

УДК 519.7:004.8

М.В. Білоіваненко, Г.А. Воскобойникова

Харківський національний університет радіоелектроніки, Харків

РОЗРОБКА КОНЦЕПЦІЇ РОЗПОДІЛЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ, ЯКА ПОБУДОВАНА НА ОНТОЛОГІЧНІЙ БАЗІ ЗНАНЬ

У роботі розглянуто проблему розподілення онтологічних інформаційних систем. Проаналізовано найбільш популярні методи розподілення інформаційних систем: організація розподілених БД, створення peer-to-peer мереж, функціональне розподілення та розподілення ресурсів. Розглянуто декілька практичних рішень цієї задачі, які вже реалізовано. На основі аналізу розглянутих матеріалів було запропоновано чотири концепції розподілення інформаційної системи, яка побудована на онтологічній базі знань. Аналіз цих концепцій виявив, що не всі класичні методи розподілення можуть бути застосовані до онтологічних систем.

Ключові слова: розподілена інформаційна система, онтологічна база знань.

Вступ

Актуальність фізичного та логічного розподілення інформаційних систем обумовлена швидкими темпами зростання масштабів інформаційних систем. Застосування таких систем дає змогу найбільш оптимально розподіляти інформаційні ресурси підприємств, вдосконалювати швидкість та якість їх обробки. Але інформаційні системи зараз перебувають в стані часткової, а декуди й повної, інтелектуалізації. Це призводить до змін у бізнес-логіці систем. Тому розподілення інтелектуальних інформаційних систем за звичними схемами значно видозмінюється і не завжди виглядає достатньо оптимально.

Огляд предметної області та постановка задачі. В Харківському національному університеті радіоелектроніки проводиться розробка «Онтологічного порталу менеджменту освітніх та наукових ресурсів МОН України». Портал являє собою web-орієнтовану розподілену систему, розроблену з використанням онтологічного підходу до побудови інформаційних систем. В основі такої інформаційної системи лежить онтологічний опис ресурсів та онтологічна база знань.

Архітектура порталу реєстрації освітніх ресурсів МОН України ґрунтується на принципах розробки розподілених систем, орієнтованих на роботу в Інтернет, для багатьох користувачів. Інформаційну основу порталу складає онтологія і пов'язаний із нею опис відповідних ресурсів.

В межах проведення дослідницької роботи у проектах «Онтологічний портал для менеджменту та оцінки національних ресурсів України в галузі освіти і науки», проект ДФФД МОН № Ф15/456-2007 (2007 р.), «Розробка Web-орієнтованої системи для підтримки процедур акредитації та ліцензування вищих навчальних закладів України», НДР № 219, ДР № 0107U001569 (2007-2008 р.р.), «Нові-

тня інформаційна технологія забезпечення прозорості акредитації університетів», проект TEMPUS SM_SCM-T020B06-2006 (UA) (2006-2007 р.р.) групою розробників було реалізовано прототип інформаційної системи.

Проблема розподілення бази знань розробленої інформаційної системи постала з необхідності подальшого впровадження онтологічного порталу для користування. Стало необхідним розподілити ресурси системи, якими є засоби зберігання та обробки інформації, через нестачу обчислювальних ресурсів та потрібних потужностей паралельного доступу до великих об'ємів інформації на існуючих носіях.

Задачею авторів було проаналізувати існуючі технології розподілення інформаційних систем та запропонувати нові, основані на існуючих, концепції розподілення, які б давали змогу найбільш ефективно реалізувати розподілення інформаційної системи, в основі якої лежить онтологічна база знань.

Концепції розподілення онтологічних інформаційних систем

Аналіз існуючих рішень для розподілення онтологічних інформаційних систем [1 – 5] показав, що на сьогодні не існує сталих архітектур чи методів розподілення інформаційних систем, що ґрунтуються на використанні інтелектуальних технологій представлення інформації. Далі наведено декілька концепцій, які стали симбіозом існуючих моделей та методів розподілення.

1. Фізичне розподілення даних. Фізичний розподіл даних – це розподіл даних та бізнес-логіки однієї системи між декількома фізичними носіями, що призводить до більш ефективної їх обробки – підвищенню швидкості обробки, спрощенню доступу до даних. Схема такого розподілення представлена на рис. 1.

