

МЕТОД ПОБУДОВИ ІНТЕЛЕКТУАЛЬНИХ АГЕНТІВ АНАЛІЗУ ТА СТРУКТУРИЗАЦІЇ ДЖЕРЕЛ ПРОСТОРУ ДАНИХ НАУКОВО-ТЕХНІЧНОЇ ІНФОРМАЦІЇ

© Литвин В., Шаховська Н., 2008

Подано визначення наукового проекту та описано методи інтелектуального аналізу та структуризації наукових даних.

Determination of scientific project is given and the methods of intellectual analysis and structured of scientific information are described.

Вступ

1. Постановка проблеми

Університети і дослідницькі інститути у всьому світі активно планують і реалізують архіви своєї наукової продукції. Крім того, використання веб-серверу наукових установ надає нові можливості для своєчасного поширення наукової інформації. Задача правильного вибору програмного забезпечення, яке б максимальне повно задовольнило потреби навчальних і наукових організацій, при мінімумі прямих і непрямих витрат є актуальною задачею сьогодення [1].

У загальній класифікації проектів науковий проект виділяється насамперед за характером цільового завдання. За своїм змістом науковий проект є одним з основних форм організації наукової діяльності, яка визначається як діяльність, скерована на отримання і використання нового знання. Отже, мета реалізації наукових проектів має двоєдиний характер і визначає основний зміст взаємозв'язаних процесів отримання і використання нових знань.

Процес отримання нових знань реалізується у формі фундаментальних і прикладних наукових досліджень, а процес їх використання реалізується у формі експериментальних розробок.

Функціонування науки, техніки і виробництва як єдиної системи забезпечується здійсненням науково-технічної діяльності, скерованої на отримання, використання нових знань для вирішення технологічних, інженерних, економічних, соціальних, гуманітарних та інших проблем.

2. Аналіз останніх досліджень та публікацій

Сьогодні не існує єдиного тлумачення поняття "Науковий проект". Відповідно до основних положень теорії управління проектами і підставами методології, **науковий проект** можна визначити як обмежений в часі цілеспрямований процес отримання, теоретичної систематизації і використання нового наукового знання зі встановленими вимогами до якості результатів, витрат ресурсів і специфічною організацією.

Загальна характеристика наукового проекту може бути представлена в наступних основних аспектах його розгляду:

- науковий проект як діяльність;
- науковий проект у межах повного життєвого циклу систем і об'єктів (ПЖЦСЮ);
- науковий проект в структурі інноваційного процесу;
- науковий проект у межах теорії управління проектами;
- науковий проект у межах теорії організаційних систем.

З позицій системного аналізу, діяльність розглядається як складна система, скерована на підготовку, обґрунтування і реалізацію вирішення проблем політичного, соціального, економічного, технічного та ін. характеру.

Аналіз даних наукових ресурсів є дуже важливою та актуальною задачею з огляду таких особливостей цієї предметної області:

- пошук необхідної інформації в Інтернет саме наукового спрямування не завжди приносить хороші результати, оскільки більшість джерел Інтернет – науково-популярні, а якісні джерела – зазвичай платні;
- одні і ті самі дані певного наукового експерименту можуть застосовуватись для різних задач, що значно пришвидшить для науковця пошук вирішення проблеми;
- оскільки кожен із розробників має свій погляд на створення системи обліку та аналізу наукових даних, то без розроблення спеціальних алгоритмів, які використовують міжнародні стандарти найменування об'єктів кожної з предметних областей, пошук необхідної інформації у локальних мережах навчально-наукових закладів стає практично неможливим;
- впорядкування та аналіз даних навчально-наукових організацій (університетів, інститутів) дасть можливість не тільки виявити перспективні напрямки науки (у яких публікується достатня кількість статей), але й визначити перспективних вчених з певної галузі.

3. Формулювання цілі статті

Мета роботи – розроблення інтелектуальних агентів аналізу джерел даних простору даних наукових досліджень, автоматизації розпізнавання їх змісту, структурування науково-технічної інформації, розроблення методів визначення змісту джерел даних та зведення їх до однорідної структури з метою подальшого аналізу їх змісту.

4. Виклад основного матеріалу

Науковий проект представляється як завершений цикл продуктивної діяльності окремої людини, колективу, організації, підприємства або спільної діяльності багатьох організацій і підприємств.

Повний життєвий цикл систем і об'єктів передбачає процес задуму, втілення і подальшого використання об'єктів – машин, механізмів, пристроїв, матеріалів, речовин, систем, комплексів і технологій аж до зняття їх з експлуатації (див. рис. 1).

