

Антон ВІТУШКО,

провід. інженер СІАЗ НБУВ

ОНТОЛОГІЇ АНАЛІТИЧНИХ СЛУЖБ У БІБЛОТЕЧНОМУ СЕРЕДОВИЩІ

Стаття присвячена питанням використання онтологій у системах керування знаннями (СКЗ). Розглядаються особливості використання онтологій в аналітичних службах. Пропонуються моделі подання знань, які забезпечуватимуть автоматизовану обробку інформації на семантичному рівні в системах керування знаннями. Висвітлюється роль онтологій в процесі вибудовування СКЗ в аналітичних службах.

Ключові слова: онтології, системи керування знаннями (СКЗ), інформаційні технології, семантичний, формальні мови, моделювання предметної галузі, дескриптивні логіки, метаописання, knowledge representation.

Онтологія¹ є ключовою технологією семантичної обробки знань. Досвід роботи аналітичних служб показав, що перспектива подальшого розвитку аналітичних служб тісно пов'язана з керуванням знаннями й заснована на застосуванні концептуальних онтологічних моделей.

Сьогодні вузьким місцем при розробці систем керування знаннями є не програмний аспект, а завдання добування, формулювання, структурування й подання інформації, тобто даних і знань. Розглянемо роль онтологій як моделі опису знання в аналітичних підрозділах.

На сьогодні діяльність аналітичних служб цілком залежить від наявних у них знань і здатності їх ефективно використовувати. Однак засоби подання цих знань поки що недостатньо досконалі й часто змушують знову й знову шукати рішення тих самих завдань.

Усе це значно ускладнює роботу, тому важливо розробляти моделі подання знань, які забезпечували б автоматизовану обробку інформації на семантичному рівні в системах керування знаннями (СКЗ).

Керування знаннями є цілеспрямованою стратегією аналітичних служб, орієнтованою на ефективне використання всієї наявної інформації, досвіду й кваліфікації співробітників, скорочення часу реакцій на оперативно мінливі потреби споживача інформаційного продукту, що

¹ Онтологія – цілісна структурна специфікація певної предметної галузі, її формалізоване подання, що включає словник показчиків на терміни предметної галузі й логічні вираження, що описують, як вони співвідносяться один з одним.

виводить функціонування аналітичних підрозділів на новий рівень.

Керування знаннями – це комплексна організаційно-технічна діяльність, спрямована на підвищення ефективності використання знання в аналітичному середовищі. При цьому знання класифікуються й розподіляються за категоріями відповідно до визначених онтологією структурованих і слабко структурованих баз даних і баз знань.

Успішне вибудовування «керування знаннями» залежить від багатьох факторів. Одним з них є використання інформаційних технологій. Для успішного керування знаннями в аналітичних підрозділах створюються спеціальні інформаційно-програмні системи керування знаннями. Такі системи можуть допомогти вирішити проблеми, пов'язані з неструктурованим станом масивів даних, і проблеми, пов'язані з недостатньо ефективним використанням наявної інформації.

Виділяють два підходи побудови СКЗ, які значно відрізняються між собою [2]. Перший підхід – «технологічний» – СКЗ будується на основі комбінування й інтеграції існуючих технологій для підтримки різних процесів роботи зі знаннями. До них належать такі інформаційні технології, що вже стали стандартними: електронна пошта, дошки оголошень, дискусійні форуми, веб-портали; а також технології, що використовуються при побудові експертних систем (розпізнавання образів і ситуацій, автоматичне анотування й класифікація і т. ін.). Даний підхід є подальшим розвитком інформаційних систем організації стосовно роботи зі знаннями.

Другий підхід – «семантичний» – заснований на використанні методів і технологій роботи зі змістом, семантикою даних, інформації й знань, таких як онтології предметних галузей, технології їхньої побудови й супроводу, семантичні метадані, семантичний пошук, системи логічного висновку, семантичне профілювання знань експертів, семантичні портали й мережі тощо. Усе це – з відповідною технологічною підтримкою в частині мов опису, моделей, програмних інструментів і систем. При цьому семантичний підхід не відкидає технологічного. Більшість елементів та інструментів семантичного підходу застосовується для підвищення рівня роботи з даними й інформацією до рівня роботи зі знаннями.