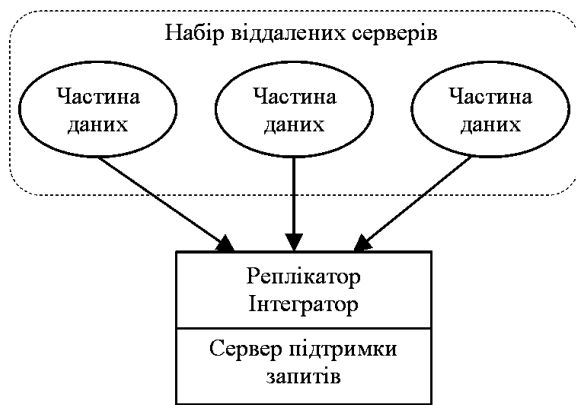


Рис. 1. Схема фізичного розподілу даних

В цій схемі показано, що для реалізації розподіленої системи потрібно створити мережу серверів, які будуть зберігати окремі частини даних та бізнес-логіки інформаційної системи та не будуть безпосередньо пов'язані між собою. Зв'язок у такій системі реалізовано через сервер, який є водночас реплікатором та інтегратором системи. Реплікатор забезпечує розподілене внесення інформації до системи та її модифікації, а інтегратор – розподілений збір даних для виконання запитів.

Фізичний розподіл даних онтологічної інформаційної системи на рівні онтології має бути реалізований як розподілення єдиної RDF-структури між декількома фізичними носіями.

Таке розподілення інформаційної системи має бути реалізовано на рівні зберігання даних. Однорідна інформація (у випадку онтологічної системи – це частина RDF-структури – клас або декілька класів з екземплярами) об'єднується в кластер СКБД. Таке рішення широко застосовується для реалізації розподілення інформації у корпоративних системах та мережах і є надійним рішенням для масштабів великої установи.

2. Розподілення інформаційної системи на основі застосування концепції розподілення даних у peer-2-peer мережах. Децентралізована обробка інформації та використання технічних засобів для цих цілей передбачає реалізацію функціональних підсистем та здійснення обробки інформації безпосередньо на робочих місцях.

Перевагами децентралізованої організації інформаційної системи є:

1) інформаційна система більш інтегрована з середовищем та краще відповідає вимогам користувачів, дані знаходяться безпосередньо у їх користувачів, які, таким чином, краще розуміють інформацію;

2) гнучкість структури, що забезпечує більшу автономність окремих користувачів;

3) окремі підсистеми менші розміром та мають простішу структуру, тому їх простіше створювати, супроводжувати та використовувати, зменшується централізований контроль за роботою підсистем;

4) цілі використання ресурсів можуть бути детально продумані, що покликане підвищити продуктивність їх опису та використання;

5) підсилення відповідальності працівників нижньої ланки.

У системі децентралізованого розподілення онтологічної системи вузол розподіленої системи (підсистема) повністю автономний в межах інформації, що зберігається та оброблюється на ньому. Кожен вузол системи повинен мати семантичну анотацію даних та функцій, які він зберігає та може надавати іншим вузлам системи.

Виходячи з поданої на вузол задачі, він самостійно повинен встановлювати зв'язки з іншими вузлами системи шляхом пошуку та опитування за своєю таблицею семантичних описів всіх вузлів системи. При цьому, отримані актуальні дані повинні бути закешовані для підвищення ефективності подальшого їх використання та обробки.

Вузли на стороні клієнта спроектовані за загальною схемою, але задачу на виконання отримують безпосередньо через інтерфейс користувача.

Загальна схема розподілення онтологічної інформаційної системи шляхом децентралізації системи представлена на рис. 2.

Взаємодія між вузлами в такій моделі здійснюється на основі вимог по виконанню задачі, що вирішується вузлом в конкретний момент роботи. Таким чином при кожному звертанні до вузла перевіряється доступ на виконання конкретної операції для конкретного вузла.