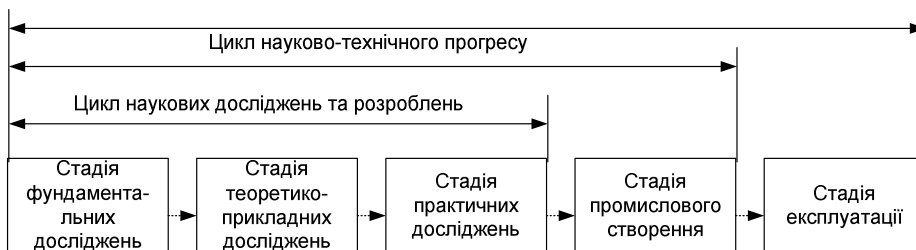


Рис. 1. Повний життєвий цикл систем і об'єктів

Під циклом досліджень і розробок (ДР) розуміється повна і впорядкована за ступенем завершеності сукупність результатів, починати з виникнення наукової ідеї до отримання її повного інформаційного подання, необхідного і достатнього для її конкретного втілення в новий продукт, аж до його передачі для промислового освоєння. Під стадією розуміється найузагальненіший характерний стан ПЖЦ, на досягнення якого орієнтовані дослідження.

Розгляд наукового проекту в структурі ПЖЦ обумовлений двоєдиним характером його мети – отримання і застосування нових знань, і наявністю трьох основних аспектів його розгляду (див. рис. 2) [1].

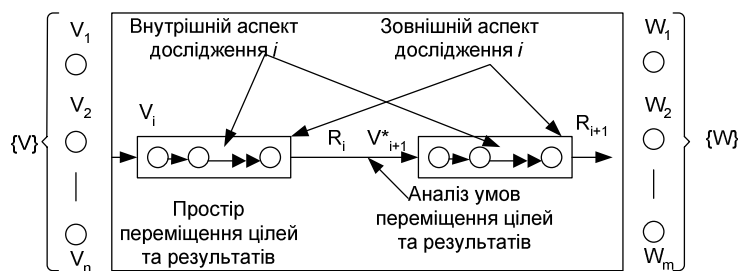


Рис. 2. Двоєдиний характер мети наукового проекту (W – множина практичних цілей, V – множина наукових цілей, V_{i+1}^* – зовнішня ціль, для досягнення якої необхідно результат R_i , R_i – результат i -го дослідження)

Відслідковувати наукові проекти, їх стан ПЖЦ, появу нових знань буде здійснювати інтелектуальний агент. Функціонування цього агента буде ґрунтуватися на онтологіях предметних областей (ПО) відповідних наукових досліджень. Оскільки наукові дослідження здійснюють наукові та навчальні заклади, то онтології, що будуть нами використовуватись, називаються організаційними. Тобто організаційна онтологія – це онтологія, що відображає знання про організаційну та функціональну структуру певного суб'єкта економічної діяльності, тобто його основні компоненти та зв'язки між ними. Вона містить інформацію про працівників підприємства, ієрархію виробничих відносин між ними; ресурси, що використовуються на підприємстві у процесі виробництва; продукцію, створення якої є наслідком функціонування підприємства, та структурні одиниці підприємства й зв'язки між ними.

Простір даних – це множина даних, поданих у різних моделях (баз даних, сховищ даних, статичних веб-сторінок, неструктурованих даних, графічних та мультимедійних даних), локальних сховищ та індексів, а також засобів інтеграції, пошуку та опрацювання інформації, об'єднаних середовищем управління моделями. Робота інтелектуального агента у просторі даних передусе застосуванню засобів опрацювання даних, оскільки структура джерел наперед невідома.

Простір даних обліку та супроводу наукових досліджень об'єднує інформацію про [7]:

- інженерні та технічні розрахунки експериментів чи досліджень, подані у вигляді таблиць Excel, текстових файлів з розділювачами тощо;
- наукові статті та тези, описані у внутрішніх базах даних кафедр чи наукових підрозділів;
- програми, що використовуються для ведення розрахунків та експериментів;
- наукові звіти кафедр та наукових підрозділів з роботи працівників;
- літературні джерела (бібліотечні фонди), подані у вигляді внутрішніх баз даних чи загальноуніверситетського сховища даних, url-посилання, мультимедійна та графічна інформація, збережена на файл-сервері університету чи подана у глобальній мережі.

Для опису простору даних уведемо поняття організації.

