

ОНТОЛОГІЇ, ЙМОВІРНІСНІ ТЕМАТИЧНІ МОДЕЛІ ТА ТЕМАТИЧНІ КАРТИ У ОНТОЛОГО-КЕРОВАНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

У статті розглянуто проблему представлення знань у онтолого-керованих інформаційних системах та підхід до її вирішення. Проаналізовано категоризацію та роль різних категорій онтологій в інформаційних системах. Запропоновано підхід до представлення онтологій у вигляді тематичних карт, що будуються за даними, отриманими машинним навчанням за допомогою ймовірнісних тематичних моделей.

Ключові слова: ймовірнісна тематична модель, онтологія, представлення знань, тематична карта.

Вступ

Із зростанням обсягів інформації та збільшення інтенсивності інформаційних потоків дедалі гостріше постає проблема створення інформаційних систем (ІС), здатних опрацьовувати ці інформаційні потоки. В контексті створення подібних систем особливе місце посідають інтелектуальні технології, які дозволяють змінити звичні методи та засоби опрацювання інформації. Не зважаючи на те, що штучний інтелект як самостійний міждисциплінарний науковий напрям з'явився ще у 50-х роках ХХ ст., тільки сьогодні інтелектуальні системи та технології справді надійшли в масове використання і набули широкого розповсюдження в усіх сферах діяльності. Разом із цим розповсюдженням розширюється і коло задач, у яких використання інтелектуальних систем і технологій є необхідним. Окрім таких грандіозних проектів як, наприклад, система Watson компанії IBM чи самокеровані автомобілі компанії Google, є безліч інших проектів і систем, що потребують формального подання знань з різних предметних областей і засобів оперування цими знаннями.

Подання деякої предметної області з реальності в інформаційних системах засновано на певній формі представлення знань. Найчастіше у якості такої форми застосовується онтологічне представлення. Онтологічне подання описує семантичні залежності між поняттями предметної області і відповідає семантичній формі подання реальності, зокрема для інтелектуальних систем довільного призначення. Онтологія у філософії – наука про буття. В інформатиці онтологія – явна специфікація концептуалізації. Концептуалізація – спрощене абстрактне подання

реальності, яку ми хочемо відобразити для певних потреб [19].

Питаннями концептуалізації та формалізації знань займається інженерія знань (knowledge engineering), яка є окремим міждисциплінарним напрямом на перетині філософії, лінгвістики, математики, психології. Проте, попри активні дослідження в галузі інженерії знань, центральна проблема – створення універсальних методів та засобів перетворення інформації на знання – досі залишається відкритою.

Онтології в інформаційних системах. Проблема онтологічної інженерії

Онтології є однією із форм подання знань у ІС. Саме поняття «онтологія» є дещо абстрактним і строго не визначено; натомість, поняття лише окреслює загальну концепцію формалізації предметних областей для застосування у ІС, для використання як людьми, так і комп'ютерними системами у складі ІС.

У [5] онтологія визначена як:

- специфікація концептів природною або формальною мовою;
- відношення між концептами;
- правила застосування відношень до концептів;
- інтерпретація кожної специфікації концепту у відповідне поняття предметної області.

Зазначимо, що під таке визначення, як і під означення онтології з [19], підпадає, наприклад тезаурус. Загалом, різні форми онтологій можуть різнитися своєю виразністю та ступенем деталізації, при цьому відповідати концептуальному значенню онтології взагалі. За Корчо та ін. [16],

спектр виразності онтологій має вигляд, наведений на рис. 1. Виділяють дві основні групи – це легкі та важкі онтології, які поділяються на вісім підкатегорій:

- список термів є список ключових слів, який зазвичай використовується для обмеження значень певних властивостей;
- тезаурус визначає відношення між термами;
- неформальна таксономія визначає явну ієрархію узагальнення та спеціалізації, проте не підтримує суворе наслідування;
- формальна таксономія забезпечує суворе наслідування;
- онтологія на основі фреймів або класів/відношень подібна до об'єктно-орієнтованих моделей: клас визначається за місцем в ієрархії та властивостями, які наслідувані з підкласів та реалізовані у сутностях;