3. Функціональне розподілення інформаційної системи. Функціональне розподілення онтології має за основу створення мережі серверів, кожен з яких буде зберігати загальну структуру онтології інформаційної системи та зберігати кожен свої унікальні ресурси. Це розподілення онтологічної інформаційної системи буде відбуватися на рівні розподілення даних системи між зовнішніми репозитаріями (рис. 3). Кожен репозитарій має можливість виконувати обробку запитів до своїх ресурсів чи перенаправляти запити до синхронізатора. Взагалі, локальні вузли можуть працювати автономно в межах наданої їм схеми даних, якщо користувач не потребує додаткової інформації. Кожен вузол має достатнє функціональне оснащення для обробки локальних запитів.

При цьому всі локальні ресурси повинні мати ідентичні версії програмного забезпечення за для коректного обслуговування запитів до всієї розподіленої системи.

Також, в такій системі повинен існувати синхронізатор для можливості одночасної зміни загальної схеми ресурсу за необхідністю. Також синхронізатор контролює та оновлює версії програмного забезпечення на локальних ресурсах.

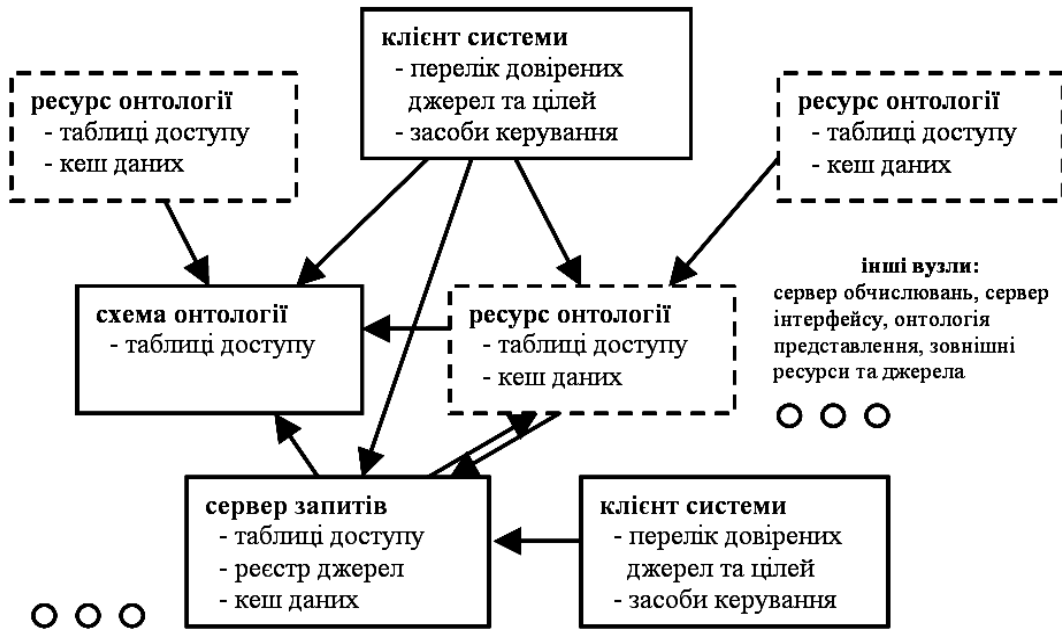


Рис. 2. Розподілення онтологічної системи шляхом децентралізації

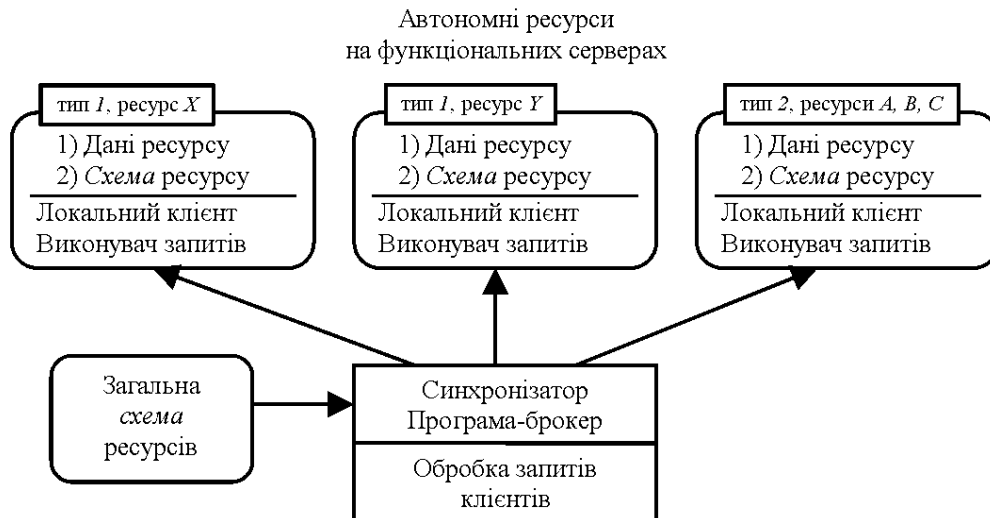


Рис. 3. Функціональне розподілення інформаційної онтологічної системи