Організація – це стійка система відношень між суб'єктами, заснована на сукупності досягнутих ними угод. Подання організації як системи відношень є однією з можливих моделей організації. У контексті організації між індивідами формується багато типів відношень – від виробничих до соціально-психологічних. Використовуються три види заснованих на онтологіях моделі організацій, що структурують та організують інформацію [1]: 1) організаційна онтологія; 2) онтологія ПО діяльності організації; 3) онтологія користувацької діяльності. Організаційна онтологія забезпечує семантичну інформацію про структуру організації. Ця складна структура часто використовує декомпозицію в окремі ієрархічні модулі.

Онтологія ПО діяльності організації проектується для того, щоб організувати й структурувати бізнес-функції і дії, що мають місце у певній ПО. Така доменна онтологія діяльності забезпечує ієрархічну структуру для класифікації записів, що документують бізнес-функції і дії для класифікації й індексації цілей.

Онтологія користувацької діяльності пов'язана з діями для пошуку інформації, здійснюваними кінцевими користувачами. З погляду аналізу, що ґрунтується на знаннях, потрібно визначити, що саме потрібно знати різним типам користувачів і про інформаційні об'єкти і дії, що використовуються у задачі пошуку інформації, і як ці знання мають бути організовані. Дії користувачів зазвичай можуть бути описані через дії, що ними виконуються, та інформаційні об'єкти, зв'язані з інформаційними потребами. Розвиток онтології користувацької діяльності починається з формування таксономічної класифікації знань про задачу й інформаційні об'єкти. Знання про задачу містять словник для подання процесу виконання дій, наприклад, пошук, перегляд, збереження. Такі онтології можуть використовуватися:

- для виконання складних інформаційних запитів, пов'язаних з опрацюванням контенту на семантичному рівні (наприклад, знайти всі проекти, які за певний період часу виконувалися працівниками, що контактували зі співробітниками певного підрозділу та мали певний рівень повноважень, тобто мали доступ до відповідних корпоративних знань);
- для формування групи експертів, що мають достатні знання та досвід роботи, щоб виконувати експертизу в області, описаній у вигляді повнотекстового документа;
- для пошуку співробітника, який відповідає за певне коло питань (незалежно від того, як у цій організації називається його посада та як сформульовані його посадові обов'язки);

- для встановлення відповідності між інформаційними потребами користувача та інформаційними ресурсами, поданими у ІС.

Організаційна онтологія наукової організації містить класи та підкласи, що характеризуються певними атрибутами (рис. 3).

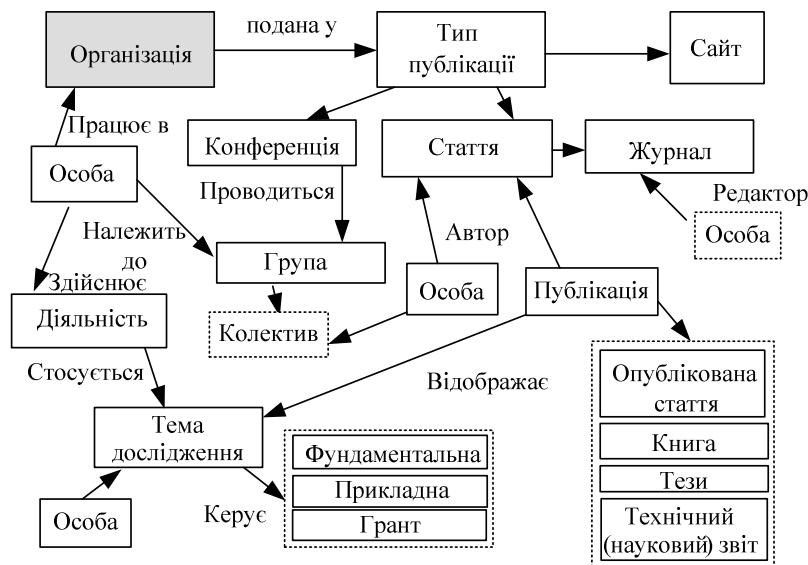


Рис. 3. Онтологія наукової організації

Одним із засобів моделювання ПО є створення її тезауруса. Для цієї задачі тезаурус можна розглядати як окремий випадок онтології. Тезаурус – це $Th = \langle T, R \rangle$, де T – множина термінів, а R – множина відносин між цими термінами. Множини T і R скінченні. Термін – це слово або словесний комплекс, який співвідноситься з поняттям певної організованої області знань (науки, техніки), що вступає в системні відношення з іншими словами і словесними комплексами й утворює разом з ними в будь-якому окремому випадку й у певний час замкнену систему, що відрізняється високою інформативністю, однозначністю, точністю й експресивною нейтральністю. Тезаурус – структура лінійно пов'язаного подання слів та їхніх значень, призначена для зіставлення концептуальних визначень у контексті слова. Множина термінів тезауруса відповідає множині концептів онтології O .

