

О. В. Касілов, К. І. Крамська

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

## МОДЕЛІ І МЕТОД СИНТЕЗУ АГЕНТНОЇ ІНФОРМАЦІЙНО-ПОШУКОВОЇ СИСТЕМИ

**Анотація.** Існуючі сучасні інформаційно-пошукові системи такі як Google, Яндекс, Yahoo орієнтовані на середньостатистичного користувача, для якого предметом пошуку є неспецифічна, науково-технологічна інформація. Тому ці пошукові системи не можуть в повному обсязі надати потрібну інформацію для професійного користувача. Внаслідок того, що дана інформація є довідковою або енциклопедичною, вона не відповідає заданому пошуку. Забезпечити якісний пошук інформації і вирішити проблему інформаційно-аналітичного забезпечення наукової і технологічної діяльності можливо за допомогою використання агентних технологій. Агентні технології ініціюються і управляються автоматично, а також функціонують без участі користувача. Тому вони мають більш високу продуктивність і можливість безперервної роботи. Розроблена модель багаторівневої агентної інформаційно-аналітичної системи у природничо-наукових і технологічних напрямках. Вона є найбільш доцільною під час тематичного обслуговування колективних і персональних користувачів. Також в статті розглянуто використання агентних технологій в інформаційно-пошукових системах науково-технічної інформації. Запропоновано методику структурно-параметричного синтезу подібних систем.

**Ключові слова:** інформаційно-пошукові системи; агентні технології; агентні системи; агентний пошук.

### Вступ

Існуючі інформаційно-пошукові системи (ІПС) широкого використання не можуть повною мірою задовольнити користувача-професіонала, тому що вони надають довідкову та енциклопедичну інформацію по предмету, що не задовольняє критеріям новизни і пертинентності. За даними International Data Corporation (IDC), обсяг згенерованих даних в 2020 року прогнозується збільшення обсягу до 40 зеттабайт [1]. Незважаючи на спроби стандартизації інформаційних ресурсів, кількість різних систем доступу до цієї інформації постійно збільшується. Нові джерела науково-технічної інформації є вкрай привабливими для професійних користувачів, так як інформація на них з'являється з мінімальними затримками [2], що також дозволяє користувачеві входити в безпосередній контакт з авторами інформаційних повідомлень різного типу, опублікованих в таких джерелах. Проблемна ситуація полягає в тому, що використання діалогових систем пошуку і структуризації цієї інформації стає практично неможливим у зв'язку з великою кількістю джерел. Збір і структуризація традиційним діалоговим способом стала неможливою у зв'язку зі збільшенням кількості інформаційних джерел і швидкості їх публікації. Велика частина завдань якісного пошуку інформації та проблеми інформаційно-аналітичного забезпечення наукової і технологічної діяльності можуть бути вирішені при використанні агентних технологій.

**Мета роботи.** Розробка методу реалізації багаторівневої агентної інформаційно-аналітичної системи з природничо-наукових і технологічних напрямків та програмно-технічної реалізації моделі тематичного обслуговування колективних і персональних користувачів.

**Постановка задачі.** *Об'єкт дослідження* - агентні технології пошуку та обробки інформації в інформаційно-аналітичних системах. *Предмет дослідження* - методи і засоби агрегування тематичної інформації, а також засоби забезпечення тематичної

і лінгвістичної масштабованості мультиагентної системи. *Методи дослідження* - методи системного аналізу, структурно-параметричного синтезу, системного проектування, багатокритеріальні методи кількісної експертної оцінки складних об'єктів, методи тестування та експертна оцінка.

### Результати досліджень

**1. Метод структурно-параметричного синтезу агентних систем.** Нові джерела науково-технічної інформації є вкрай привабливими для професійних користувачів тому, що мають такі властивості:

1. Інформація на кожному сайті має унікальну структуру, яка визначається незалежними розробниками.

2. Власники інформації висувають різні умови для її придбання.

3. По кожному науковому тематичному напрямку кількість інформаційних джерел настільки велике, що окремий користувач фізично не може з ними усіма ознайомитися.

Перераховані вище властивості інформаційних джерел призводять до необхідності вирішення наступних завдань:

1. Визначення тематичних груп інформаційних джерел.