Програма-брокер системи транслює запити клієнтів до відповідних локальних ресурсів розподіленої системи для отримання доступу до необхідних для виконання запитів ресурсів. При цьому доступ до інформації локального ресурсу може здійснюватися через центральний сервер або безпосередньо через інтерфейс локального ресурсу.

4. Розподілення інформаційної системи на рівні ресурсів з урахуванням характеру співвідношень між ними. Розподілення на рівні ресурсів в онтологічній системі передбачає розподілення на рівні онтології, яка в межах однієї системи містить усі її ресурси. В такому випадку кожен ресурс (частина єдиної онтології) повинен мати анотацію свого наповнення та бути підпорядкований загальній схемі ресурсів інформаційної системи. Загальна схема ресурсів знаходиться на окремому сервері, який за-

безпечує управління усіма ресурсами системи, доступ до них користувача, а також зберігає реєстр довірених віддалених джерел, на які можуть посилатися ресурси системи. Наповнення ресурсів може мати посилання на будь-які інші ресурси, але реалізація цих посилань проходить тільки через головний сервер управління ресурсами.

Ресурси верхнього рівня за цією схемою (рис. 4) можуть відсилати запити інформації на інші ресурси, які прямо не пов'язані з сервером. Сервер при цьому встановлює прямі зв'язки з віддаленими джерелами після звертання до ресурсу по ідентифікатору.

Таким чином, концепція розподілення ресурсів онтологічного порталу, яка представлена на рис. 4, базується на створенні незалежних віддалених серверів онтологій. Загальна схема системи виглядає таким чином, як на рис. 5.

