

Антон Вігушко,

наук. співроб. СІАЗ НБУВ

ОНТОЛОГІЇ ЯК ЗАСІБ ОРГАНІЗАЦІЇ ВЕБ-ПОРТАЛІВ АНАЛІТИЧНИХ СТРУКТУР КОМУНІКАЦІЇ

У роботі розглядаються можливості використання онтологій як засобів організації веб-порталів. Наводяться різні аргументи на користь даного підходу. Пропонуються програмні інструменти, які сприятимуть використанню онтологій для проектування веб-порталів. Розглядаються можливості використання онтологій для пошуку, добування, подання й зберігання інформації.

Ключові слова: веб-портал, Інтернет, Інтранет, семантичні мережі, онтології, тезаурус, семантика, релевантність.

На сьогодні в мережі Інтернет представлені значні обсяги інформації, які рік у рік неконтрольовано зростають. Більша частина цієї інформації стає практично недоступною через неефективну роботу пошукових систем, що не враховують семантику запитів. Проблема погіршується ще й через те, що як основне завдання пошукових систем сьогодні розглядається індексація ресурсів мережі при повній відсутності засобів аналізу змісту представленої в них інформації. Це призводить до того, що дедалі більшу частину інформаційного простору завойовують великі й середні веб-портали, які використовують механізми виборчого пошуку й об'єднання інформаційного змісту інтернет/інтранет-мереж. Однак через швидке нагромадження інформації користувачі мережі зіштовхуються з проблемою практично повної відсутності її структуризації, що ускладнює як пошук, так і логічне осмислення змісту. Досить часто доводиться долати певні труднощі під час пошуку потрібної інформації навіть у межах одного порталу. Це, як правило, є наслідком слабкої структурованості веб-порталу. Але ж одна з основних характеристик будь-якого веб-ресурсу – це можливість легко й швидко знайти будь-яку інформацію.

Інтернет-технології, які існують на сьогодні, включаючи портали, змушують користувача працювати в слабоструктурованому середовищі, перекладаючи на нього основний вантаж переходу від примітивних до більш складних форм й інтелектуальних методів роботи з мережевою інформацією [7].

Отже, потребує вирішення питання про організацію веб-порталів на основі онтологічних методик концептуалізації й систематизації слабоструктурованої інформації, а також розробки і використання ефективних засобів пошуку й аналізу інформації.

Сьогодні проводиться велика кількість досліджень в галузі онтологій, у тому числі й у роботах з інтелектуалізації інформаційного пошуку, у першу чергу – у середовищі Інтернет. Загальною метою таких проектів є розробка нових підходів до побудови просторів знань і засобів роботи з ними [5].

Особливістю онтології веб-порталів, орієнтованих на пошук інформації в Інтернеті, є наявність у ній опису мережевих ресурсів поряд з традиційним описом предметної області. Така онтологія на основі попереднього індексування групує посилання на інтернет-ресурси в категорії й співвідносить їх з поняттями, які в ній описані. На основі властивостей мережевих ресурсів онтологія відслідковує взаємозв'язки між інформаційними джерелами для співвіднесення їх з певною тематикою, групою, інтересами й іншими параметрами.

Онтології веб-порталів можуть використовуватися не тільки для локалізації пошуку в проіндексованих даним порталом ресурсах, але й для уточнення формулювання запиту під час виконання пошуку в усій мережі.

Зазначимо, що для вирішення завдання пошуку інформації також повинні бути створені спеціальні засоби поповнення онтології та її інтеграції з іншими предметними онтологіями.

Отже, веб-портали потребують інструментів структуризації й концептуалізації їх інформаційного наповнення. Використання онтологій як одного з таких інструментів має ряд досить серйозних переваг.