За роллю у кінцевій ІС онтології можна поділити на два великі класи: онтології, що є самостійними базами знань (наприклад, як центральні компоненти експертних систем, або як цілком автономні бази знань (БЗ) для обслуговування зовнішніх запитів); та онтології у онтолого-керованих системах (ОКІС) [6; 7; 24], де кінцева функціональність системи не пов'язана безпосередньо із онтологією і представленими у ній знаннями, а лише скеровується онтологією. Прикладом ІС другого класу може бути онтолого-керована інформаційно-пошукова система, у якій кінцевого користувача не цікавить внутрішній механізм поведінки системи, а лише вирішення задачі пошуку релевантної інформації. Проте система здійснює пошук саме на основі БЗ, що представлена внутрішньою онтологією.

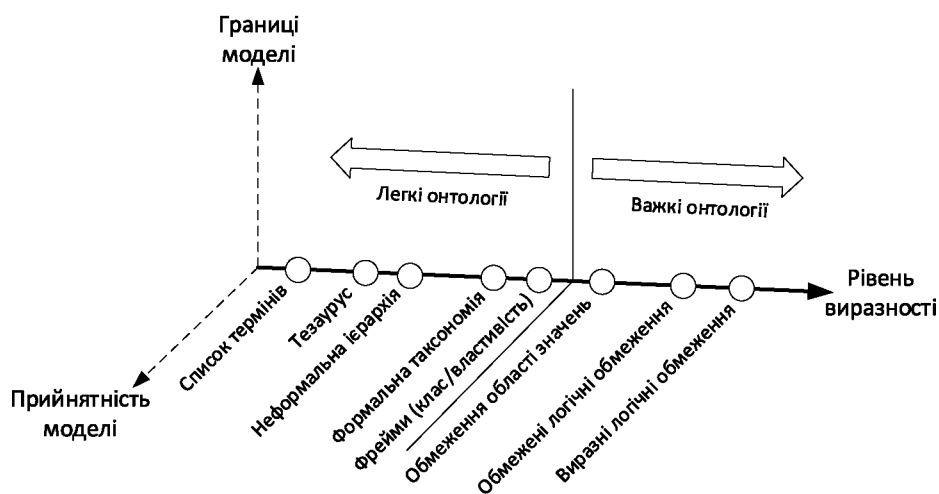


Рис. 1. Категорії онтологій

- обмеження області значень набуває сили для властивостей. Обмеження можуть бути за типом даних або за предметною областю;
- за допомогою логічних умов значення властивостей можна ще більше обмежувати;
- онтології з найбільшою виразністю часто використовують обмеження першого ступеня: відношення ціле-частина, інверсні відношення, диз'юнктивне покриття тощо.

Проблема онтологічної інженерії, по суті, полягає у стандартизації методологій та автоматизації процесів будівництва онтологій, що є метою будь-якої інженерної дисципліни. У ширшому контексті онтологічна інженерія тією чи іншою мірою містить різні теми – дизайн, розробку, застосування, поширення та повторне використання знань і базові питання – як це проілюстровано у [17].

Онтології як самостійні бази знань. Для онтологій першого класу (де онтології – автономні БЗ) характерний сценарій використання, коли користувач (людина чи комп'ютерна система) виконує запити безпосередньо до онтології або через спеціальний проміжний рівень із метою обстежувати предметну область за її поданням у онтології. Онтологія цікавить користувача сама по собі. За таких умов повнота та виразність онтології відіграє ключову роль. Від точності понятійної моделі, ієрархії відношень та правил виводу напряму залежить якість видобутих із онтології знань. На рис. 1 таким онтологіям відповідають ті, що розташовуються у напрямі важких онтологій (на малюнку – справа від умовної роздільної лінії). Наприклад, експертні системи [14; 21; 25] будуються на правилах та механізмах логічного виводу, із виразними логічними обмеженнями і досить складною ієрархією відношень між концептами. Від повноти

та виразності онтології безпосередньо залежить, наскільки машина виводу експертної системи буде здатна повертати точні та коректні результати.