У момент виникнення в аналітичних підрозділах проблеми керування знаннями, безліч досліджень у сфері штучного інтелекту були спрямовані на вирішення проблем подання знань (knowledge representation) і виконання логічного висновку на знаннях, які поєднували ранній досвід концептуального моделювання з використанням формальних логік. Центральним елементом цих досліджень було поняття «онтологія». Дослідники онтологій розглядають керування знаннями як одну з найбільш важливих галузей застосування онтологій.

Для розуміння користі від використання онтологій у керуванні знаннями аналітичних служб потрібен пошук таких процесів і сценаріїв роботи зі знаннями, які можуть виграти від застосування онтологій. Для цього більш ретельно розглянемо їхні основні властивості. Виходячи з найбільш часто цитованого визначення, онтологія – це спільно використовувана, формальна класифікація предметної галузі [3].

Можна виділити три специфічні особливості онтологій:

- онтології будуються на основі спільного й однозначного розуміння предметної галузі в межах співтовариства. Це розуміння формалізується угодою експертів з приводу понять і відносин, які є в предметній галузі (людський фактор у керуванні знаннями, заснований на онтологіях);

- онтології описують зв'язки між сутностями предметної галузі. Повноцінні онтології мають більше можливостей для опису відносин, ніж прості таксономічні ієрархії, які є в сучасних рішеннях щодо керування знаннями. Відносини (зв'язки між поняттями) значно збільшують кількість способів здійснення навігації й пошуку предметною галуззю, виконання аналізу, класифікації й візуалізації знань (структурний фактор у керуванні знаннями, заснований на онтологіях);

- онтології використовують спосіб подання, що дає можливість комп'ютерної обробки накопичених знань аналітичної служби. Тобто зберігають з використанням формальних мов, таких як Ontolingua, CycL, мов, заснованих на дескриптивних логіках, – LOOM, мов, заснованих на фреймах, – OKBC, OCML, Flogic, мов, заснованих на веб-стандартах, – XOL, SHOE, UPML [5]. Такий вид подання дає змогу передавати онтології між комп'ютерами (мови для обміну онтологіями через Web – RDF(S), DAML, OIL, OWL), зберігати онтології, перевіряти погодженість онтологій, виконувати логічний висновок на онтологіях і за допомогою онтологій (комп'ютерний фактор у керуванні знаннями, заснованими на онтологіях).

Ці особливості можуть підняти рівень інтелектуальності програмних систем, що використовуються аналітичною службою, шляхом включення до них знань про предметну галузь у формі онтологій. Відзначимо взаємозв'язок зазначених вище факторів – онтологія інтегрує розуміння предметної галузі людиною й комп'ютером, – у той час як комп'ютери можуть маніпулювати й виконувати логічне виведення на знаннях про предметну галузь за допомогою онтологій, розуміння того, що позначають ці символи й правила, залишається функцією людини. Тому онтологія не може існувати без співтовариства, що її підтримує.

Важливість онтологічного підходу обумовлена також тим, що знання, які не використовуються й не зростають, в остаточному підсумку стають застарілими й марними. Знання ж, які поширюються, генерують нові знання.

Метою керування знаннями є систематичне відновлення знань як ресурсу аналітичної служби, важливість якого постійно зростає. Ресурси знань розрізняються залежно від галузей індустрії, але, як правило, включають керівні документи, листи, новини, інформацію про замовників, відомості про конкурентів і технологію, що нагромадилися в процесі розробки. Самі ці ресурси можуть перебувати в різних місцях: у базах даних, базах знань, архівах, технічних бібліотеках, в експертів і можуть бути розосереджені по всій організації. При цьому часто один відділ служби дублює роботу іншого саме тому, що співробітникам відділу важко знайти й використати інформацію, що міститься в сусідніх підрозділах.