Приклад тезауруса області наукових досліджень поданий на рис. 4.

- людина** (ім'я (STRING), по батькові (STRING), прізвище (STRING), рік народження (INTEGER));
- співробітник** (... , посада (STRING), працює в (підрозділ), ідентифікаційний код (INTEGER));
- науковий співробітник** (... , науковий ступінь (STRING), працює за темою (тема), науковий стаж (DATE), публікації (публікація));
- інженер** (... , має кваліфікацію (STRING));
- аспірант** (... , рік вступу (DATE), науковий керівник (науковий співробітник), публікації (публікація));
- підрозділ** (назва (STRING), керівник (співробітник));
- інститут** (... , адреса (STRING));
- відділ** (... , відноситься до (інститут));
- лабораторія** (... , відноситься до (відділ));
- тема** (шифр (STRING), назва (STRING), керівник (науковий співробітник), дата початку (DATE), дата закінчення (DATE), виконавці (співробітник));
- комплексна тема** (... , складається з (тема));
- публікація** (назва (STRING), автори (людина), рік публікації (STRING), мова (STRING), кількість сторінок (INTEGER));
- наукова стаття** (... , УДК (STRING), анотація (STRING), назва видання (STRING));
- монографія** (... , рецензент (науковий співробітник), назва видавництва (STRING));
- тези конференції** (назва конференції (STRING), дата ПО ведення (DATE), місце ПО ведення (STRING)).

Рис. 4. Тезаурус онтології наукових досліджень

Структура тезауруса визначена стандартами ANSI Z39.19, ISO 2788-1986, ISO 5964-1985, ГОСТ 7.25-2001, ГОСТ 7.24-90. Для врахування ефектів, пов'язаних з розходженнями суб'єктивних знань приймача і передавача в комунікаційних процесах, що є наслідками різних обсягів знань у ПО, використовують тезаурусну модель, що зв'язує семантичні властивості інформації зі здатністю користувача сприймати інформацію.

Модель інформаційних ресурсів простору даних наукових досліджень подана на рис. 5.