Онтології у ОКІС. Онтолого-керовані системи є різноманітнішими за своїм призначенням та ключовими принципами внутрішньої будови, ніж БЗ, засновані на онтологіях. Роль онтологій у таких системах різниться від системи до системи, а відтак різниться значущість тих чи інших характеристик онтології для кінцевого результату роботи системи. Варто відзначити, що автономні знання-орієнтовані системи, як, наприклад, експертні, здебільшого використовуються у контекстах окремих предметних областей. У більшості випадків створення специфічної онтології під конкретну предметну область є виправданішим і доречнішим, оскільки дозволяє оптимально використати ресурси на розробку і досягти рівня виразності, необхідного для ефективного використання експертних систем у певній області.

Натомість ОКІС, як, наприклад, онтолого-керовані інформаційно-пошукові системи, найчастіше затребувані там, де є перетин різних предметних областей або створення предметно-специфічних систем не є виправданим. Наприклад, створення пошукової системи для великих репозиторіїв вимагає розробки єдиної інтегрованої системи, яка була б здатною оперувати довільною предметною областю із представлених у колекції документів. Або інженерія продуктової лінійки (software product line) онтологічно-керованих систем, що функціонують за спільними принципами і різняться тільки предметною областю з поправкою на окремі компоненти, які є вільно замінюваними для кожної системи в лінійці. У таких продуктових лінійках теж є неможливим опортуністський (ad-hoc) підхід у створенні кожного нового продукту. Натомість доречним є знаходження потрібного балансу між виразністю онтології та зусиллями на її побудову, в ідеальному випадку – балансу між якістю кінцевого результату та автономністю системи (наскільки ОКІС здатна розбудовувати онтологію самостійно, з мінімальним втручанням експерта або без нього). Зазначимо, що питання автоматизації розбудови онтологій та формальні методи оцінки зазначеного балансу виходять за рамки цієї роботи. Зокрема, щодо питання про автоматизацію методів та засобів онтологічної інженерії можна звернутися, наприклад, до вітчизняних [3; 2; 1; 4] і закордонних [18; 22; 27; 28] праць.

Розглянемо один підхід до можливого балансу між виразністю онтології та автономністю ОКІС на основі ймовірнісних тематичних моделей та тематичних карт у контексті ОКІС.

Ймовірнісні тематичні моделі

Ймовірнісні тематичні моделі (ТМ) – це алгоритми для виявлення тематичного наповнення документів у великих неструктурованих колекціях [11]. Найпростішою тематичною моделлю є приховане розміщення Діріхле (latent Dirichlet allocation) [13]. Зазначимо, що всі тематичні моделі відносяться до класу породжувальних (generative models), тобто є моделями, що генерують спостережувані дані за деякими прихованими параметрами. За цим принципом тематичні моделі є засобами машинного навчання. Окремо варто наголосити на тому, що ідею про породжувальну модель інформації висловив Клод Шеннон у 1948 р. у праці «Математична теорія зв'язку» [26]. Цікаво, що, окрім математичного підґрунтя породжувальних моделей, у контексті інформації природною мовою породжувальне моделювання також підкріплюється ідеями з напряму генеративної лінгвістики, який виник у той самий час (1950-ті рр.) завдяки Ноаму Хомському, філософу та лінгвісту, автору ієрархії формальних граматики.

Породжувальний процес у ТМ зображений на рис. 2. Слід окремо наголосити, що у загальному випадку не варто ототожнювати



Рис. 2. Породжувальний процес у тематичних моделях

тему у ТМ із темою у загальноприйнятому розумінні – як тему статті чи розмови тощо. Тема у ТМ є умовною одиницею розбиття усього словника на групи, в рамках яких є ймовірнісні розподіли на словах, що з певною ймовірністю зустрінуться у документі. Оскільки ТМ є суто математичними алгоритмами, утворені теми не можна трактувати як тезауруси предметних областей. Натомість слід розглядати ці теми як набори слів, з яких автор утворює документ, ніби обираючи слово за словом з кожного набору.