2. Рішення задач доступу до джерела.

3. Розробка алгоритмів автоматизованого сканування інформаційних джерел.

Також існують такі завдання, які є загальними для роботи з будь-якими електронними джерелами інформації, а саме, визначення першоджерела і оцінка достовірності інформації. Інформація, що міститься в електронних джерелах, може бути цільовим чином змінена, а в деяких випадках можуть бути наміряно спотворені метадані інформаційного повідомлення, такі як - дата публікації, автор, джерело і т. д. Сучасні інформаційно-пошукові системи (ІПС), такі як Google, Яндекс, Yahoo, орієнтовані на користувача, для якого спеціалізована природничо-наукова і технологічна інформація не є предметом пошуку [3]. При пошуку науково-технічної інфор-

мації перед користувачем неминує виникати завдання освоєння синтаксису пошукової системи, за допомогою якої можна будувати складні пошукові запити і отримувати найбільш релевантну видачу, ґрунтуючись на правильно підібраних ключових словах і логічних операціях між ними [4]. Рішення задач по освоєнню синтаксису пошукових систем - трудомістка робота, що вимагає постійної практики, тому середньостатистичний користувач мережі Інтернет не здатний використовувати сучасні інформаційно-пошукові системи на максимальному рівні [5]. Середня кількість ключових слів в типовому запиті користувача до пошукової системи складає від двох до трьох, а синтаксичні і логічні оператори практично не використовуються. Перераховані вище проблеми доступу користувачів до професійно-значущої інформації формуються зараз різними авторами як проблема Big Data.

Одним з класичних методів стиснення, концентрації і класифікації науково-технічної інформації є розробка реферативних баз даних, наприклад, Web of Science, Scopus (США). Ця база містить корисну інформацію про публікації в обраних ними джерелах інформації, наводять різні статистичні характеристики про процеси публікацій з різних напрямків, але їх істотним недоліком є відсутність текстів самих публікацій.

Велика розмірність завдань реферування, різний авторитет джерел інформації призводять концентраторів до необхідності оцінки авторитетності джерела інформації, яка може і не збігатися з оцінками конкретного користувача. Зокрема, концентратором може аналізуватися інформація, видана тільки на англійській мові з англомовних джерел, а джерела на інших мовах можуть не розглядатися, якщо відсутній переклад на англійську мову. Це призводить до зниження характеристики повноти інформації з тих чи інших тематичними напрямками.

Для професійних користувачів науково-технічної інформації важливі й інші відомості, наприклад ті, що містяться в патентних базах даних, або інформація про стан різних об'єктів користувальницького інтересу. Тобто необхідно враховувати можливість вилучення з первинної текстової інформації фактографічних даних, структурованих відповідно до професійних інтересів користувача.

Велика частина завдань якісного пошуку інформації та проблеми інформаційно-аналітичного забезпечення наукової і технологічної діяльності можуть бути вирішені при використанні агентних технологій. Агентні технології ініціюються і управляються автоматично, функціонують без участі користувача, тому мають більш високою продуктивністю і можливістю безперервної цілодобової роботи.

**2. Типологія користувачів і агентів.** В інформаційному описі науково-технічної діяльності найбільш стійко визначена її тематична спрямованість. Це властивість науково-технічної інформації використовується і в класичній десяткової класифікації, використовуваної в бібліотечних системах [6]. Стійкість тематичної класифікації пояснюється тривалістю життєвого циклу об'єктів наукового інтересу,

спеціалізацією науковців і інженерів. Тематична стійкість науково-технічної діяльності дозволяє фахівцям виробляти підмножина професійних мов, область за допомогою яких максимально точно і повно описуються об'єкти наукового дослідження і конструкторське-технологічної діяльності. Фактично це професійні підмножини природних мов. Терміни цих мов також мають високу стійкість, що дозволяє використовувати їх в якості ключових слів в автоматизованих інформаційно-пошукових системах. У розробці пошукових приписів автоматизованих систем відкривається можливість діяти за таким сценарієм: відобразити тематичні інтереси користувача ключовими словами з тематичного тезауруса; вказати мережеві джерела, що містять шукану тематичну інформацію, а далі проводити процес сканування джерел, виділення і обробку тематичної інформації без участі користувача; надавати користувачеві знайдену і структуровану інформацію. Тобто рішенням інформаційно-пошукових завдань в мережі Інтернет буде займатися не користувач, а його представник - агент [7].

