

Нестеренко О.В., к.т.н., доц., провідний науковий співробітник
Зайцев С.М., к.ф.-м.н., провідний науковий співробітник
Артеменко О.Л., провідний науковий співробітник
 Державне підприємство «Український науковий центр розвитку інформаційних технологій», м. Київ, Україна

УСЕОХОПЛЮЮЧИЙ ІНТЕРНЕТ, ВЕЛИКІ ДАНІ І КОМП'ЮТЕРНІ ОНТОЛОГІЇ

Глобалізація у сучасному світі - це використання для аналітики і прийняття ефективних рішень значних, вельми значних об'ємів даних. Ідеї, що з'являються в результаті проведення аналізу даних, використовуються для поліпшення клієнтського обслуговування, зниження витрат або навіть для запровадження нових бізнес-моделей, які «висаджують» галузі і витісняють конкуренцію. Але більш ніж будь-коли дані, на яких тримається це розуміння, стрімко множаться та поширюються від постійно зростаючої кількості джерел.

Навіть поверховий огляд різних видів і джерел даних в сучасній економіці малює картину виключно диференційованого ландшафту даних. Почнемо з традиційних джерел даних, що відносяться до систем управління підприємствами - системи ERP, CRM, керування ланцюжками постачань (*Supply Chain*), управління життєвим циклом продукту (PLM), колл-центри обслуговування клієнтів і т.ін. Додамо електронну пошту, статистику сайту компанії і матеріали електронної комерції. У підсумку маємо поєднання значної кількості структурованих і неструктурованих даних. Далі включимо сюди дані з нових каналів взаємодії з клієнтами - платформи продуктів, мобільні застосування, соціальні мережі.

Але все це дрібниці коли мова заходить про Інтернет Речей (*Internet of Things, IoT*) та Всеохоплюючий Інтернет, або Всеосяжний Інтернет (*Internet of Everything, IoE*). Якщо компанія почала використовувати датчики у виробничому устаткуванні, носимому/возимому приладді або у товарах на полицях - вважайте це ще одним важливим джерелом даних для управління. Але залишається одна проблема – як опанувати цей величезний об'єм даних?

Коли з'являється багато даних, по суті, старі способи управління ними для аналітики і зберігання вже не підходять. Тому ще наприкінці 90 років у обіг був уведений термін Великі дані (*Big Data*), під яким на сьогодні розуміються інтелектуальні технології аналізу значних об'ємів даних з різних джерел (бажано з усіх можливих у предметній області) з великою швидкістю і в режимі реального часу.

Стратегія «цифрового перетворення» (*Digital Transformation*) як нового ядра для тих, хто прагне конкурувати в цифровій економіці, надає можливостей значній частині великих світових компаній вже сьогодні отримувати дохід від послуг, що надаються на основі накопичених даних (*data-as-a-service*), таких як продаж необроблених даних, метрик, виявленого розуміння (знання) та рекомендацій, отриманих на основі технологій бізнес-аналітики (*Business Intelligence, BI*) та інтелектуального аналізу даних (*Data mining*).

Технології Big Data дійсно дозволяють знаходити всілякі кореляції в будь-яких даних. Проте, в багатьох випадках відповідь може бути не настільки очевидною – чи спостерігаємо ми причинно-наслідковий зв'язок, чи випадковий збіг, що є суттєвою системною проблемою.

На цей час існують певні технології, зокрема когнітивні семантичні технології і штучний інтелект (машинне навчання), що допомагають вирішувати проблеми аналізу великих даних. На підході й нові технологічні інструменти управління даними. Візьмемо, наприклад, обробку даних в пам'яті (*in-memory*) – швидкість обробки приголомшує. Гарною новиною є й те, що існують хмарні обчислення (*cloud computing*).

Однак наявність цього потужного інструментарію все ще не дозволяє вирішити вищезазначену проблему. У суспільних і соціальних сферах, зокрема в економіці, в державному управлінні – там, де широке поле для Великих даних – насправді немає скільки-небудь строгої теорії структуризації понять і даних, теорії в тому сенсі, як це розуміють, наприклад, природничі науки. Адже нова революція полягає не в комп'ютерах, які обробляють дані, а в самих даних і в тому, як ми їх використовуємо.