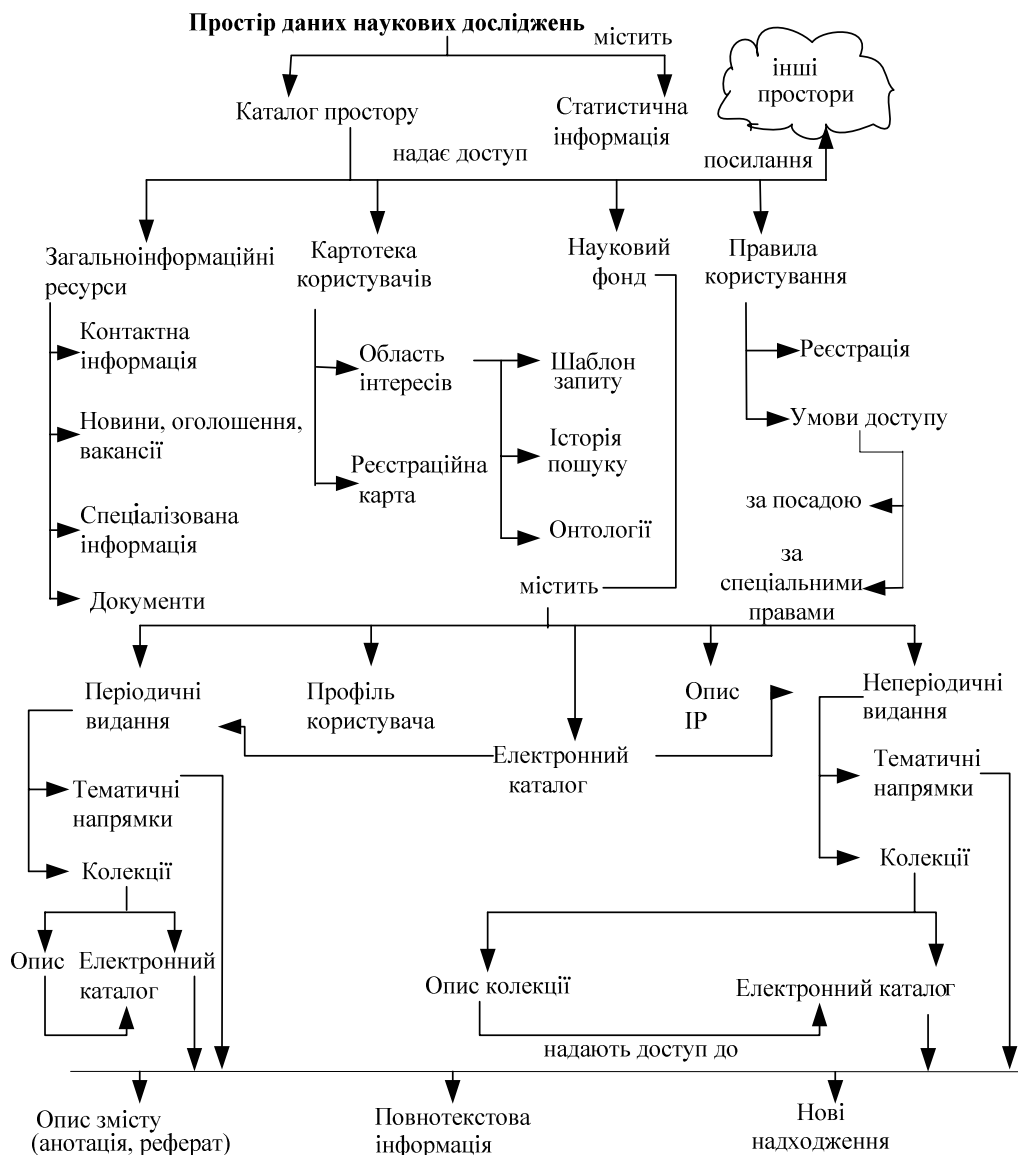


Рис. 5. Модель інформаційних ресурсів простору даних наукових досліджень

Науковий фонд містить періодичні видання та книжковий фонд і є сукупністю електронних колекцій наукової спрямованості.

Електронна колекція – це систематизована сукупність IP, які об'єднані стосовно певного критерію. Критерієм може бути будь-яка властивість електронного IP або їх ієрархічна структура. Областю визначення критерію часто розглядається набір елементів опису IP DUBLIN Core, який останнім часом набуває статусу стандарту, розширений певними елементами: наприклад, користувач, власник, призначення, цінність, стандартні класифікатори та рубрикатори ББК, УДК, класифікатор ВАК.

Опис колекції складається з її історії (найважливіших дат), переліку вмісту або електронного каталогу, посилань на засоби пошуку в колекції та поповнення й редагування колекції.

Вміст *електронного каталогу* електронної колекції збігається зі структурою критерію, за яким було упорядковано колекцію. Вміст електронної колекції є сукупність окремих джерел, що є *описом змісту* (анотація, реферат тощо) та *повнотекстовою інформацією*. Ми повинні розробити Простір даних наукових досліджень, який містив би інформацію ПО джерела результатів наукових досліджень та здійснював автоматичне розпізнавання структури джерела.

До основних задач, що ставимо перед собою, належать:

- доступ до електронних версій всіх журналів, на друковані версії яких користувачі мають права (передплату);
- доступ на перегляд передплачених журналів за алфавітом, видавництвом, темою;
- доступ до електронних версій статей за наявності передплати раніше, ніж статті з'являються у друкованому вигляді;
- вільний доступ до архіву за наявності передплати;
- аутентифікація користувача;
- пошук ресурсу:
 - пошук за ключовими словами, автором чи назвою, а також словами з анотації;
 - швидкий пошук за ключовими словами, автором, редактором, серією, словами в назві, анотації, повному тексті;
 - за фіксованим, обмеженим переліком критеріїв, наприклад, за назвою, видавцем, а також розширений пошук за прізвищем, національністю, датами народження та смерті автора;
 - складний пошук за багатьма критеріями, пов'язаними логічними зв'язками;
 - розширений пошук за багатьма критеріями з можливістю обмеження області пошуку;
 - швидкий та розширений пошук за автором, назвою, темою, серією, датою публікації, мовою, а також областю пошуку (колекція), типом ресурсу (журнал, довідник, база даних, книжка), резюме, весь текст, всі поля;
 - розширений пошук, що надає можливість задання фільтрів, формату виведення, шаблонів пошуку, обмеження області пошуку (вся база даних, в межах однієї теми або їх множини, булевського пошуку, пошуку за словами з помилками);
 - з використанням пошукового запиту, в який користувач заносить значення параметрів.

