

Використання онтологій для підвищення якості пошуку інформації для поповнення баз знань інтелектуальних систем

Вороной Олексій

Кафедра програмного забезпечення автоматизованих систем, Державний університет інформатики і штучного інтелекту, УКРАЇНА, м.Донецьк, пр.Б.Хмельницького, 64, E-mail: vrnalexey@gmail.com

This work shows the concept of building multiagent system for search of knowledge in distributed unstructured sources of information. Contents, main functions of agents communication during solving task of creating knowledge in intellectual system is proposed. To improve quality of knowledge search requests to search systems are created according to known subject.

Ключові слова – multiagent system, distributed sources, intellectual system.

I. I. Вступ

Системи, засновані на знаннях, активно застосовуються в наукових дослідженнях, бізнесі, навчанні. Виникнення мережевих технологій та Інтернет сприяє використанню цих ресурсів в якості джерел отримання знань. Однак, крім позитивних сторін, процес одержання знань супроводжується низкою проблем, пов'язаних зі збільшенням обсягу надлишкової інформації внаслідок багаторазового дублювання, слабкої структурованості інформації, представленням її природною мовою. Покликана вирішити ці проблеми концепція семантичного Web в теперішній час знаходиться на стадії розвитку і основна маса інформаційних ресурсів, накопичених у Web першого покоління, продовжує залишатися у вигляді, який не пристосований для пошуку знань. У зв'язку з цим до процесу вилучення знань пред'являється ряд вимог: висока швидкість обробки великих обсягів даних; гнучкість щодо постійно мінливого оточення; адаптованість до типів інформаційних ресурсів та їх змісту; робота з неструктурованими природно-мовними джерелами.

II. II. Технологія багатоагентних систем

Для вирішення проблем вилучення знань з розподілених джерел інформації в даний час застосовуються технології багатоагентних систем (БАС) [1, 2]. У порівнянні з іншими засобами організації розподілених обчислень вони мають низку переваг: зменшення навантаження на мережу; автономне та асинхронне виконання обчислень; адаптація до умов виконання і т.і. Під багатоагентною системою розуміється безліч розподілених по мережі інтелектуальних агентів, що пересуваються в пошуках доречних даних і знань, а також кооперуються в процесі вироблення рішень [3]. Під агентом розуміється система, яка містить як програмні, так і апаратні компоненти, спрямована на досягнення певної мети і здатна до взаємодії із середовищем та іншими агентами. До основних властивостей агентів відносяться: здатність до навчання, автономність, можливість взаємодіяти з іншими агентами різними способами, здатність до міркування (наявність механізму виводу), комунікативність, мобільність.

В роботі розглядається вирішення завдання пошуку нових знань з неструктурованих джерел для інтелектуальних систем (ІС) з базами знань. Для цього пропонується концепція побудови засобів поповнення баз знань предметних областей (ПО) ІС яка заснована на використанні колективів агентів.

В агентно-орієнтованій інтелектуальній системі (АОІС) процеси отримання та оновлення знань забезпечуються двома групами агентів – багатоагентними системами: пошуку інформації, що вміщує знання, в розподілених віддалених джерелах – БАС ПЗ та вилучення і редагування знань з колекцій документів з предметних областей – БАС ВРЗ.

1. Розглянемо склад та завдання, що вирішують інтелектуальні агенти БАС ПЗ та БАС ВРЗ.

2. До складу БАС ПЗ входять: агент - менеджер пошукових агентів (1), агент пошукових систем (2), агенти перевірки існування документів (3), агент ранжування документів (4), агенти отримання документів.

3. Граф взаємодії агентів БАС ПЗ при обробці запитів пошуку документів для поповнення знань в ПО з використанням трьох пошукових систем приведений на рис.1. Вершина з номером 6 відображає агента – менеджера колекції документів (КД) ПО.

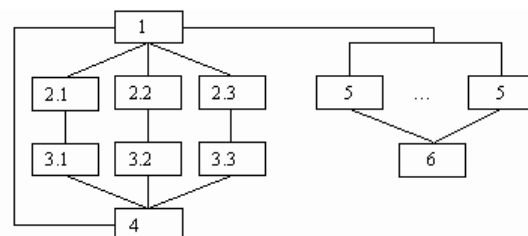


Рис. 1. Граф взаємодії агентів БАС ПЗ.