Будь-яка система автоматизації зачіпає проблеми зберігання знань, але тільки системи керування знаннями орієнтовані на це в явному вигляді, тим самим сприяючи збереженню цього найціннішого ресурсу. Роль онтології в процесі вибудовування СКЗ в аналітичних службах полягає у формуванні цілісного підходу до керування ресурсами, при якому досягаються:

- системність – онтологія представляє цілісний погляд на предметну галузь, дає можливість структурувати й класифікувати всі знання підприємства;
- стандартність – матеріали, представлені в єдиній формі набагато краще сприймаються й відтворюються;
- науковість – побудова онтології дає змогу відновити відсутні логічні зв'язки в усій їхній повноті.

Важливість онтології визначає її місце в процесі розробки системи керування знаннями аналітичних служб. Зокрема, при розробці таких систем можна виділити такі етапи [7, 8]:

1. Стихійне й безсистемне нагромадження інформації в організації.
2. Процес виявлення джерел даних і знань, їх опис.
3. Структурування – розробка онтологічної моделі знань аналітичної служби з метою структурування масиву інформації. Вибір на даному етапі між простою таксономією та повноцінною онтологією визначає подальше використання та цінність розроблювальної системи керування знаннями, а також складність подальших етапів процесу.
4. Формалізація й програмна реалізація – автоматизація процесів керування знаннями, подання структурованої інформації у форматах машинної обробки, тобто на мовах опису даних і знань, і організація автоматизованої обробки й пошуку інформації з запиту.
5. Обслуговування – коректування формалізованих знань (додавання, відновлення, видалення застарілої інформації і т. ін.).

Таким чином, онтологія створюється на третьому етапі розробки системи керування знаннями й надалі є основою для автоматизації

процесів керування знаннями та забезпечення даних процесів. Зрозуміло, що онтологія є стартовим майданчиком для тих організацій, які хочуть створити інтегровану систему, а не «мозаїку» окремих функціональних блоків.

У застосуванні онтології для керування знаннями можна виділити такі типові проблеми:

1. Проблема автоматичного породження онтології з існуючих масивів даних або аналогічну їй проблему прив'язки існуючого масиву даних до розробленої онтології. Частково цю проблему дають змогу вирішити засоби автоматичного анотування й побудови тезаурусів.

2. Вибір методики моделювання предметної галузі таким чином, щоб забезпечити можливість використання вже створених онтологій різних предметних галузей. Забезпечення перевірки погодженості онтології, що розроблюється, із вже створеними.

3. Проблема узагальнення онтології. Розробку підстав для об'єднаного використання даних, документів і неформального знання – побудови метаонтологій, що поєднують різні дані й знання.

4. Проблема об'єднання логічного висновку й інформаційного пошуку. Об'єднане використання формальних і неформальних подань знань і даних – це послідовне зближення логічних методів і методів інформаційного пошуку й індексації даних.

5. З'єднання аналітичних процесів і керування знаннями. Остаточна мета полягає в тому, щоб виявляти інформаційну потребу протягом виконання аналітичних процесів і визначати доречне знання в специфічному контексті завдання.

Створення єдиної онтології для детального опису моделі знань організації є тривалим процесом із залученням колективу експертів, знання яких повинні охоплювати всі напрями діяльності аналітичної служби. Рішенням цієї проблеми може бути відмова від глибокої декомпозиції системи та включення в онтологію тільки найбільш значущих понять із розглянутих предметних галузей або виділення одного з напрямів діяльності служби й створення вузькоспеціалізованої онтології. Зрозуміло, що перший підхід дає занадто грубу й узагальнену модель, а другий не дає змоги використовувати модель у масштабах всієї служби.

