включає в себе узгодження цілей, інтересів і стратегій різних агентів, координацію їх дій і вирішення конфліктів шляхом переговорів. Теоретичною основою тут є результати психології малих груп і організаційної соціології.

Важливим розділом РІІ є кооперативне розподілене розв'язання задач (CDPS).

Вони вирішують проблеми, які виходять за межі їх індивідуальних можливостей. Різні вузли такої мережі зазвичай мають різну експертизу (знання, погляди) і різні ресурси. Кожен вузол повинен вміти змінювати свою поведінку в залежності від ситуації і планувати стратегії комунікації та співпраці з іншими вузлами. Тут показниками рівня кооперації є характер розподілу завдань, інтеграція різних точок зору і, звичайно, можливість вирішення загальної проблеми за заданий час.

Розподілене рішення задач декількома агентами можна розділити на наступні етапи.

- 1) менеджер агента (центральний орган) декомпозує вихідну проблему на окремі завдання;
- 2) завдання розподіляються між агентами-виконавцями;
- 3) кожен агент-виконавець вирішує своє завдання і, можливо, розбиває її на під завдання
- 4) здійснюється композиція приватних результатів, відповідних поставленим завданням, для отримання загального результату.



Рис. 1. Класифікація мультиагентних систем

Другий напрямок – «Штучне життя» – більшою мірою пов'язане з трактуванням розумної поведінки в контексті виживання, адаптації і самоорганізації в динамічній і ворожій середовищі сходиться до робіт Ж. Піаже [40]. В контексті ІЖ глобальне розумне поведінка всієї системи спирається на ряд простих, не обов'язково розумних агентів. Тут також використовуються терміни «колективний інтелект» та «роєвий інтелект». Прибічники цього напрямку, в частности Р. Брукс, Л. Стілс, Дж. Денебергі інші, спираються на такі положення:

- 1) МАС являє собою взаємозалежну групу агентів;
- 2) кожен агент самостійно визначає свої реакції на події в локальному оточенні і взаємодія з іншими агентами;
- 3) зв'язки між агентами горизонтальні, інтагента-куратора, що контролює взаємодію інших агентів;
- 4) не існує точних правил, що визначають глобальну поведінку агентів;
- 5) поведінка, властивості і структури на колективному рівні формуються тільки результатів телокальних взаємодій між агентами.

Реакції на навколишнє середовище і механізми локальної взаємодії, як правило, не включають в себе такі аспекти, як прогнозування, планування пізнання, хоча вони можуть бути здатні вирішувати складні проблеми.

Часто проводиться принципова відмінність між розподіленим і децентралізованим ІІІ.

У цьому випадку вихідним об'єктом дослідження є загальна складна проблема, для її вирішення формується група агентів, будується загальна концептуальна модель і вводяться глобальні критерії досягнення мети.

Основним об'єктом дослідження тут є вже не розподілене рішення загальних завдань, а діяльність автономних агентів (і координація діяльності різних агентів) в динамічному багато агентному світі. Наряду з розподіленими знаннями і ресурсами описуються локальні завдання окремих агентів, які вирішуються на базі локальних концептуальних моделей і локальних критеріїв.

Інтелектуальний агент – це програма, яка автономно виконує завдання, задані користувачем комп'ютера, через певні проміжки часу. Інтелектуальні агенти використовуються для допомоги операторам збору інформації. Прикладом завдання, що виконується агентом, є завдання постійного пошуку і збору необхідної інформації в Інтернеті. та навчання.

Агенти в розподіленому штучному інтелекті можуть розглядатися без урахування характеристик інших агентів, і проблема узгодженості знань поступається місцем проблемі забезпечення співпраці та комунікації між агентами. У багатьох випадках необхідний також фізичний розподіл завдань, наприклад, при використанні груп проботів.

Типова схема розподіленого розв'язання задач декількома агентами включає наступні кроки:

- Підлеглий агент (агент-менеджер) (супервізор, центральний орган) декомпозує вихідну проблему на окремі завдання.

– Ці завдання розподіляються між агентами-виконавцями.

– Кожен агент-виконавець вирішує своє завдання, іноді розбиваючи її на підзавдання.