Агент 1 БАС ПЗ отримує запити на пошук документів за темами ПО та на їх підставі видає завдання агентам пошукових систем в вигляді запитів на пошук. До складу БАС ПЗ входять агенти пошуку, які використовують пошукові системи в тематичних колекціях документів за предметними областями, багатоцільові пошукові системи та бази даних «прихованого» Web. Агенти перевірки існування документів встановлюють актуальність посилань, виданих пошуковими агентами. Агенти ранжування документів виключають посилання що повторюються та впорядковують за ступенем релевантності посилання на всі документи, що знайдені пошуковими системами. Ступінь релевантності обчислюється на основі анотацій, які видаються пошуковими системами. Менеджер пошу-

кових агентів на основі ранжированого списку послань формує завдання агентам отримання документів. Агенти отримання документів взаємодіють з агентом – менеджером КД ПО.

БАС ВРЗ включає агентів експертів ПО, агентів роботи з локальними знаннями, які реалізують процедури наповнення та коригування баз знань в процесі діалогу з агентами експертів, агентів роботи з віддаленими знаннями, що реалізують алгоритми автоматичного вилучення знань з колекцій документів з ПО.

У режимі обробки локальних знань користувачу – експерту представляється зручний графічний інтерфейс для введення знань в базу в формі їх представлення, засоби редагування знань і роботи з ними.

У режимі обробки віддалених знань може виконуватись оновлення або коригування вже існуючої бази знань. При цьому експерт може сформулювати запит на пошук нових знань у віддалених інформаційних ресурсах, аналізувати інформацію здобуту агентами БАС ПЗ за раніше підготовленими запитом та на її підставі виконувати оновлення або коригування бази знань.

Неструктурованість інформації в мережевих джерелах суттєво ускладнює пошук релевантних документів. Вилучення необхідних знань пов'язано з переробкою великої кількості документів, які повертають на запити пошукової системи. Одним з напрямків підвищення якості пошуку є формування запитів пошуку з урахуванням семантики ПО. Засобом відображення семантики ПО може бути онтологія [4].

Онтологія це структурна специфікація ПО, її формалізоване уявлення, яке включає словник (або імена) покажчиків на терміни предметної області та логічні зв'язки, які описують, як вони співвідносяться один з одним. Таким чином, онтології забезпечують словник для представлення та обміну знаннями з певної предметної області і безліч зв'язків, встановлених між термінами в цьому словнику.

Онтології використовуються як бази знань інтелектуальних систем, або на їх підставі створюються бази з продукційними або фреймовими моделями знань. Це дає можливість корегування запитів на пошук нових знань з допомогою онтології. Для цього з користувальницького запиту виділяються смислові структури: значущі слова і терміни предметної області. Ці смислові структури потім використовуються для формування пошукового образу із застосуванням евристичних правил і виведення на онтології. Образ релевантного документа являє собою опис бажаного результату роботи пошукової системи, яка включає в

себе набір термінів, які повинні включатися в документ, характеристики документа, набір вимог до результату пошукової системи.

На етапі виведення на онтології виконується перетворення користувальницького запиту в з'єднаний логічними зв'язками набір термінів і понять, які будуть використовуватися пошуковою системою. Після перетворення розширений і уточнений таким чином запит автоматично модифікується в запит до пошукової системи. При цьому задаються параметри пошуку, специфічні для кожної системи.

Для реалізації розглянутого підходу підвищення якості пошуку знань до складу БАС ПЗ необхідно ввести агента корегування запитів згідно з онтологією. Також змінюється алгоритм роботи агентів ранжування результатів пошуку. Аналіз відповідності документів має ґрунтуватися на виділенні ключових слів з таких частин документа як заголовки, посилання і тіло документа та використанні онтології. В разі виявлення нових термінів можливе корегування онтології користувачем – експертом.

III. ВИСНОВОК

Запропоновані типи агентів та багатоагентна організація засобів автоматизації створення і поповнення баз знань інтелектуальних систем реалізується в проєкті інтелектуального навчального середовища для циклу дисциплін напряму підготовки студентів «Комп'ютерні науки».

- [1] Левашова Т.В., Пашкин М.П., Смирнов А.В., Шилов Н.Г. Принципы построения систем для быстрой интеграции знаний из распределенных источников // Труды международного конгресса «Искусственный интеллект в XXI веке». — Дивногорское, Россия, 2001. Т. 1. — С. 105–119.
- [2] Ланин В.В. Интеллектуальное управление документами как основа технологии создания адаптируемых информационных систем // Труды международной научно-технической конференции «Интеллектуальные системы» (AIS'07). Т. 2 / М.: Физматлит, 2007. С. 334-339.
- [3] Agent projects. Europe's Network of Excellence for Agent-based Computing. <http://www.agentlink.org/resources/-agentprojects-db.html>, 2001.
- [4] Гаврилова Т.А., Лещева И.А., Лещев Д.В., 2000. Использование онтологий в качестве дидактического средства // Ж. "Искусственный интеллект" №3. - с.34-39.