Для вирішення завдання побудови онтології аналітичної служби пропонується структурувати онтологічну модель у такий спосіб [1]:

$$O = \{O_o, O_r, O_z\},$$

де O_o – онтологія організації, O_r – онтологія інформаційних ресурсів, а $O_z = \{O_1, \dots, O_t\}$ – ієрархічно організована, послідовно розширювана система онтологій основних галузей знання O_i (онтологія предметної галузі), значущих для роботи підприємства.

Виділення ієрархії галузей знання служби дає можливість створюю-

вати окремо онтології різних підгалузей знань, які можуть мати різну детальність, залежно від цілей моделювання.

Онтологія організації O_o включає основні поняття, які описують структуру, склад елементів і роботу організації.

Онтологія інформаційних ресурсів O_r включає опис усіх видів ресурсів даних і інформації організації (документи, файли, бази даних, програми та ін.).

Онтологія предметної галузі O_i може бути:

- специфічна для прикладної галузі. У багатьох дисциплінах розробляються стандартні онтології, які можуть бути використані експертами за предметними галузями (доменами) для спільного використання й анутовання інформації у своїй галузі;

- специфічна для завдання. Це онтологія, що використовується конкретно прикладною програмою й має терміни, які використовуються при розробці ПЗ, що виконує конкретне завдання. Вона має специфіку додатка, але може також містити деякі загальні терміни.

Цикл життя моделі знань аналітичної служби може бути розділений на три етапи: створення, керування й підтримка.

На стадії створення ідентифікуються мета, можливості й вимоги до моделі знань аналітичних служб, у тому числі вибір або розробка спеціальної мови для подання онтології. Коли всі вимоги ідентифіковані, можна розпочинати виділення концептів – базових понять даної предметної галузі для конкретної служби. Основні джерела визначення концептів – експерти аналітичної служби, різні документи, інші онтології. Далі визначаються властивості концептів і відносини між ними.

Наступним кроком даної стадії потрібно проаналізувати початкову версію моделі знань аналітичних служб, проконсультуватися з різними фахівцями для усунення суперечностей. Цей процес інтерактивного проектування триватиме протягом усього життєвого циклу моделі знань. За результатами цього аналізу модель знань може бути структурована на онтологію служби, онтологію інформаційних ресурсів і онтологію предметних галузей.

На стадії керування на основі моделі знань аналітичної служби розгортається система керування знаннями. На даному етапі здійснюється заповнення значень елементів онтології екземплярами. Термін «екземпляр» використовують для подання елементів у предметній галузі, тобто елемента даного концепту. Онтологія разом з безліччю окремих екземплярів становить базу знань аналітичної служби. Таким чином, на даній стадії життєвого циклу онтологія вирішує завдання керування знаннями й пошуку знань.

На стадії підтримки експерти можуть додавати, оновлювати й видаляти елементи онтології через користувальний інтерфейс.

Структура СКЗ, заснована на єдиній моделі знань організації [1], представлена як:

1. Єдина модель знань служби.
2. Метаописання всіх об'єктів щодо моделі знань.
3. Підсистема пошуку знань (оцінка семантичної близькості метаданих).
4. Об'єкти – контейнери знань.
5. Функціональні підсистеми.

Модель знань є концептуальною основою СКЗ: вона визначає безліч понять і відносин, а також правила їхнього спільного використання. База знань СКЗ (екземпляри понять) формується з метаописів всіх об'єктів, які можуть містити знання.

Підсистема пошуку знань обробляє метадані об'єктів і відбирає ті з них, які задовольняють запит користувача. Можливості підсистеми пошуку знань використовуються функціональними підсистемами порталу СКЗ, які надають користувачам різні сервіси (навігація за елементами бази знань і репозитарію документів).

Робота різних підсистем порталу (пошук, категоризація знань) пов'язана з оцінкою семантичної близькості пар об'єктів, а точніше їх метаописів. Для виконання логічного виведення використовуються дескриптивні логіки. Дескриптивні логіки описують знання в термінах понять і обмежень ролей, які використовуються для автоматичного виведення.