LDA й інші тематичні моделі є ймовірнісними моделями, які трактують дані як результат перебігу певного породжувального процесу, що визначає сумісний розподіл спостережуваних та прихованих величин. Для тематичних моделей проблема полягає в обчисленні апостеріорного розподілу – умовного розподілу прихованих величин (пропорції тем) при спостережуваних (документи). Формально генеративний процес LDA подається як сумісний розподіл спостережуваних та прихованих величин:

$$p(\beta_{1:K}, \vartheta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\vartheta_d)$$

$$\left(\prod_{n=1}^N p(z_{d,n} | \vartheta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right) \quad (1)$$

$\beta_{1:K}$ – теми, розподіли на словнику

$\vartheta_{1:D}$ – пропорції тем у документах

$z_{1:D}$ – співставлення кожного слова у документі із темою

$w_{1:D}$ – слова у документах

Згаданий вище апостеріорний розподіл прихованих величин при спостережуваній величині (слова у документах) має вигляд:

$$p(\beta_{1:K}, \vartheta_{1:D}, z_{1:D} | w_{1:D}) = \frac{p(\beta_{1:K}, \vartheta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad (2)$$

Знаменник дробу трактується як ймовірність отримати спостережуваний корпус при кожній можливій моделі тематичної структури корпусу. Обчислення цієї ймовірності означає розрахунок сумісного розподілу по усіх можливих варіантах тематичної структури, що є дуже складно. Через знаменник обчислити апостеріорний розподіл неможливо, і тому тематичні моделі націлені на знаходження ефективних апроксимацій. Двома основними методами апроксимації в тематичних моделях є семплінг за Гібсом та варіаційні алгоритми [11; 13].

Тематичні карти

Тематичні карти [8; 29] є однією з форм представлення онтологій. Тематична карта (ТК) виражає онтологію у вигляді концептів (тем) та відношень (асоціацій) між ними. Ієрархія тем може бути як завгодно складною. Тематичні карти багато в чому подібні до стандарту W3C RDF [9; 10], хоча останні орієнтовано на ресурси, а не на теми як такі. Онтологія в OKIC може мати будь-яку форму і походження, але ми зосередимось на тематичних картах з двох причин: по-перше, вони дозволяють інтуїтивно моделювати зміст документів, інтерпретуючи їх як розкриття певних тем у деяких пропорціях; по-друге, тематичні карти придатні до автоматизованої побудови з мінімальною участю людини-експерта.

Тематичні карти складаються з двох просторів: простору тем та простору ресурсів. Зв'язки ТК поєднують теми в ієрархію та теми з ресурсами. Простори є потенційно незалежними один від одного, що в перспективі дозволяє використовувати існуючі ТК на нових колекціях документів. Простір тем у ТК представляє модель предметної області. ТК є високорівневою моделлю знань з предметної області, що представлені документами (ресурсами). ТК за виразністю тяжіють до легких онтологій, проте дозволяють досліджувати тематичну ієрархію довільних колекцій ресурсів і видобувати інформацію, рухаючись цією ієрархією у різних напрямках (від загальних тем до конкретніших чи навпаки, між спорідненими темами одного підрівня тощо).

Тема у ТК є машинно-читабельним представленням певного концепту. Не існує жодних обмежень на природу концептів, що можуть позначатися темами, проте є чотири основні форми ідентифікації тем у межах ТК:

1. Ідентифікатор теми як ресурсу у серіалізованій ТК: таким ідентифікатором виступає URI (Uniform Resource Identifier за стандартом RFC 3986) і є унікальним у межах ТК.
2. Ідентифікатор теми у вигляді людино-читабельного ярлика: тема може мати довільну кількість назв, доступних людині для розуміння.
3. Ідентифікація за посиланням: для ресурсів, що мають власний URI ідентифікатор пов'язаної теми, є похідним від ідентифікатора ресурсу, що дозволяє означувати кожну тему за відповідним їй ресурсом.
4. Ідентифікатор за описом: деякі теми можуть позначати концепти, що не є ідентифікованими за URI (наприклад, люди), але асоціюються із певними описовими

сутностями (реєстраційні картки, фотографії, анкети тощо), а відтак ідентифікуються цими сукупностями описової інформації.