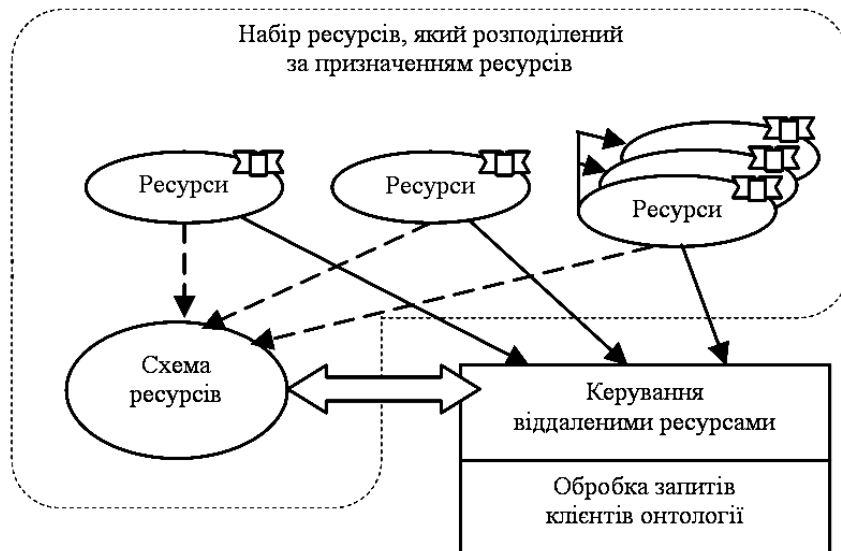


Рис. 4. Розподілення онтологічної системи на рівні ресурсів

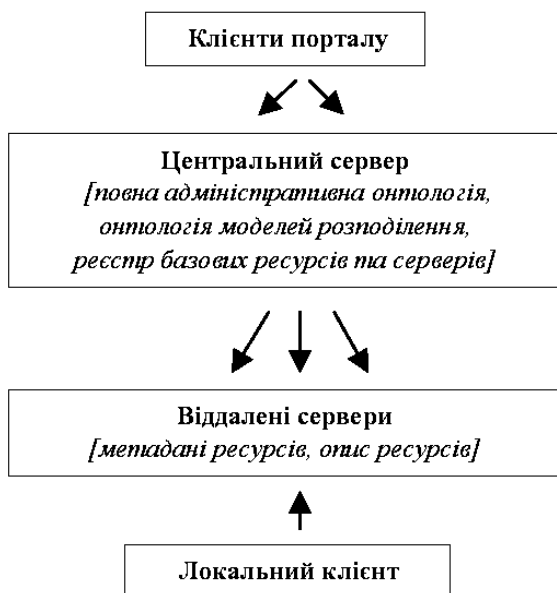


Рис. 5. Загальна схема розподілення онтологічної системи на рівні ресурсів

Центральна онтологія, яка містить схему усіх ресурсів системи, також є адміністративною частиною системи, містить моделі розподілення даних та реєстр віддалених серверів, до яких існує доступ.

Реєстр базових ресурсів містить посилання на ресурси верхнього рівня, створюється автоматично, на основі метаданих, які представлені у віддалених онтологіях. Також у реєстрі міститься колекція ресурсів верхнього рівня, які в онтологічному порталі представлені ВНЗ, та довірені сервери онтологій, які містять всі джерела даних.

Віддалена онтологія відрізняється за своєю структурою від центральної онтології системи. Основною та найважливішою частиною віддаленої онтології є дані по ресурсам інформаційної системи. Окрім цього віддалена онтологія зберігає службову

інформацію, що дозволяє їй функціонувати як частині єдиної розподіленої онтологічної системи, – метадані про ресурси, що зберігаються.

Розподілення онтологічної бази знань відбувається шляхом розділення опису класів онтології, екземплярів ресурсів та підпорядкованих ним екземплярів. При цьому створюються таблиці, які містять всі характеристики локальних ресурсів, та репозитарій, що зберігає джерела даних локальних онтологій. Таким чином проходить й розподілення системи на рівні зберігання заархівованих даних, коли частина ресурсів системи перенесена до окремої онтології, яка розміщена на власному сервері.

Для впровадження такої схеми розподілення в існуючому онтологічному порталі слід створити модуль розподіленого доступу до даних, які містяться на віддалених серверах. Також потрібні модулі кешування та проксування ресурсів, що дасть змогу зменшити кількість звертань між серверами онтологій системи при обробці однакових чи однотипних запитів. Також потрібно реалізувати модуль підтримки реєстру віддалених серверів, які мають зберігати частини загальної онтології системи.

До функції центрального серверу потрібно віднести наступне:

- 1) підтримка шифрування транзакцій системи, бо вона є розподіленою не тільки логічно, а й фізично; також реалізація авторизації запитів по парі ключів;

- 2) обслуговування реєстру довірених віддалених серверів, зберігання даних, які необхідні для підключення до них; ця функція повинна реалізуватися напряму адміністратором порталу;

- 3) автоматичне створення колекції віддалених онтологій: автоматичне створення колекції ресурсів верхнього рівня через опитування метаданих всіх віддалених онтологій; ця функція виконується при

зміненні розподіленої конфігурації, зміненні списку ресурсів верхнього рівня;

4) виконання процедури розподіленого пошуку ресурсів на основі використання їх унікального ідентифікатора або заданих характеристик ресурсу; повний ідентифікатор ресурсу в розподіленій моделі може складатися наступним чином:

*http://[адреса_віддаленого_серверу]/
[назва_віддаленої_онтології]#
[локальний_ідентифікатор_ресурсу]*

5) виконання прозорого проксування, що має забезпечувати доступ до віддалених ресурсів; також виконання кешування інформації в межах однієї сесії роботи інформаційної системи;

б) виконання процедури віддаленої модифікації даних на довірених серверах в межах розподіленої інформаційної системи.