Візуалізація онтологій дає змогу зробити видимими семантичні мережі пам'яті людини. Ці мережі складаються з вузлів – різних концептів, а також з різних зв'язків або співвідношень, які описують взаємодії між ними. У зв'язку з цим розробка семантичних мереж має на увазі глибокий аналіз структурних взаємодій між цими концептами. У свою чергу, аналіз дає змогу розроблювачеві більш оптимально структурувати веб-сайт, тому що він буде чітко уявляти собі, як пов'язані між собою окремі вузли семантичної мережі (зокрема, як пов'язані окремі розділи portalу один з одним). Ще одна перевага використання онтологій як інструменту концептуалізації веб-порталів є системний підхід до питань, пов'язаних з їх проектуванням.

У найпростішому випадку побудова онтології зводиться до:

– виділення концептів – базових понять даної предметної області;

– побудови зв'язків між концептами – визначення співвідношень і взаємодій базових понять;

– зіставлення побудованої онтології з наявними – проведення паралелей з іншими галузями знань.

Є безліч різних підходів і методологій щодо створення онтологій. Більшість з них базується на принципах, запропонованих Грубером [3]:

1. Ясність (Clarity) – онтологія повинна ясно передавати зміст уведених термінів (вони не повинні викликати почуття когнітивного дисонансу в користувача).

2. Погодженість (Coherence) – онтологія не повинна бути логічно суперечливою.

3. Розширюваність (Extendibility) – онтологія повинна допускати можливість розширення.

4. Мінімум впливу кодування (Minimal encoding bias) – концептуалізація, що лежить в основі створюваної онтології, повинна бути специфікована на рівні представлення, а не символічного кодування.

5. Мінімум онтологічних зобов'язань (Minimal ontological commitment) – онтологія повинна містити тільки найбільш істотні припущення про модулювання світу.

Кожний веб-портал може мати ієрархічну або мережеву структуру, тобто складатися з більш спеціалізованих порталів, пов'язаних певним форматом, наприклад, співвідношенням «загальне – частка». Така архітектура робить портал знань гнучким і легко розширюваним [7]. В основі порталів знань лежать онтології, що містять опис обладнання й типології відповідних мережевих ресурсів.

Виділимо три основні компоненти веб-порталу: онтологія, колекціонер онтологічної інформації про ресурси, конструктор запитів і формувач відповідей.

Онтології веб-порталів містять як опис предметної області (ПО), так і опис релевантних їй ресурсів. Опис предметної області включає сукупність термінів і відносин, семантично значущих для даної предметної області, а також правил, згідно з якими можна будувати твердження про елементи ПО. Опис ресурсу мережі включає список посилань на сторінки й сайти, а також опис зв'язків між ними.

Колекціонер онтологічної інформації про ресурси, використовуючи стандартні механізми пошуку, послідовно обробляє посилання на документи, сторінки, сайти, визначаючи їх тематику, структуру, ступінь релевантності між тематикою й інформацією, яка в них утримується. Отримані відомості накопичуються в онтології. Колекціонер повинен аналізувати нові джерела й перевіряти старі.

Формувач відповідей, обробляючи джерела, відслідковує ступінь їх релевантності поняттям і видає список найбільш підходящих для даного поняття ресурсів.

Конструктор запитів. Портал знань має мову запитів, близьку до природної мови. Це потребує наявності в ньому відповідного лінгвістичного забезпечення, одним з основних компонентів якого є словник термінів, органічно пов'язаний з онтологією [2].

На основі такого лінгвістичного забезпечення повинні вирішуватися такі важливі завдання:

- завдання лінгвістичної інформації про дану предметну область природною мовою. Для цього необхідна побудова словника, організованого з урахуванням проблемно орієнтованої лексики, морфології, словотвору, синонімії;

- опис системи знань про предметну область у вигляді комплексу пов'язаних між собою понять.

Для опису вхідної мови, близької до природної, можна використовувати спеціалізований словник-тезаурус, у якому відбиті у вигляді ієрархічних лексичних шаблонів усі можливі входження термінів (слів і словосполучень) конкретної предметної області, що дає змогу задавати в запиті терміни не тільки українською та російською, але й іншими мовами. Надалі ця ієрархія понять може бути використана для інтеграції словника-тезауруса й онтології понять в єдину мережу знань.