Важливо зауважити, що хоча тема може мати довільну кількість ідентифікаторів, кожний окремий ідентифікатор повинен однозначно вказувати на конкретну тему.

На користь тематичних карт зазначимо, що тематичні карти забезпечено стандартизованим XML синтаксисом [30] і специфікацією прикладного програмного інтерфейсу (API) з різноманітними імплементаціями (Common Topic Map Application Programming Interface). Також тематичні карти стандартизовані за ідентифікатором ISO/IEC 13250.

Моделювання онтологій як тематичних карт засобами тематичних моделей

Модель LDA є найпростішою тематичною моделлю. Як одна з перших ймовірнісних моделей для текстів, LDA базується на ряді припущень, що практично унеможливають її використання у розроблюваних ОКІС. Модель спирається на подання текстів як «мішка слів» («bag of words»), тобто ігнорує порядок та зв'язок слів у документах. Окрім того, породжувальний процес моделі передбачає лише однорівневе моделювання тем без врахування ієрархічних зв'язків між темами, що є недоречним у практичних застосуваннях. Натомість у контексті побудови ієрархічних тематичних карт для ОКІС цікавими є корельована тематична модель [12] і модель розміщення патінко (pachinko) [23]. Моделі описують значно складніший породжувальний процес, що покликаний відтворити значно природніший стан речей у тематичній ієрархії, тобто врахувати, що окремі теми можуть бути підтемами інших. Як наслідок, зазначені моделі можна використати для навчання на еталонних колекціях документів, а навчені моделі застосувати для виведення (inference – процес, коли навчена модель опрацьовує довільні колекції ресурсів) тематичних структур (ієрархій) на кінцевих репозиторіях ресурсів. Внаслідок такого виведення для кожного ресурсу утворюється набір асоціацій із певними темами. З отриманих даних будується тематична карта, яка об'єднує одночасно інформацію щодо розподілу на темах для кожного ресурсу, розподіли на словнику для кожної теми (якщо це текстовий ресурс), розподіли на темах для відношень «супер-тема – підтема». Приклад такої структури зображено на рис. 3.

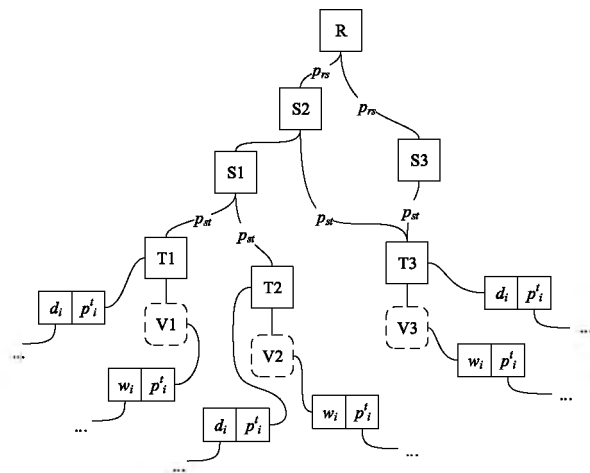


Рис. 3. Структура тематичної карти. S , T – супер-теми, теми; V – розподіли на словнику, де w_i – слово, p'_i – ймовірність слова i для теми t ; темі відповідає список ID документів d_i із ймовірностями p'_i з розподілу на темах (ймовірність теми t для документу i); p_{si} та p_{pi} – розподіли на підтемах