Основним ресурсом, який забезпечує ефективність наукових досліджень, є актуальна науково-технічна інформація. Її опрацювання потребує відповідного інструментарію. У складі автоматизованого робочого місця (АРМ) наукового працівника є його функціональне ядро – так званий інтелектуальний агент, здатний в процесі самонавчання адаптуватися до конкретних інформаційних потреб користувача та виявляти, зберігати і використовувати релевантні до відповідних задач знання. Інтелектуальний агент може бути реалізований у межах продукційної, об'єктної, нейромережевої парадигми чи їх певного поєднання. Вибір розробника відштовхується від конкретних експлуатаційних вимог. При цьому доцільно максимально відкласти рішення щодо програмно-апаратної реалізації, пов'язані з обмеженнями у виборі операційної системи чи комп'ютерної архітектури. У будь-якому разі основою інтелектуального агента є його база знань, а ядром бази знань є онтологія (див. рис. 6). **Загальні** властивості інтелектуального агента визначаються його онтологією, яка задає спосіб представлення знань, механізми міркувань та прийняття рішень.

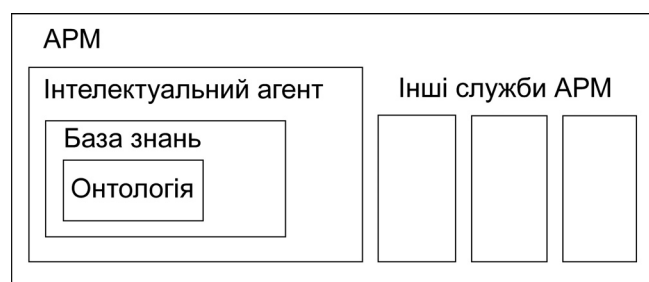


Рис. 6. Місце інтелектуального агента у загальній структурі АРМ

Особливості функціонування спеціалізованого інтелектуального агента визначаються його інтересом – вектором оцінок бажаності можливих станів агента.

Для опису **інтересу агента**, за допомогою якого він розрізняє стани довколишнього світу та позиціонує себе у ньому, застосовується функція корисності, котра є числовою оцінкою його бажаності для агента. Корисності об'єднуються з імовірностями дій для визначення **очікуваної корисності** кожної дії.

Нехай $U(S)$ – корисність стану S з погляду агента, що приймає рішення щодо вчинення деякої дії A . Довільна недетермінована дія може спричинити результуючий стан $Result_i(A)$, де індекс i пробігає по усіх можливих результатах. Перш ніж вчинити дію A , агент оцінює імовірність $P(Result_i(A) | Do(A), E)$ кожного з можливих результатів, де E – сукупність доступних агенту параметрів його стану, а $Do(A)$ – висловлювання, згідно з яким в поточному стані виконується дія A . Отже, можна обчислити умовну корисність дії $EU(A|E)$ з врахуванням відомих параметрів стану:

$$EU(A | E) = \sum P(Result_i(A) | Do(A), E) \cdot U(Result_i(A)).$$

Якщо раціональний інтелектуальний агент керується принципом **максимальної очікуваної корисності** (Maximum Expected Utility – MEU), він змушений вибрати дію, яка максимізує очікувану корисність для агента. Так функціонує механізм мотивації поведінки раціонального інтелектуального агента незалежно від сфери його застосування.

У випадку інформаційно-пошукового агента його інтерес може бути заданий через оцінку новизни отриманих повідомлень, яка потребує застосування методів інтелектуального аналізу природомовних текстів, описаних нижче.

Звіт про роботу агента записується у базу знань, схема відношень якої подана на рис. 7.

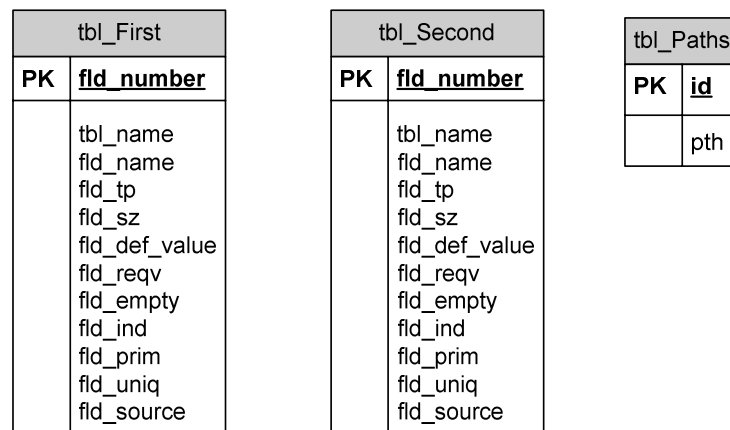


Рис. 7. Схема відношень бази знань інтелектуального агента

Відношення tbl_First та tbl_Second містять, відповідно, інформацію про структуру першого джерела даних та другого джерела. Tbl_Path зберігає шляхи або сторінки усіх джерел простору даних наукових досліджень.