До функцій віддаленого серверу розподіленої системи слід віднести наступне:

1) підтримка шифрування та авторизації по парі ключів для всіх користувачів віддаленого ресурсу – для зовнішніх та внутрішніх транзакцій;

2) занесення, модифікація та оновлення віддаленої онтології; ці функції мають виконуватися виконуватися тільки через центральний сервер, аби запобігти порушення цілісності розподіленої онтологічної системи;

3) автоматичне керування ресурсами та метаданими в межах локального клієнту, що знімає загальне навантаження по обслуговуванню онтологічної системи з центрального серверу.

Аналіз запропонованих концепцій

При використанні розподілення за допомогою кластерів СКБД для окремих частин даних інформаційної системи виникає проблема розмежування доступу до інформації в межах одного кластера. Для зовнішнього сервера кластер є єдиною інформаційною одиницею, тому розмежувати рівні доступу стає неможливим. Це є недопустимим, насамперед, для частини RDF-структури, бо один екземпляр такої структури містить в собі вичерпну інформацію про будь-який об'єкт, але ця інформація в межах бізнес-логіки інформаційної системи може бути доступна не всім користувачам.

Наприклад, інформація про окремого викладача ВНЗ має бути повністю доступна відділу кадрів цього ВНЗ та частково доступна сторонньому користувачу (наприклад, студенту).

На сьогодні не існує рішень практичного масштабу розподілу даних на рівні онтологічних структур або RDF-графу. Основні причини цього такі:

а) реалізація розподілу самостійних файлів онтології неможлива, бо на сьогодні ще не вирішена проблема узгодженої віддаленої модифікації онтологічної структури даних;

б) складною є реалізація інтегратора даних, бо це потребує реалізацію одночасної обробки великих об'ємів інформації, що виключає для онтологічної системи основну вигоду від розподілення даних на фізичному рівні – прискорення обробки та спрощення доступу до даних.

Недоліками децентралізованої організації інформаційної системи є:

а) необхідність дублювання ресурсів;

б) можливість неефективного використання ресурсів;

в) проблеми з сумісністю та стандартизацією ресурсів за рахунок великої кількості користувачів, які створюють ресурси;

г) проблеми керування такою системою;

д) нерівномірність розвитку рівнів інформаційної системи в різних її підрозділах та ін.

Децентралізований підхід до організації інформаційної системи може бути застосований, якщо:

а) організація, яка користується інформаційною системою має значний розмір;

б) децентралізація має підтримку органів керування;

в) існує потреба у швидкості та гнучкості інформаційного обслуговування;

г) в межах однієї організації виконуються суттєво різні операції, різні інформаційні сервіси мають бути підпорядковані зазначеним підрозділам та ін.

При функціональному розподіленні кожен вид серверів системи орієнтований на окремий клас ресурсів. Створення універсального локального сервера не передбачено, що суттєво зменшує гнучкість системи.

Розподілення інформаційної системи на рівні ресурсів з урахуванням характеру співвідношень між ними дає можливість створити гнучку та ефективну модель розподілення ресурсів онтологічної системи, яка базується на загальних принципах Semantic Web.

При розподіленні інформаційної системи на рівні ресурсів стає можливим досягнення наступних цілей:

1) гнучке розподілення навантаження на декілька серверів;

2) паралельна обробка складних запитів;

3) автономне керування даними кожного окремого ресурсу без підключення до центрального серверу;

4) створення віддаленого серверу зовнішнього джерела, що дає можливість інтегрувати нові ресурси в існуючу базу даних системи.

Висновки

В роботі було розглянуто існуючі архітектури та методи розподілення інформаційних систем за для підвищення ефективності їх роботи. Аналіз цих мо-

делей було проведено у контексті вирішення задачі розподілення інформаційної системи, що побудована на онтологічній базі знань. Задача розподілення такої системи не є тривіальною, бо складність структури інформації системи сильно відрізняється від структури інформації звичайних інформаційних систем, в основі яких лежать реляційні методи представлення даних.

На основі аналізу існуючих методів розподілення інформаційних систем та опиту, що було надбано при вирішенні задач розподілення онтологічних систем, авторами було запропоновано чотири концепції розподілення онтологічних систем. Це розподілення на фізичному рівні даних, розподілення на основі застосування концепції peer-2-peer мереж, функціональне розподілення та розподілення на рівні ресурсів з урахуванням характеру співвідношень між ними. У межах цих концепцій було розглянуто організацію та розміщення структури та наповнення онтологічної бази знань або її частин.