У нашому випадку агент - це пошукова програма (або пошуковий робот), самостійно активізується в системі за заданим розкладом або запитом користувача і виконує дії як в мережі Інтернет, так і всередині агентної системи за задалегідь заданими приписами.

Користувачів науково-технічної інформації можна класифікувати на колективних і персональних користувачів.

У кожній з перерахованих груп користувачів своя роль в науці і свої інформаційні інтереси і вимоги до якості інформаційного обслуговування.

Перша група користувачів зацікавлена в повноті надходить науково-технічної інформації з усіх світових джерел. Агентні технології дозволяють вирішувати цю проблему шляхом автоматичного сканування всіх джерел.

Друга група користувачів зацікавлена не тільки в повноті, а й у тематичній точності інформації. Високий рівень цієї характеристики агентні технології забезпечують за рахунок автоматичної фільтрації всього вхідного потоку інформації.

Якість обслуговування персонального користувача характеризується параметром пертіненції одержуваної інформації. Цей параметр відображає ступінь задоволеності конкретного користувача отриманими із системи даними. Пертіненція забезпечується персональними агентами, які надають фахівцям інформацію за різними тематичними напрямками, які входять до сфери його професійних інтересів.

Таким чином, встановлюється відповідність між типами користувачів і типами агентів, які здійснюють інформаційне обслуговування. Це три типи агентів, що виконують різні функції в агентних пошукових системах.

1. Агент-листоноша (AP) - агент, який взаємодіє з інформаційними ресурсами мережі Інтернет за заданим розкладом і агрегує отриману інформацію в текстову базу даних.

2. Агент кур'єр (АК) - агент, який взаємодіє з агентом-листоношою і текстовою базою даних і систематизує інформацію за відповідними тематичними напрямками.

3. Агент-референт (AR)- агент, який взаємодіє з користувачем і агентом-кур'єром в цілях доставки пертінентної інформації з баз даних системи користувачеві.

Кожен з перерахованих вище типів агентів поділяється на функціональний і цільовий. Функціональний агент (індекс  $F$ , наприклад, функціональний агент-кур'єр – АКФ) виконує стандартизовану операцію за заданим розкладом. Цільовий агент (індекс  $C$ , наприклад, цільовий агент-кур'єр – АКС) виконує операцію по налаштованим користувачем параметрам. Особливості інформаційно-аналітичних систем з природничо-наукових і технологічних напрямках визначають і методологічні особливості їх розробки, а саме:

- створення ядра системи (одне або кілька тематичних напрямків і мов);
- тематичне масштабування (розширення) системи і лінгвістичне масштабування;
- введення декількох національних мов для моніторингу національних джерел науково-технічної інформації.

Ядро системи може бути розроблено відносно швидко і з порівняно невеликими фінансовими і трудовими затратами [8]. Подальше масштабування системи переводить її в стаціонарне стан, в якому регулярно реалізуються агентні інформаційні технології і цільові інформаційно-аналітичні дослідження, підтримувани класичними технологіями діалогового типу.

**3. Обґрунтування значень цільових параметрів агентної системи.** Параметрами якості інформаційно-аналітичного обслуговування виділених вище двох груп користувачів. Якість роботи автоматизованих інформаційно-пошукових систем оцінюється значеннями наступних параметрів: повнота, точність, пертінентність [9].

Введемо наступні позначення:  $A$  – видані документи, пертінентні запитом користувача;  $B$  – чи не видані документи, пертінентні запитом користувача;  $C$  – видані документи, які не пертінентні запитом користувача;  $D$  – чи не видані документи, які не пертінентні запитом користувача.