– Для отримання загального результату проводиться синтез інтеграція окремих результатів, відповідних поставленим перед ними завданням. За синтез і інтеграцію загального результату відповідає агент-інтегратор (агент-інтегратор) результатів відповідає агент-інтегратор (у більшості випадків той же агент-субкоординатор (агент-менеджер)).

Два, найбільш важливі аспекти розподіленого штучного інтелекту – це розподіл завдань між агентами і інтеграція результатів. Після отримання приватних результатів виникає проблема їх узгодження та інтеграції.

Тут основними критеріями ефективності розподіленого розв'язання завдань є час, що витрачається на рішення, і сумісність під завдань з можливостями конкретного агента-виконавця.

У разі розподіленого рішення завдань залежний агент (агент-менеджер) може дотримуватися двох протилежних стратегій:

– вибір кращого агента-виконавця для вирішення конкретної підзадачі (вибора агентів для вирішення підзавдань);

– вибір підзавдання, найбільш підходящою для даного агента-виконавця (вибір підзавдання агентів).

Отже, основними характеристиками соціальної групи у розподіленому штучному інтелекті, тобто. групи, що складається з штучних агентів, що працюють спільно для досягнення загальної мети, є соціальна структура і розподіл ролей між агентами. У даному випадку соціальна структура формується в результаті розподілу ролей. «повноваження» та «відповідальність». Відповідно до цього підходу, організації створюються і розвиваються «зсередини» системи.

До основних теоретичних питань розподіленого штучного інтелекту ставляться такі:

– Опис, декомпозиція та розподіл завдань між агентами в багатоагентних системах. Синтез (композиція) рішень;

– Забезпечення взаємодії та комунікації агентів у багатоагентних системах. Побудова комунікаційних мов і протоколів;

– Координація прийняття рішень та координація поведінки агентів. Планування в багатоагентних системах. Вирішення конфліктів між агентами з приводу ресурсів;

– Представлення внутрішніх станів агентів і виведення знань, планів і дій інших агентів;

– Представлення різних перспектив, цілей і переваг агентів в багатоагентній системі.

Залежно від концепцій, обраних для організації МАС, зазвичай виділяють три основні класи архітектур:

– Архітектури, засновані на принципах та методах штучного інтелекту;

– Архітектури, засновані на поведінкових моделях з можливістю реагування на зовнішні події;

– Гібридні ієрархічні архітектури на основі поведінкових моделей методів ШІ.

Існує кілька причин, яким програмні агенти, мультиагентні системи (МАС) і технології на основі агентів є необхідними і корисними. Основною ідеєю програмних агентів є делегування повноважень. Для реалізації цієї ідеї агент повинен вміти взаємодіяти з власником або користувачем, отримувати відповідні завдання і повертати результати, орієнтуватися в середовищі виконання і приймати рішення, необхідні для виконання завдання.

Оскільки онтологія є точним представленням описуваної предметної області, агенти, які використовують онтології у своїй роботі, класифікуються як *deliberative*.

У динамічному підході МАС-додатки використовують парадигму мобільних агентів.

Мобільний агент – це програма, здатна переміщатися по мережі, наприклад по World Wide Web. Мобільний агент покидає комп'ютер клієнта і відправляється на віддалений сервер для виконання будь-якої дії, а потім повертається назад.

Існуючі МАС також можна розділити на централізовані, коли обробка даних здійснюється централізовано на одному вузлі, і розподілені, коли дані обробляються на різних вузлах залежно від характеру обробки інформації. Наприклад, первинна обробка відбувається на різних вузлах системи, після чого отримані дані відправляються на вузол.

До недоліків централізованих систем пошуку інформації в Інтернеті належать:

1) усі дані повинні бути повністю зібрані на якомусь вузлі системи для подальшої обробки, що призводить до пересилання по мережі великих обсягів даних;

2) обробка і зберігання великих обсягів даних на одному вузлі системи пов'язані з додатковими труднощами;

3) необхідність у потужних обчислювальних ресурсах;

4) проблеми узгодження даних;



5) виходу з ладу центрального вузла системи перешкоджатиме роботі всієї системи.