Завдяки інтеграції словника-тезауруса й онтології веб-портал стає здатним «розуміти» різномовні ресурси й сприймати запити на різних мовах. При цьому онтологія стає незалежною від конкретних мов, тому що опис мовних (лінгвістичних) властивостей понять відбивається в тезаурусі, водночас як в онтології подана інформація тільки про їх семантику [4].

Як уже було сказано вище, онтологія веб-порталу містить як опис предметної області, так і опис релевантних їй ресурсів.

Частина онтології, що описує конкретну предметну область, містить у собі сукупність термінів і відносин, семантично значущих для даної предметної області, а також правил, згідно з якими можна будувати твердження про елементи ПО.

Процес побудови онтології розпадається на серію підпроцесів. Спочатку складається глосарій термінів, який надалі використовується для дослідження властивостей і характеристик представлених у ньому термінів, потім природною мовою створюється список точних визначень. Після цього на основі таксономічних відносин будуються дерева класифікації понять.

Для фіксації значущих відносин між термінами виділяються основні зв'язки між ними, які можна графічно відобразити за допомогою діаграм бінарних відносин. Такі зв'язки надалі можуть стати основою для інтеграції різних онтологій.

Опис ресурсу мережі містить посилання на ресурси, опис сторінок, сайтів і зв'язків між ними [1]. Для цього в онтологію заносяться такі характеристики ресурсу, як назва, адреса, автор, структура, тематика, мова, тип інформації (текст, графіка, мультимедіа, гіперпосилання, дата відновлення, кількість відвідувачів). Важливою характеристикою ресурсу є його ступінь релевантності даному поняттю ПО. Вона може бути представлена деякою числовою величиною від 0 до 1, що визначає на діаграмі вагу дуги, яка, у свою чергу, пов'язує дане поняття з ресурсом.

Таким чином, в онтології стає можливе описування таких властивостей понять, як «посилання на релевантний ресурс у мережі Інтернет» із вказівкою значення атрибута «ступінь релевантності». З кожним ресурсом можуть бути пов'язані й інші властивості, корисні під час пошуку інформації, наприклад «посилання іншого ресурсу на ресурс» тощо.

Розглянемо кілька видів пошуку:

- локальний (коли пошук інформації проводиться безпосередньо серед посилань на ресурси, які є в онтології);

- глобальний (коли запит перед передачею тій або іншій пошуковій системі попередньо уточнюється за допомогою онтології й перекладається у відповідний формат з використанням ключових слів (термінів) і логічних зв'язків) [6].

Вибір виду пошуку здійснюється в користувацькому інтерфейсі веб-порталу, який забезпечує також можливість складання пошукового запиту природною мовою й перегляду результатів пошуку. Тому що при глобальному звертанні до пошукової системи користувач повинен мати можливість одержати у відповідь ресурси, релевантні змісту запиту, та їх пошук повинен бути семантично орієнтованим. Для цього засобу пошуку відповідної до запиту інформації пропонується організувати на основі онтології, що містить опис семантики ресурсів.

Відомо, що семантика інтернет-ресурсів дуже різноманітна, отже, здійснювати пошук необхідної інформації простіше, коли наявна специфічна предметна область. Унаслідок цього на практиці можна обмежитися побудовою онтології однієї конкретної області.

Як уже було сказано вище, для побудови онтології потрібне формальне декларативне представлення чітко організованих конструкцій, які б містили в собі словник термінів тематичної області, опис визначень цих термінів та існуючі взаємозв'язки між ними, їх теоретично

можливі й неможливі взаємозв'язки. Описані в такий спосіб онтології пропонуються застосувати як посередника між користувачем і пошуковою системою. Взаємодія з онтологією передбачається на таких етапах:

- 1) формування пошукового образу релевантного документа;
- 2) побудова запиту до пошукової системи;
- 3) формування списку релевантних документів.