Здійснивши аналіз запропонованих концепцій розподілення, автори надійшли висновку, що найбільш оптимальними для розподілення онтологічних інформаційних систем будуть методи функціонального розподілення та розподілення на рівні ресурсів з урахуванням характеру співвідношень між ними.

Список літератури

1. Vdovjak Richard. *A Model-driven Approach for Building Distributed Ontology-based Web Applications* / Richard Vdovjak. – Eindhoven: Technische Universiteit Eindhoven, 2005. Proefschrift, ISBN 90-386-0584-6, NUR 983, SIKS Dissertation Series No. 2005-08. – 180 с.

2. Quilitz Bastian. *Querying Distributed RDF Data Sources with SPARQL* / Bastian Quilitz, Ulf Leser – Springer Berlin/Heidelberg, *The Semantic Web: Research and Applications*, ISBN 978-3-540-68233-2 Volume 5021/2008. – P. 524-538.

3. Cai Min. *RDFPeers: A Scalable Distributed RDF Repository based on A Structured PeertoPeer Network* / Min Cai, Martin Frank. – *Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, 2004*, ISBN:1-58113-844-X – P. 650-657.

4. Haas Christine. *RDFS Reasoning and Query Answering in distributed environments* / Christine Haas. – *Seminar Peer-to-Peer Netzwerke, Summer term 2009*. – P. 7.

5. Langegger Andreas. *SemWIQ – Semantic Web Integrator and Query Engine* / Andreas Langegger, Wolfram Wob – Springer Berlin/Heidelberg, *The Semantic Web: Research and Applications*, ISBN 978-3-540-68233-2, Volume 5021/2008. – P. 493-507.

6. Таненбаум Э. *Распределенные системы. Принципы и парадигмы* / Э. Таненбаум, М. Ван Стеен. – СПб.: Питер, 2003. – 877 с.

7. *Звіт про науково-технічну роботу „Розробка і впровадження розподіленої архітектури онтологічного порталу МОНУ для надійного, безпечного та ефективного менеджменту та інтеграції освітніх ресурсів України”, № IT/534-2009 від 20 липня 2009 р., Етап 1. Розробка загальної апаратно-програмної архітектури розподіленої версії онтологічного порталу (проміжний), Харківський національний університет радіоелектроніки, каф. штучного інтелекту, 10-2009.*

Надійшла до редколегії 12.12.2009

Рецензент: д-р техн. наук, проф. С.І. Кучеренко, Харківський національний університет радіоелектроніки, Харків.

РАЗРАБОТКА КОНЦЕПЦИИ РАСПРЕДЕЛЕНИЯ ИНФОРМАЦИОННОЙ СИСТЕМЫ, КОТОРАЯ ПОСТРОЕНА НА ОНТОЛОГИЧЕСКОЙ БАЗЕ ЗНАНИЙ

М.В. Белоиваненко, А.А. Воскобойникова

В работе рассмотрена проблема распределения онтологических информационных систем. Проанализированы наиболее популярные методы распределения информационных систем: организация распределенной БД, создание peer-to-peer сетей, функциональное распределение и распределение ресурсов. Рассмотрено несколько практических решений этой задачи, которые уже реализованы. На основе анализа рассмотренных материалов были предложены четыре концепции распределения информационной системы, которая построена на онтологической базе знаний. Анализ этих концепций обнаружил, что не все классические методы распределения могут быть применены к онтологическим системам.

Ключевые слова: распределенная информационная система, онтологическая база знаний.

CREATION THE CONCEPT OF INFORMATION SYSTEM DISTRIBUTION, WHICH BUILT ON ONTOLOGY KNOWLEDGE BASE

M.V. Bilovanenko, G.A. Voskoboynikova

The problem of distributing of the ontological informative systems is considered in work. The most popular methods of distributing of the informative systems are analysed: organization of distributed DB, creation of peer-to-peer networks, functional distributing and allocation of resources. A few practical decisions of this task, which are already realized, are considered. On the basis of analysis of the considered materials four conceptions of distributing of the informative system which is built on the ontological base of knowledges were offered. The analysis of these conceptions discovered that not all classic methods of distributing can be applied to the ontological systems.

Keywords: distributed informative system, ontological base of knowledges.