З практичної точки зору структура має такий характер. Якщо обмежитися розглядом лише текстових ресурсів, дерево тем, яке є аналогією словника у класичному текстовому індексі, значно компактніше. Як відомо, закон Хіпса [20] дозволяє приблизно оцінити кількість термінів за кількістю лексем, і для мільйона лексем дає оцінку в ~39 тис. термінів. Кількість тем, навпаки, не зростає настільки швидко з розміром колекції і не перевищує кількох сотень для колекції в 20 000 документів (точні значення залежать від колекції та задачі). Важливо, що за таких умов тематичною картою можна вільно оперувати в оперативній пам'яті. Що ж стосується нашого обмеження текстовою інформацією, то це є лише питання вибору відповідних ТМ (розробки або адаптації необхідних під конкретний тип даних і задачу), різні типи яких проілюстровано, зокрема, оглядово у [11].

Висновки

У роботі ми розглянули можливість моделювання онтологій у онтолого-керованих системах за допомогою ймовірнісних тематичних моделей. Варто зауважити, що доречність використання описаного підходу щодо моделювання онтологій як тематичних карт засобами ймовірнісних тематичних моделей у кожному окремому випадку із розроблюваною ОКІС вимагає більш строгого, формального обґрунтування. Вочевидь є деякий клас ОКІС, у яких виразність та повнота онтологій не є основним критерієм, або навіть є недоречною через складність автоматизованої розбудови і подальшого використання. Натомість, зміщення у бік

легких онтологій, як-то тематичні карти, дозволяє вжити більше інженерний, ніж опортуністський підхід. У подальшому варто приділити увагу дослідженню питання ефективності використання

тематичних карт у певних класах ОКІС, особливо там, де першочерговими є задачі опрацювання неструктурованої гетерогенної інформації, як, наприклад, автономні пошукові системи.