Припустимо, що загальна кількість документів, оброблюваних при пошуковому запиті –  $E$ , тоді  $A + C$  – кількість виданих документів,  $B + D$  - кількість не виданих документів, тобто

$$E = A + B + C + D. \quad (1)$$

Повнота інформаційного пошуку  $R$  визначається відношенням числа знайдених пертінентних документів до загальної кількості пертінентних документів, наявних у вихідній множині документів [10]:

$$R = A / (A + B). \quad (2)$$

Точність інформаційного пошуку  $P$  визначається відношенням кількості знайдених пертінентних документів  $A$  до загальної кількості документів, виданих на запит користувача [9]:

$$P = A / (A + C). \quad (3)$$

Наявність серед відібраних за результатами запитів користувача непертінентних документів – інформаційний шум системи. Коефіцієнт інформаційного шуму  $K$  є таким [9]:

$$K = C / (A + C). \quad (4)$$

Для кожного запиту до інформаційної бази значення показників будуть різними, тобто, строго кажучи, це випадкові величини, характеристики яких можуть бути отримані шляхом статистичних експериментів.  $A, B, C, D$  – параметри «видачі», тобто відповідь пошукової системи на запити (пошукові приписи) -  $Z$ , які є функцією:

$$Z_j = F_j(k_1, \dots, k_n, \dots, k_n), \quad (5)$$

де  $n$  – кількість ключових слів в конкретному запиті,  $k_i$  – ключове слово, або в нашому випадку - поняття з тематичного тезауруса обсягу  $N$ ,  $j$  – індекс національної мови ( $j = 1, 2, \dots, l$ , де  $l$  – загальна кількість мов в тезаурус).

Параметри «видачі» є випадковими величинами, тому оцінку їх числових значень будемо здійснювати методом статистичних випробувань. На обраному обсязі  $E_v$  текстової бази даних, сформованої повідомленнями, проводилися пошукові процедури з різними пошуковими завданнями, фіксувати випадкові значення параметрів видачі і обчислювати значення параметрів системи за формулами (2)–(4).

Відомо, що параметри повноти і точності автоматизованих діалогових інформаційно-пошукових систем є суперечливими. Повноту завжди можна підвищити до одиниці (при дуже низькій точності), повертаючи всі документи на всі запити. При цьому при збільшенні кількості знайдених документів повнота не убуває, з іншого боку, точність зазвичай знижується. Показник, що дозволяє знайти баланс між точністю і повнотою пошуку, називається  $F$ -заходом і є їх середнім гармонійним зваженням:

$$F = 1 / (\alpha * (1 / P) + (1 - \alpha) * (1 / R)) = ((\beta^2 + 1) \cdot PR) / (\beta^2 \cdot P + R), \quad (6)$$

де  $\alpha$  – вага параметра  $P$ ,

$$\beta^2 = (1 - \alpha) / \alpha, \alpha \in [0, 1], \text{ тобто } \beta^2 \in [0, \infty). \quad (7)$$

У агентну систему закладаються різні показники повноти і точності виходячи з інтересів користувача, тому  $\alpha = 1/2$ , або  $\beta = 1$ , формула спрощується.

$$F = 2 PR / (P + R). \quad (8)$$

З огляду на, що  $P$  і  $R$  є стохастичними параметрами, що визначаються формулами (2) і (3), вигляд  $F$ -заходів для системи визначається таким чином:

$$F_{\text{сист}} = [2 (M[P] \pm KP) \times (M[R] \pm KR)] / [(M[P] \pm KP) + (M[R] \pm KR)]. \quad (9)$$

Цільове значення  $F$ -міри синтезованої системи встановимо на рівні 0,7.

**4. Модель динамічної візуальної понятійно-термінологічної інформаційної системи.** Назвемо буквою ідентифікатор  $b$ , що є фрагментом потоку даних певного розміру, однозначно визначає знак в природній мові. Назвемо слово ідентифікатор  $w$ , що складається з послідовності символів, які є літерами. Назвемо символом кінця пропозиції елементи множини { ".", "?", "!" }. Назвемо (синтаксичною) пропо-

зицією  $S$  множини слів  $\{w_1, w_2, \dots, w_N\}$  синтаксично пов'язаних один з одним (що знаходиться в одній природно-мовній пропозиції, що закінчується символом кінця пропозиції). Нехай  $\epsilon$  текст - множина синтаксичних пов'язаних упорядкованих ідентифікаторів, які є словами  $w \in \{W\} \subset \{\omega\}$ , де  $w$  – слово,  $W$  – множина слів в тексті,  $\omega$  – множина допустимих слів у мові. Допустима множина слів в мові визначається словниками і є каркасом онтології. Для тексту проводиться сегментація на семантично близькі компоненти. Для кожного слова проводиться нормалізація морфологічної форми з використанням функції  $m$  морфологічного аналізу:  $m(w) = b, w \in \{W\}$ , де  $\{W\}$  – множина словоформ одного слова,  $b$  – нормальна форма слова (ім. падіж, однини для іменників, 1 л., однини, тепер. час для дієслів і т.д.).