Нижче наводиться список найпоширеніших рішень для керування знаннями, які можна створити на основі онтології [2, 3]:

- системи ухвалення рішення. Ці системи використовують технології штучного інтелекту для подання рішень. Як база даних використовуються типові ситуації або сценарії, що мали місце в роботі аналітичної служби. Такі сценарії описують майбутній досвід у тій чи іншій ситуації;

- бази й архіви даних. Бази даних – це сукупність структурованих даних, які можна використати за допомогою одного або декількох додатків. Архіви даних – це база даних, у якій зберігаються поточні й колишні дані, що використовувалися на всіх рівнях організацій для управлінського аналізу й прийняття рішень;

- інформаційна проходка – це проведений за допомогою спеціальних програмних інструментів аналіз зібраних даних для виявлення певних закономірностей;

- система керування документообігом. Система використовується для зберігання найрізноманітніших документів, файлів, зображень, а також їх властивостей, історії їх життя й забезпечення їх життєдіяльності (створення, зберігання версій, публікація, маршрутизація й контроль виконання, передача документа для зберігання в архів);

– експертні системи. Експертні системи містять знання з вузької і спеціальної галузі;

– екстранет, інтранет – це приватна мережа, доступ до якої має обмежена кількість зовнішніх користувачів;

– програмне забезпечення групового користування. Як і інші, засновані на базі Інтернет інструменти, це програмне забезпечення створює умови для обміну інформацією, надаючи групам користувачів можливість працювати спільно в різних формах.

Отже, створення онтологій аналітичних служб є перспективним напрямом сучасних досліджень з обробки інформації. При створенні онтологій виникає ряд проблем, які необхідно послідовно вирішувати. Варто підкреслити, що сьогодні існує ряд великих онтологій, побудованих як у межах окремих предметних галузей, так і для незамкнених галузей знання. Найперспективнішою є автоматизація створення онтологій, однак на даному етапі ще не розроблені ефективні процедури, застосування яких дасть змогу скоротити частку помилок. Тому процес створення онтологій є дещо трудомістким. Однак на сьогодні існує ряд успішних розробок з використання онтологій. Тому необхідно продовжити подальше вивчення цієї проблеми, проводити аналіз уже існуючих онтологій для побудови й використання онтологій аналітичними службами бібліотек.

Список використаних джерел

1. *Тузовский А. Ф.* Онтолого-сематические модели в корпоративных системах управления знаниями / А. Ф. Тузовский. – Томск : ТПУ, 2007. – 381 с.

2. *Gruber T. A.* Translation Approach to Portable Ontology Specifications / T. A. Gruber // Knowledge Acquisition. – 1993. – № 5 (2). – Р. 199–220.

3. *Гаврилова Т. А.* Использование онтологий в системах управления знаниями [Электронный ресурс] / Т. А. Гаврилова – Режим доступа: http://big.spb.ru/publications/bigspb/km/use_ontology_in_suz.shtml. Загл. с экрана.

4. *Гладун А. Я.* Онтология в корпоративных системах // [Электронный ресурс] / А. Я. Гладун, Ю. В. Рогушина. Корпоративные системы. – 2006. – № 1. – Режим доступа: <http://www.management.com.ua/ims/ims116.htm/>. Загл. с экрана.

5. *Клещев А. С.* Классификация свойств онтологий. Онтологии и их классификации : препринт. / А. С. Клещев, Е. А. Шалфеева. – Владивосток : ИАПУ ДВО РАН, 2005. – 19 с.

6. *Гаврилова Т. А.* Онтологический инжиниринг [Электронный ресурс] / Т. А. Гаврилова – Режим доступа: <http://big.spb.ru>. Загл. с экрана.

7. *Гаврилова Т. А.* Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб.: Питер, 2000. – 384 с.

8. *Джанетто К.* Управление знаниями. Руководство по внедрению корпоративной системы управления знаниями / К. Джанетто, Э. Уиллер. – М.: Добрая книга, 2007. – 175 с.