Цінність результату аналітичної діяльності вочевидь напряму залежить від повноти обробленої інформації [1]. Об'єм інформації, що опрацьовується, зазвичай складається з трьох частин: $I_A = I_P + I_D + I_O$, де I_A - обсяг опрацьованої інформації; I_P - обсяг пертинентної інформації; I_D - обсяг додаткової інформації, яка є релевантною, але не є пертинентною; I_O -

особиста інформація, яка згенерована власне аналітиком у процесі дослідження даних за рахунок індивідуальних знань, існуючого інструментарію аналізу даних і творчого підходу до аналітичних дій.

Найважливішою умовою успішної аналітичної роботи експерта є наявність інформаційного поля досліджуваної предметної області (ПДО), яке має уявляти множину структурованих і неструктурованих інформаційних масивів (інформаційних ресурсів), необхідних для витягу з них необхідних даних. Робота аналітика починається зі збирання інформації щодо проблеми, що розглядається, з різних джерел шляхом використання інформаційно-пошукових систем. Хоча можливості сучасних пошукових систем постійно зростають, однак поки що вони, як і раніше, є ще далекими від ідеалу, тому об'єм I_d іноді може значно переважати об'єм I_n , що ускладнює роботу аналітика і знижує ефективність систем аналізу. Водночас для формування I_o необхідно провести аналіз даних, що накопичені в базах даних автоматизованих систем, які функціонують в організації. Проблема зростає у зв'язку із тим, що об'єми цих БД постійно і стрімко зростають, і вже часто-густо стають для аналітика неосяжними.

Розвиток наукових досліджень і досвідів в сферах моделювання аналітичної діяльності свідчить, що одним з інструментів, за допомогою якого можливо досить ефективно спроектувати та реалізувати механізми управління ієрархією, яка відображає взаємодію усіх інформаційних компонентів, може бути онтологічна модель [2,3]. Вона у своїй інформаційній основі має механізм динамічного формування та використання ієрархій у вигляді певних таксономій.

Предметну область (ПрО) безпосередньо складають певні концепти та їх властивості. Концепти складають кінцеву множину $X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$, а множина властивостей R утворюється множиною декартових добутків множини X самої на себе: $R = \prod_i^n X_i$. Будемо

вважати, що властивості R є інтерпретацією відношень, тобто існує перетворення, яке кожному відношенню встановлює відповідність певної властивості. Тоді множина F може бути утвореною декартовим добутком множин X і R : $F = X \times R$.

Таким чином онтологія деякого операційного середовища в загальному випадку формально представляється впорядкованою трійкою $O = (X, R, F)$.

На основі концептів X (понять, термінів) ПрО формується предметна складова операційного середовища аналізу даних. Функції інтерпретації F (визначень) X та/або R складають функціональну частину операційного середовища.

Розгляд граничних випадків множин: $R = \emptyset$; $R \neq \emptyset$; $F = \emptyset$; $F \neq \emptyset$ у всіх чотирьох комбінаціях значень R і F дає різні варіанти онтологічних конструкцій, починаючи від простого словника і таксономії до формальної структури концептуальної бази знань для високоінтелектуальних знання-орієнтованих систем.

ЛІТЕРАТУРА

1. Пархоменко Б.В. Роль творчості в процесі обґрунтування прийняття рішень / Б.В. Пархоменко // XII Международная научно-практическая конференция «Построение информационного общества: ресурсы и технологии», 6-7 июня (2007, Киев, Украина). - К.: УкрИНТЭИ, 2007. С. 106-108.
2. Емельянов С.В. Многокритериальные методы принятия решений / С.В. Емельянов, О.И. Ларичев. М.: Знание, 1985. 32 с. – (Новое в жизни, науке, технике. Сер. «Математика, кибернетика», №10).
3. Стрижак О. Є. Онтологічні інформаційно-аналітичні системи / О. Є. Стрижак // Радіоелектронні і комп'ютерні системи. – 2014. – № 3 (67). – С. 71–76.