Властивості функції морфологічного розбору:

$$m(b) = b;$$

$$\forall w (w \notin \{W\}) \Rightarrow m(w) = b, b \notin W\}.$$

Для кожного слова в нормальній формі будуються рейтинговий розподіл  $\xi_w$ , яке використовується в наступних процедурах фільтрації при підготовці до візуалізації. Для слів будуються відносини  $R1$  і  $R2$  за таким правилом:  $(w_1, w_2) \in R1$  якщо  $w_1, w_2$  – послідовні слова в реченні, тоді

$$\forall w_1 \in S \ \& \ \forall w_2 \in S \Rightarrow (w_1, w_2) \in R2.$$

Для відносин будуються рейтингові розподілу  $\xi_{R1}$  і  $\xi_{R2}$  зв'язків елементів.

Відношення  $R1$  використовується для визначення складних термінів, що складаються з декількох послідовних слів.

Якщо  $\xi_{R1}(w_1, w_2) > \epsilon_1$ , то слова  $w_1$  та  $w_2$  вважаються єдиним терміном і для нього повторюється вся обробка, при якій слова  $w_1$  та  $w_2$ , розташовані послідовно, розглядаються як одне слова  $w_1 + 2$ . Якщо  $\xi_{R1}(w_1, w_2) > \epsilon_2$ , то слова  $w_1$  і  $w_2$  вважаються семантично пов'язаними,  $\epsilon_1$  та  $\epsilon_2$  – значення порогів візуалізації. Для візуалізації будуються множина  $V2$ :

$$\xi_{R2}(w_1, w_2) > \epsilon_2 \Rightarrow (w_1, w_2) \in V2.$$

Множина  $V2$  візуалізується у вигляді графа, ребрами якого є елементи  $(w_1, w_2) \in V2$ .

Параметр  $\epsilon_2$  може динамічно змінюватися користувачем для інтерактивного вивчення семантичного оточення необхідного терміну.

Експериментально показано, що найбільш ефективно сприйняття людиною інформації відбувається при одночасній візуалізації не більше 20 об'єктів.

Для досягнення цього візуалізуються перші 20 елементів множини  $V1$ , упорядкованих по спадаючій частоті  $\xi_{R2}$  їх використання.

**5. Модель агентного пошуку та обробки науково-технічної інформації.** Основна ідея трирівневої моделі агентного пошуку та обробки інформації полягає в наступному:

1. На першому рівні слід забезпечити головним чином повноту пошуку інформації в зовнішніх джерелах, чого можна домогтися за рахунок повноти маршрутної тематичної бази даних;

2. На другому рівні забезпечується точність інформації за рахунок повноти тематичних тезаурусів системи;

3. На третьому рівні досягається максимальна пертинентність пошуку інформації відповідно до вимог конкретного користувача-фахівця за рахунок персональних словників користувача.

Модель агентного пошуку і структуризації НТІ включає в себе такі рівні:

1. Збір інформації з зовнішніх джерел мережі;
2. Фільтрація і тематична рубрикація;
3. Вторинна фільтрація по персональним словникам.

Модель агентної інформаційно-аналітичної системи представлена на рис. 1.

На першому рівні вирішується завдання збору науково-технічної інформації в повнотекстову базу даних. Основними елементами є *APF*, база даних управління агентним пошуком і повнотекстова база даних. У базі даних управління агентним пошуком містяться інформаційні джерела з різних напрямків. *APF*, керовані даною базою доставляють текстову інформацію і необхідні атрибути (автор, дата публікації тощо) за відповідним графіком сканування з інформаційних джерел в повнотекстову базу даних [11].