Принцип роботи агента полягає у попарному переборі усіх джерел, завантаженні інформації про них у відношення бази знань та їх порівнянні. Результат порівняння двох джерел (реляційних баз даних для прикладу) подано на рис. 8. Обираються два джерела, для яких ступінь збігу відношень та полів буде максимальним.

Система інтелектуального аналізу природомовних текстів

Особливістю АРМ наукового працівника є наявність у наборі його засобів системи інтелектуального аналізу природомовних текстів (ІАПТ), яка забезпечує користувача можливістю автоматичного виділення з усього доступного масиву електронних текстових документів деякої підмножини документів, найбільш релевантних до цієї предметної області. Зокрема, актуальною

практичною проблемою залишається виділення корисної електронної кореспонденції з загального потоку електронної пошти, перевантаженого поштовими повідомленнями рекламного характеру – так званим спамом.

	A	B	C	D
1	ТАБЛИЦІ, яких нема в одній з баз:			
2	rf_state	нема у першій базі		
3				
4	ПОЛЯ, яких нема в одній з баз:			
5	ob_disc	crtdate	відсутнє у першій базі	
6	ob_disc	crtdate	відсутнє у першій базі	
7	od_person	crtdate	відсутнє у першій базі	
8	od_person	crtdate	відсутнє у першій базі	
9	od_person	note	відсутнє у першій базі	
10	rf_admunitype	crtdate	відсутнє у першій базі	
11	rf_admunitype	crtdate	відсутнє у першій базі	
12				

	A	B	C	D	E	F	G	H	I	J	K	L
1	Таблиця	Поле	База	Властивості								
2	rf_part			Тип поля	Розмір поля	За замовчуванн	Обовязкове	не порожне	індекс	первинний ключ	унікальне	джерело стрічок
3		minvalue	1	Текстовий	50					ИСТИНА		
4			2	Довге ціле	0					ЛОЖЬ		
5		step	1	Текстовий	50					ИСТИНА		
6			2	Довге ціле	0	1				ЛОЖЬ		

Рис. 8. Результат роботи інтелектуального агента

Тенденції розвитку інформаційних засобів вказують на необхідність „інтелектуалізації” методів обробки даних, яка полягає у застосуванні алгоритмів штучного інтелекту для автоматизації роботи, пов’язаної з пошуком документів, їх класифікацією, рангуванням за заданими користувачем критеріями, кластеризацією за змістом. Отже, зміст природомовного тексту з цією метою необхідно належним чином розпізнати.

Розпізнавання змісту тексту інтелектуальним агентом

Читачем природомовного тексту є інтелектуальний агент. Тому сам текст, який своєю чергою, створений інтелектуальним агентом, побудований як повідомлення. Структура повідомлення орієнтована на сприйняття іншим агентом, тому складається з двох частин (рис. 9): констатуючої частини, за якою адресат оцінює релевантність повідомлення (1) та визначає його контекст (2), та конструктивної частини – потенційно нових для читача знань у даному контексті (3).

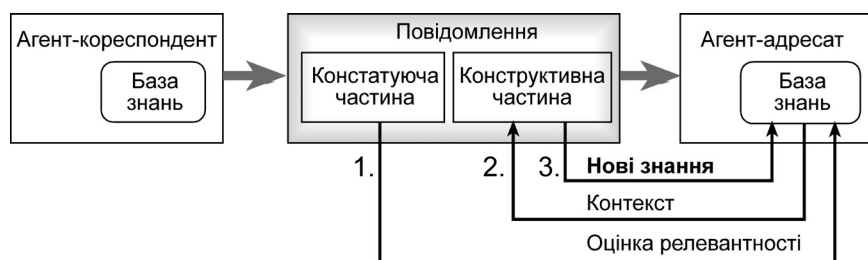


Рис. 9. Розпізнавання нових знань у повідомленні з метою наповнення ними бази знань

Кожне відоме інтелектуальному агенту знання характеризується власною функцією корисності. Розпізнане нове знання має дві частини: загальну (спільну, подібну) частину з прототипом знання, упізнаним серед відомих агенту апріорних знань та частину відмінностей, в якій функція корисності обчислюється або вказується явно (рис. 10).