До недоліків розподілених систем можна віднести необхідність введення додаткових коштів, що гарантують взаємодію та координацію роботи кожного вузла системи.

Враховуючи перелічені переваги та недоліки різних типів архітектур мультиагентних систем, була обрана розподілена система, що складається з кількох серверів і клієнтів, що взаємодіють за допомогою передачі повідомлень. В якості агентної структури були обрані мобільні агенти, які «переміщуються» по мережі до серверів системи і виконують певні завдання. Такий вибір був зумовлений наступними міркуваннями:

1) процес обробки даних може бути прискорений за рахунок розподілу обчислювального навантаження за різними вузлами системи;

2) усі дані обробляються на одному злі, клієнтам мережі передаються лише результати

обробки, що виключає необхідність пересилання великих обсягів даних мережі;

3) висока надійність – при виході з ладу одного або декількох вузлів системи інші продовжують функціонувати, тому працездатність всієї системи не знижується.

При розробці МАС необхідно враховувати кілька принципів, які застосовувалися при побудові МАС для отримання файлів по локальній мережі. Перерахуємо ці принципи:

1) Розподілена структура МАС.

2) Механізм взаємодії компонентів МАС за допомогою повідомлень.

3) Механізми передачі агентів по мережі.

**Висновки.** У проведеному дослідженні розглянуто основних особливостей характеристик ІПС, а також різних методів пошуку інформації в ІПС та проаналізовано проблеми побудови ІПС Інтернету та можливих шляхів їх вирішення.

#### Список літератури:

1. Zong Woo Geem. Recent Advances in Harmony Search Algorithm. *Studies in Computation Intelligence*. 2016. P. 51–75.
2. Автоматизовані інформаційно-пошукові мови URL: <http://ubooks.com.ua/books/00092/inx13.php>.
3. Sergey Brin, Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, Stanfor. 1998. P. 45–48.
4. Google PageRank. URL: <https://ahrefs.com/blog/google-pagerank/>.
5. Page, Larry: PageRank: Bringing Order to the Web. Stanford Digital Library Project. 2002. P. 47–52.
6. Taher Haveliwala & Sepandar Kamvar: The Second Eigenvalue of the Google Matrix. Stanford University Technical Report. 2008. P. 20–34.
7. Омечинська Н. В., Гуйда О. Г., Прокопенко І. Ю. «Дослідження технологій інформаційного пошуку для використання в семантичній мережі», The 4th International scientific and practical conference “Current challenges of science and education” (December 11-13, 2023) MDPC Publishing, Berlin, Germany.
8. Омечинська Н.В., Дичко А.ОІ., Мінаєва Ю.Ю. «Проектування пошукового сервісу». URL: <https://www.calameo.com/books/003168372749b97cf65cb>

#### **Ometsynska N.V., Bozhenko M.I. PERSONALIZATION OF SEARCH BASED ON ANALYSIS USER REQUESTS**

*The article examines the improvement of the search efficiency of modern information and search systems (IPS). Today (IPS) is the main tool for finding information on the World Wide Web. Therefore, the concept of relevance is important for improving the quality of information search in the IPS.*

*The article determined that the personalization of search and the intellectualization of personalization are steadily growing. For this purpose, the main players of the modern Internet search engine market are actively implementing various services and tools for personalization and intellectualization of search. These search engines are designed to eliminate the main disadvantages of traditional contextual keyword searches.*

*It has been found that there are still many unsolved issues in the development of personalized intelligent IPS search systems, the development of these systems has many advantages, the mechanisms of intelligent search and search personalization should be used in combination with traditional search tools.*

*It is clear that there are many unsolved problems in the ways of creating personalized intelligent IPS. Development of PS systems is very important. This is due to the fact that such a system has many advantages, but the most important thing is that it brings us closer to solving the problem of selecting relevant information.*

*It was established that such systems bring us closer to solving the problem of selecting relevant information; it should be emphasized that in order to increase the effectiveness of the IPS, intelligent and personalized search mechanisms should be used in an integrated form together with traditional search tools.*

**Key words:** information search system (IPS), artificial intelligence (AI), database (DB), library of precedents (BP), search engine (PS), information search (IS), distributed artificial intelligence (DAI), artificial life (AL).