На даному рівні максимізувати параметр повноти ( $R_m$ ), що досягається за рахунок охоплення максимальної кількості інформаційних ресурсів за заданим тематичним напрямком. На другому рівні з повнотекстовою базою даних взаємодіють *AKF*. *AKF*, керовані тематичними тезаурусами, виробляють фільтрацію інформації за відповідними тематичними групами повідомлень. На цьому етапі *AKF* забезпечують граничну точність, відбираючи максимально пертинентну інформацію для користувачів за до-

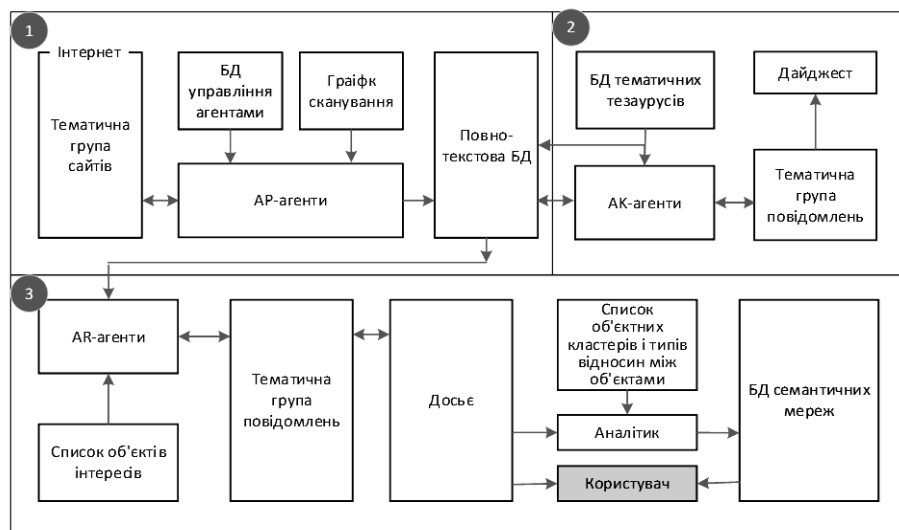


Рис. 1. Модель агентної інформаційно-аналітичної системи

помогою тематичних тезаурусів. Формуються тематичні групи повідомлень, в яких показник точності повинен бути максимізувати ( $P_{max}$ ). За підсумками угруповання інформаційних повідомлень в автоматизованому режимі формується перший тип документа - дайджест за тематичним.

На третьому рівні здійснюється персоніфікована обробка інформації. В якості керуючих даних виступає список об'єктів, що становлять професійний інтерес для користувача і персональні тезауруси кожного з фахівців. ARF на третьому рівні виконують технологічні операції, які полягають у виділенні з текстів інформації, пов'язаної з іменами об'єктів професійного інтересу, і передачі цих даних в досьє кожного з об'єктів. На третьому рівні формуються такі типи звітів для користувача:

1. Досьє на об'єкт інтересу - документ описує стан динамічного об'єкта в різні періоди часу. Під об'єктом інтересу маються на увазі технологія, проєкт, організація і т. д. Досьє на об'єкт інтересу зберігає посилання на відповідні документи - джерела інформації.

2. Семантична мережа - граф, вершинами якого є об'єкти, а дуги означають різні типи відносин між об'єктами.

Отримані семантичні мережі зберігаються у відповідній базі даних. При необхідності користувач має можливість ввести нові об'єкти в семантичну мережу, і вона буде перебудована.

Таким чином, запропонована модель відображає наступні основні принципи структуризації науково-технічної інформації, розсіяною в мережі Інтернет:

1. Інтеграція тематичної інформації з усіх джерел мережі в єдиній повнотекстової бази даних.

2. Тематична рубрикація інформації, шляхом фільтрації повнотекстової бази даних з використанням тематичних тезаурусів.

3. Формування користувацьких баз даних на основі персональних тезаурусів.

## Висновки

Розглянуто об'єктивні передумови використання агентних технологій в інформаційно-пошукових системах науково-технічної інформації. Запропоновано і реалізовано методику структурно-параметричного синтезу подібних систем. Як параметричних вимог до системи обрані й обґрунтовані чисельні значення основних характеристик пошукових систем по повноті і точності.