Проблема полягає в тому, щоб зробити пошук динамічним і зручним для користувача. Для будь-якого типу запиту, що виникає в людини в практичній діяльності, повинні бути знайдені адекватні знання в інформаційному просторі Інтернету. При цьому мова для формулювання пошукової вимоги не повинна бути занадто складною. Зокрема, спілкування користувача з пошуковою системою можна зробити більш простим, наблизивши мову запиту до природної мови.

При такій організації пошуку на етапі формування образу релевантного документа з користувацького запиту виділяються значеннєві структури: значущі слова й терміни предметної області. Ці значеннєві структури потім використовуються для формування пошукового образу із застосуванням евристичних правил і висновку на онтології. Образ релевантного документа є описом бажаного результату роботи пошукової системи, який містить у собі:

- 1) набір термінів, які повинні міститися в документі;
- 2) набір характеристик документа;
- 3) набір вимог до результату пошукової системи, таких як кількість документів тощо.

На етапі побудови запиту до пошукової системи здійснюється висновок на онтології. При цьому виконується перетворення користувацького запиту в з'єднаний логічними зв'язуваннями набір термінів і понять, які будуть використовуватися пошуковою системою [8].

При локальному пошуку спочатку заданий природною мовою пошуковий запит обробляється системою, яка здійснює лексичний аналіз за допомогою зіставлення тексту запиту з безліччю лексичних зразків-шаблонів, заданих у тезаурусі. Потім отриманий у результаті лексичного аналізу набір термінів (понять) предметної області зазнає семантичного аналізу, який виконується на основі онтології, що становить зі словником-тезаурусом єдину мережу знань. При цьому розглядаються як властивості понять предметної області й зв'язки між ними, так і властивості представлених в онтології ресурсів, зіставлених зазначеним у запиті поняттям предметної області. Кінцевий список посилань на ресурси формується з урахуванням ступеня релевантності кожного ресурсу отриманої сукупності понять.

Отже, запропонований підхід дає змогу підвищити ефективність пошуку в мережі Інтернет шляхом організації веб-порталів, кожний з яких надає користувачеві доступ до ресурсів певної тематики. В основі таких порталів знань лежать онтології, що містять поряд з традиційним описом предметної області опис структури й типології відповідних мережевих ресурсів. Головною перевагою такого підходу є те, що портали знань дають змогу значно скоротити час обробки запиту користувача й кількість виданих ресурсів за рахунок більш точного визначення ступеня їх релевантності й зберігання посилань на них безпосередньо на порталі знань.

Список використаних джерел

1. Онтология (информатика) [Электронный ресурс] // Энциклопедия «Википедия». – Режим доступа: [http://ru.wikipedia.org/wiki/Онтология_\(информатика\)](http://ru.wikipedia.org/wiki/Онтология_(информатика)). – Загл. с экрана.
2. KBSI. Knowledge based systems incorporated, Technical Report IDEF5 Method Report [Electronic resource] / HRGA, Wright-Patterson Air Force Base, Ohio, September 1994. – Mode of access: <http://www.idef.com/pdf/Idef5.pdf>. – Title from the screen.
3. Gruber T. R. A translation approach to portable ontologies [Electronic resource] / T. R. Gruber // Knowledge Acquisition. 1993. – № 5 (2). – С. 199–220. – Mode of access: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>. – Title from the screen.
4. Gruber T. R. Toward Principles for the Design of Ontologies / T. R. Gruber // International Workshop on Formal Ontology, March, 1993, Padova. Italy.
5. Ясницкий Л. Н. Введение в искусственный интеллект / Л. Н. Ясницкий. – М. : Академия, 2005. – 176 с.
6. Клещев А. С. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия «онтология» / А. С. Клещев // Науч.-техн. информация. Серия 2 : «Информационные процессы и системы». – 2001. – № 2. – С. 20–27.
7. Гаврилова Т. А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб. : Питер, 2001. – 384 с.
8. Клещев А. С. Математические модели онтологии предметной области. Часть 2. Компоненты модели / А. С. Клещев // Науч.-техн. Информация. Серия 2 : «Информационные процессы и системы». – 2001. – № 3. – С. 19–28.