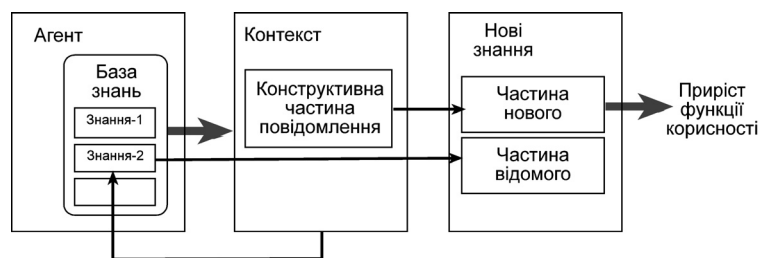


Рис. 10. Обчислення функції корисності для нових знань як різниця між функцією корисності знання-прототипу і функції корисності, оціненої з контексту пропонуваного в повідомленні нового алгоритму

Отже, новизна знань оцінюється, враховуючи приріст максимальної очікуваної корисності інтелектуальної системи загалом.

Якщо нове знання є не повним алгоритмом, а лише окремим фактом чи правилом, що все ж вносять уточнення до вже відомих агенту алгоритмів, зміна їх функції корисності є оцінкою новизни цього факту (правила) та їх важливості для агента.

Представлення знань у формі семантичної мережі передбачає, що будь-яке можливе узагальнення, тобто комплексне, складене поняття завжди явним чином артикульоване, назване і як окремий концепт фігурує в базі знань. Отже, функція корисності присутня у БЗ інтелектуального агента як властивість його конкретних та узагальнених алгоритмів розв'язання задач, які стоять перед ним. Узагальнення відповідають контейнерним класам відповідно організованих алгоритмів з результирующим вигрaшем, обчисленим у розмірності ресурсів. Для кожного складового суб-алгоритму рекурсивно оцінюється його власна частина загального вигрaшу.

5. Висновки

Тобто можна зробити такий загальний висновок:

1. Для підвищення ефективності наукових досліджень необхідно вести оперативний облік затрат та забезпечити об'єктивний аналіз і оцінку результатів досліджень.
2. Сучасний розвиток інформаційних технологій створює необхідні передумови впровадження програмно-технічних систем обліку та аналізу ресурсних затрат і отриманих результатів.
3. Суб'єктивний фактор відіграє на етапі впровадження сучасних систем підтримки прийняття рішень надзвичайно важливе значення.
4. Ресурси, що підлягають обліку, мають різноманітну природу, традиційно по-різному обліковуються і нерідко їх облік розпоршений між різними службами наукової установи.
5. Надзвичайно динамічна трансформація інформаційного суспільства потребує гнучких методів і засобів обліку, контролю та аналізу ресурсів наукової установи: фінансових, матеріальних, людських та інтелектуальних, з метою їх найбільш ефективного використання.
6. Таким засобом може слугувати автоматизоване робоче місце наукового співробітника.

1. *Методологические вопросы управления научными проектами*, – 10.2006, – <http://www.quality.edu.ru/quality/met/493/>. 2. Даревич Р.Р., Досин Д.Г., Литвин В.В., Мельничок Л.С. *Застосування інформаційних технологій для координації наукових досліджень* – Львів: Сполом, 2008. – 240 с. 3. Даревич Р.Р., Досин Д.Г., Литвин В.В. *Метод автоматичного визначення інформаційної ваги понять в онтології бази знань // Відбір та обробка інформації*. – 2005. – Вип. 22(98). – С.105–111. 4. Dosyn D.G., Darevych R.R., Lytvyn V.V. *Modelling of the intelligent text recognition agents based on dynamic ontology*. // *Тези доп. IV міжнар. конф. „Інтернет – Освіта – Наука – 2004”*, Збірник матеріалів конференції. – Вінниця: УНІВЕРСУМ – Вінниця, 2004. – Т. 2. – С.577–579. 5. Dosyn D., Darevych R., Lytvyn V., Dalyk U. *New knowledge evaluation using message model of NLT document // Proceedings of the international conference on computer science and information technologies CSIT-2006*. – Lviv, Ukraine. – P. 118–119. 6. Даревич Р.Р., Досин Д.Г., Литвин В.В., Назарчук З.Т. *Оцінка подібності текстових документів на основі визначення інформаційної ваги елементів бази знань // Штучний інтелект*. – № 3. – 2006. – С.500–509. 7. Шаховська Н.Б. *Простір даних області наукових досліджень // Моделювання та інформаційні технології*. – 2008. – № 45. – С.132–140.