Запропоновано модель агентного пошуку інформації, що забезпечує досягнення високих показників повноти, тематичної точності і пертинентності інформації.

Тезауруси предметних областей пропонується у вигляді візуального інтерактивного середовища. Розроблено та обґрунтовано функції агентів, що працюють на кожному з рівнів системи, і методи формування пошукових приписів для «АР», «АК», «АР». Запропоновано метод агрегування баз даних керуючих агентним пошуком, по кожному з тематичних напрямків а також агрегування тематичних тезаурусів для пошуку інформації в прикордонних тематичних областях для скорочення обсягів видачі.

## Вдячність

Цю роботу було частково профінансовано Європейським Союзом у контексті проєкту «dComFra – Digital competence framework for Ukrainian teachers and other citizens» (Project Number: 598236-EPP-1-2018-1-LT-EPPKA2-CBHE-SP) за програмою ERASMUS+. Підтримка Європейською Комісією створення цієї роботи не означає повного схвалення її змісту, який віддзеркалює лише погляди авторів. Комісія не може нести відповідальності за будь-яке використання інформації, яку розміщено в цій роботі.

## СПИСОК ЛІТЕРАТУРИ

1. Асеева Н.Н., Ванская Г.П. Библиотечно-библиографическая классификация. М.: Либерея, 1999.
2. Jaynarayan L.H., Hender J. Control of Agent Based Systems (CoABS) & DARPA Agent Markup Language (DAML). DARPA, 2000. Presentation.
3. Будзко В.И., Леонов Д.В., Николаев В.С., Оныкий Б.Н., Соколина К.А. Развитие информационно-аналитической поддержки научно-технической деятельности в Национальном исследовательском ядерном университете «МИФИ», *Системы высокой доступности*. Т. 7, № 4. С. 4-17.
4. Маннинг Д., Рагхаван П., Шютце Х. Введение в информационный поиск. Вильямс, 2014.
5. Kent A., Berry M.M., Luehrs F.U., Perry J.W. Machine literature searching VIII. Operational criteria for designing information retrieval systems. *American documentation*. Vol. 6, No. 2. April 1955. pp. 93-101.
6. Rijsbergen C.J. Information Retrieval. London: Butterworths, 1979.
7. Artamonov A.A., Ananieva A.A., Tretyakov E.S., Kshnyakov D.O., Onyky B.N., Pronicheva L.V. A three -tier model for Structuring of Scientific and Technical Information. *Journal of Digital Information Management*. Vol. 14, № 3, June 2016. pp. 184-193.
8. Roussopoulos N.D. A semantic network model of data bases. University of Toronto, Department of Computer Science, Toronto, Dissertation TR No 104, 1976.
9. Larissa T. Moss, Shaku Atre. Are Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley Professional, 2003.
10. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. М.: Финансы и статистика, 1983.
11. Artamonov A., Ananieva A., Onyky B., Ionkina K., Galin I., Kshnyakov D. Thematic Thesauruses in Agent Technologies for Scientific and Technical Information Search. *Procedia Computer Science*. 2016. Vol. 88. pp. 493-498.

## REFERENCE

1. Aseeva, N.N. and Vanskaya, G.P. (1999), *Library and bibliographic classification*, Liberia, Moscow.

2. Jaynarayan, L.H. and Hendler, J. (2000), *Control of Agent Based Systems (CoABS) & DARPA Agent Markup Language (DAML)*, DARPA, Presentation.
3. Budzko, V.I., Leonov, D.V., Nikolaev, V.S., Onyky, B.N. and Sokolina, K.A. (2011), "The development of information and analytical support for scientific and technical activities at the National Research Nuclear University MEPhI", *High Availability Systems*, Vol. 7, No. 4, pp. 4-17.
4. Manning, D., Raghavan, P. and Schutze, H. (2014), *Introduction to the information search*, Williams.
5. Kent, A., Berry, M.M., Luehrs, F.U. and Perry, J.W. (1955), "Machine literature searching VIII. Operational criteria for design-information information retrieval systems", *American documentation*, Vol. 6, No. 2, pp. 93-101.
6. Rijsbergen, C.J. (1979), *Information Retrieval*, Butterworths, London.
7. Artamonov, A.A., Ananieva, A.A., Tretyakov, E.S., Kshnyakov, D.O., Onyky, B.N. and Pronicheva L.V. (2016), "A three-tier model for Structuring of Scientific and Technical Information", *Journal of Digital Information Management*, Vol. 14, No. 3, pp. 184-193.
8. Roussopoulos, N.D. (1976), *A semantic network model of data bases*, University of Toronto, Department of Computer Science, Toronto, Dissertation TR No 104.
9. Larissa T, Moss and Shaku, Atre (2003), *Are Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*, Addison-Wesley Professional.
10. Ferster, E. and Renz, B. (1983), *Methods of correlation and regression analysis*, Finance and statistics, Moscow.
11. Artamonov, A., Ananieva, A., Onyky, B., Ionkina, K., Galin, I. and Kshnyakov D. (2016), "Thematic Thesauruses in Agent Technologies for Scientific and Technical Information Search", *Procedia Computer Science*, Vol. 88, pp. 493-498.