Список літератури

1. Розробка методів автоматизованого розширення та добудови онтологічних баз знань / А. В. Анісімов, М. М. Глибовець, П. П. Кулябо та ін. // Наукові записки НАУКМА. – 2009. – Т. 99. – С. 50–53.
2. Глибовець Н. Н. Применение онтологий и методов текстового анализа при создании интеллектуальных поисковых систем / Н. Н. Глибовець, А. Н. Глибовець, А. С. Шабинский // Проблемы управления и информатики. – 2011. – № 6. – С. 96–102.
3. Бобко О. О. Методи автоматичної генерації онтологій / О. О. Бобко, М. М. Глибовець // Наукові записки НАУКМА. – 2012. – Т. 138. – С. 60–67.
4. Литвин В. В. Автоматизація процесу розвитку базової онтології на основі аналізу текстових ресурсів / В. В. Литвин // Вісник Національного університету «Львівська політехніка». – 2010. – Т. 673. – С. 319–324.
5. Мейтус В. Ю. Інтелектуальні системи, онтології та онтологічні простори / В. Ю. Мейтус // Наукові записки НАУКМА. – 2009. – Т. 99. – С. 4–14.
6. Палагін А. В. К проектированию онтологоуправляемой информационной системы с обработкой естественно-языковых объектов / А. В. Палагін, Н. Г. Петренко // Математичні машини і системи. – 2008. – № 2. – С. 14–23.
7. Палагін О. В. Архітектурно-онтологічні принципи розбудови інтелектуальних інформаційних систем / О. В. Палагін, М. Г. Петренко // Математичні машини і системи. – 2006. – № 4. – С. 15–20.
8. Ahmed K. An Introduction to Topic Maps / K. Ahmed, G. Moore // The Architecture Journal. – 2005. – № 5. – P. 3–9.
9. Beckett D. RDF/XML Syntax Specification [Електронний ресурс] / D. Beckett, B. McBride. – 2004. – Режим доступу: <http://www.w3.org/TR/REC-rdf-syntax>
10. Berners-Lee T. Notation3 (N3) : A readable RDF syntax [Електронний ресурс] / T. Berners-Lee. – 2006. – Режим доступу: <http://www.w3.org/DesignIssues/Notation3>
11. Blei D. Probabilistic Topic Models / D. Blei // Communications of the ACM. – 2012. – Vol. 55. – № 4. – P. 77–84.
12. Blei D. A correlated topic model of SCIENCE / D. Blei, J. Lafferty // The Annals of Applied Statistics. – 2007. – Vol. 1. – № 1. – P. 17–35.
13. Blei D. Latent Dirichlet Allocation / D. Blei, A. Ng, M. Jordan // Journal of Machine Learning Research. – 2003. – № 3. – P. 993–1022.
14. Buchanan B. Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project / Bruce Buchanan, Edward Shortliffe. – Addison-Wesley Longman Publishing Co., Inc., 1984.
15. Common Topic Map Application Programming Interface [Електронний ресурс]. – Режим доступу: <http://www.tnapi.org>
16. Corcho O. Methodologies, tools and languages for building ontologies. Where is their meeting point? / O. Corcho, M. Fernandez-Lopez, A. Gomez-Perez // Data & Knowledge Engineering. – 2003. – № 46. – P. 41–64.
17. Devedzic V. Understanding Ontological Engineering / V. Devedzic // Communication of the ACM. – 2002. – Vol. 45. – № 4. – P. 136–144.
18. Fortuna B. System for Semi-automatic Ontology construction / B. Fortuna, M. Grobelnik, D. Mladenic // Proceedings of the 3rd European Semantic Web Conference ESWC-2006. – 2006. – Budva, Montenegro.
19. Gruber T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing / T. Gruber // International Journal Human-Computer Studies. – 1993. – Vol. 43. – № 5. – P. 907–928.
20. Heaps H. Information Retrieval: Computational and Theoretical Aspects / Harold Heaps. – Academic Press, 1978.
21. Jackson P. Introduction to expert systems / Peter Jackson. – Addison-Wesley Longman Publishing Co., Inc., 1990.
22. Jiang X. Mining Ontological Knowledge from Domain-Specific Text Documents / X. Jiang, A. Tan // Fifth IEEE International Conference on Data Mining (ICDM'05); 2005; Houston, Texas.
23. Li W. Pachinko Allocation: DAG-Structured Mixture Models of Topic Correlations / W. Li, A. McCallum // Proceedings of the 23rd International Conference on Machine Learning. – Pittsburg, 2006.
24. Qiu R. Towards ontology-driven knowledge synthesis for heterogeneous information systems / R. Qiu // Journal of Intelligent Manufacturing. – 2006. – Vol. 17. – № 1. – P. 99–109.
25. Russell S. Artificial Intelligence: A Modern Approach / Stewart Russell, Peter Norvig. – Prentice Hall, 2003.
26. Shannon C. A Mathematical Theory of Communication / C. Shannon // The Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423, 623–656.
27. Stavrianou A. Overview and semantic issues of text mining / A. Stavrianou, P. Andritsos, N. Nicoloyannis // ACM SIGMOD Record. – 2007. – Vol. 36. – № 3. – P. 23–34.
28. Wong W. Ontology Learning from Text: A Look Back and into the Future / W. Wong, W. Liu, M. Bannamoun // ACM Computing Surveys. – 2012. – Vol. 44. – № 4. – article.no.20.
29. ISO/IEC JTC1/SC34/WG3. Topic Maps – Part 1: Overview and Basic Concepts [Електронний ресурс]. – Режим доступу: <http://www.itscj.ipsj.or.jp/sc34/open/1045.htm>.
30. ISO/IEC JTC1/SC34/WG3. Topic Maps – XML Syntax [Електронний ресурс]. – Режим доступу: <http://www.isotopicmaps.org/sam/sam-xtm/>.

A. Shabinskiy

ONTOLOGIES, PROBABILISTIC TOPIC MODELS, AND TOPIC MAPS IN ONTOLOGY-DRIVEN INFORMATION SYSTEMS

The paper considers the problem of knowledge representation in ontology-driven information systems and an approach to solving it. Categories of ontologies and their respective roles in different information systems is analyzed. An approach of representing ontologies as topic maps is proposed, based on the data learned from raw resources with topic models.

Keywords: probabilistic topic model, ontology, knowledge representation, topic map.

Матеріал надійшов 15.09.2013