Received (Надійшла) 24.03.2020

Accepted for publication (Прийнята до друку) 20.05.2020

#### ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

**Касілов Олег Вікторович** – кандидат технічних наук, доцент, професор кафедри систем інформації; Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;

**Oleg Kasilov** – PhD, Associate Professor, Professor of the Information Systems Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

e-mail: [okasilov@hotmail.com](mailto:okasilov@hotmail.com); ORCID ID: <http://orcid.org/0000-0002-8524-2345>

**Крамська Катерина Ігорівна** – аспірант кафедри «Інформаційних технологій і систем колісних та гусеничних машин ім. О.О. Морозова»; Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;

**Kateryna Kramska** – graduate student of the department "Information Technologies and Systems of Wheel and Track Machines", National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

e-mail: [katerina.kramskaya@ukr.net](mailto:katerina.kramskaya@ukr.net); ORCID ID: <http://orcid.org/0000-0002-9694-0552>

#### Модели и метод синтеза агентной информационно-поисковой системы

О. В. Касилов, Е. И. Крамская

**Аннотация.** Существующие современные информационно-поисковые системы, такие как Google, Яндекс, Yahoo ориентированы на среднестатистического пользователя, для которого предметом поиска является неспецифическая, научно-технологическая информация. Поэтому эти поисковые системы не могут в полном объеме предоставить нужную информацию для профессионального пользователя. Вследствие того, что данная информация является справочной или энциклопедической, она не соответствует заданному поиску. Обеспечить качественный поиск информации и решить проблему информационно-аналитического обеспечения научной и технологической деятельности возможно посредством использования агентных технологий. Агентные технологии иницируются и управляются автоматически, а также функционируют без участия пользователя. Поэтому они имеют более высокую производительность и возможность непрерывной работы. Разработана модель многоуровневой агентной информационно-аналитической системы в естественно-научных и технологических направлениях. Она является наиболее целесообразно для тематического обслуживания коллективных и персональных пользователей. Также в статье рассмотрено использование агентных технологий в информационно-поисковых системах научно-технической информации. Предложена методика структурно-параметрического синтеза подобных систем.

**Ключевые слова:** информационно-поисковые системы; агентные технологии; агентные системы; агентный поиск.

#### Models and synthesis method of agent information retrieval system

O. Kasilov, K. Kramska

**Annotation.** Existing modern information retrieval systems such as Google, Yandex, Yahoo are aimed at the average user, whose subject of the search is not specific, scientific and technological information. Therefore, the aforesaid search systems cannot fully provide the necessary information for the professional user. Due to the fact that this information is reference or encyclopedic which does not correspond to the specified search. It is possible to ensure a high-quality information search and to solve the problem of information and analytical support for scientific and technological activities with the using of agent technologies. Agent technologies are initiated and controlled automatically, and also function without user's participation, therefore they have higher productivity and the possibility of continuous operation. This article has developed the architecture of a multi-level agent information-analytical system in natural-scientific and technological areas. The authors proposed software and hardware implementations of a thematic service model for collective and personal users. Also, the article discusses the use of agent technologies in information retrieval systems for scientific and technical information. The technique of structural-parametric synthesis of such systems is proposed as well.

**Keywords:** information retrieval systems; agent technologies; agent systems; agent search.