

УДК 311.21:004

ЗАДОРОЖНА Р.П., канд. екон. наук
Білоцерківський національний аграрний університет
ruslana.zadorozhna@gmail.com

BIG DATA – НОВИЙ ІНФОРМАЦІЙНИЙ ФЕНОМЕН ЦИФРОВОЇ ЕПОХИ

Досліджено зміст та взаємозв'язки понять «дані», «інформація», «знання» в рамках концепції інформаційної ієрархії DIKW та згідно з термінологічним стандартом ISO/IEC 2382:2015. На найвищому рівні узагальнення розкрито сутність даних як зареєстрованих сигналів. Проаналізовано існуючі підходи до тлумачення поняття інформації; значення контексту та необхідність застосування відповідних методів оброблення даних для перетворення їх в інформацію. Розглянуто поділ отриманої з даних інформації на аналогову та цифрову.

Охарактеризовано сутність Big Data як різновиду даних, що виник внаслідок розвитку комп'ютерних технологій. Розглянуто основні особливості (3V) великих даних: їх обсяг (volume), швидкість (velocity) та різноманітність (variety). Окреслено потенціал інтелектуального аналізу даних (Data Mining) як засобу отримання прихованих знань зі структурованих та неструктурованих масивів великих даних.

Ключові слова: дані, інформація, знання, DIKW, Big Data, Data Mining, неструктурована інформація.

Постановка проблеми. Еволюційний процес розвитку людства, його хід і результати знаходять своє відображення в різноманітних фактах, відомостях, даних, інформації тощо. У різні історичні епохи ті відомості, що визнавалися доцільними для подальшого збереження, у доступний спосіб фіксувалися, завдяки чому могли бути використані не лише сучасниками, але й наступними поколіннями. Таким чином, у найбільш узагальненому сенсі людство постійно продукує, накопичує і використовує певний «інформаційний продукт». З плином часу та еволюційним поступом його тематика постійно диверсифікувалася, а обсяги безупинно зростали. Проте в останні десятиріччя в результаті стрімкого розвитку комп'ютерних технологій та їх проникнення чи не в усі сфери життя характер динаміки цього процесу кардинально змінився. Це знайшло відображення в тому, що сучасний етап розвитку цивілізації описується в термінах «інформаційна епоха», «інформаційне суспільство», «інформаційна економіка», «інформаційна революція» тощо [1].

Характерною особливістю сучасного «оцифрованого» життя стала поява Big Data – «великих даних», які наразі привертають до себе все більше уваги і стають предметом вивчення дедалі ширшого кола дослідників – від аналітиків даних до економістів, соціологів, маркетингологів, медиків і т.д.

На перший погляд, судячи з назви, йдеться просто про значні за розміром інформаційні масиви. Однак великий обсяг – лише одна з особливостей феномену Big Data, у якому, з одного боку, знайшли втілення комп'ютерно-інформаційні тренди останніх десятиріч; з іншого – він сам здатен впливати й реально трансформує існуючі уявлення та напрацьовані впродовж тривалого часу практики й моделі поведінки як окремих індивідів, так і складних організаційних структур.

Аналіз останніх досліджень і публікацій. Вивченню різноманітних аспектів феномену Big Data присвячені дослідження таких вітчизняних і зарубіжних науковців та практиків: Р. Акофф, О. Грібініченко, Е. Даміані, В. Дашонок, О. Журенков, П. Іванов, І. Ільїн, Р. Казаков, О. Колесниченко, К. Кук'єр, Д. Лані, В. Майер-Шенбергер, Л. Мазелис, Г. Смородин, П. Хофманн.

Мета дослідження полягає у з'ясуванні передумов виникнення, сутності та особливостей Big Data як різновиду даних.

Матеріал і методика дослідження. Для досягнення мети дослідження використано такі методи: історичний, діалектичний, абстрактно-логічний, порівняння, аналіз, узагальнення.

Основні результати дослідження. Для того, щоб з'ясувати сутність та особливості «великих даних», варто спочатку розглянути поняття власне даних і тісно пов'язане з ним поняття інформації.

Перш за все слід звернути увагу на відсутність загальноприйнятих визначень даних та інформації, що є зрозумілим з огляду на багатоваріантність форм їх існування та різні погляди

на досліджувану проблематику. При цьому поняття даних та інформації зазвичай визначаються одне через інше, що може створити труднощі стосовно однозначної їх ідентифікації та спричинити розуміння їх як взаємозамінних синонімів, що не відповідає дійсності.

Наразі досить розповсюдженою є запропонована Р. Акоффом [2] концепція інформаційної ієрархії DIKW (data – information – knowledge – wisdom) (див. рис. 1). Попри те, що ця модель не уникла критики, вона дає уявлення про першооснову виникнення знань, ілюструючи поступове ускладнення пізнавального процесу.

Як видно з рисунка 1, в основі інформаційної піраміди перебувають дані (data), котрі можуть бути трансформовані в інформацію (information), з якої, в свою чергу, можуть бути здобуті знання (knowledge) і, нарешті, останнім кроком є перетворення знань у мудрість (wisdom). Кожен вищий рівень ієрархії є більш довершеним, ніж попередній, завдяки додаванню певних властивостей до рівня нижчого порядку – дані перетворюються на інформацію завдяки додаванню контексту; знання додає опцію «як» (механізм використання); мудрість – «коли» (умови використання) [3].

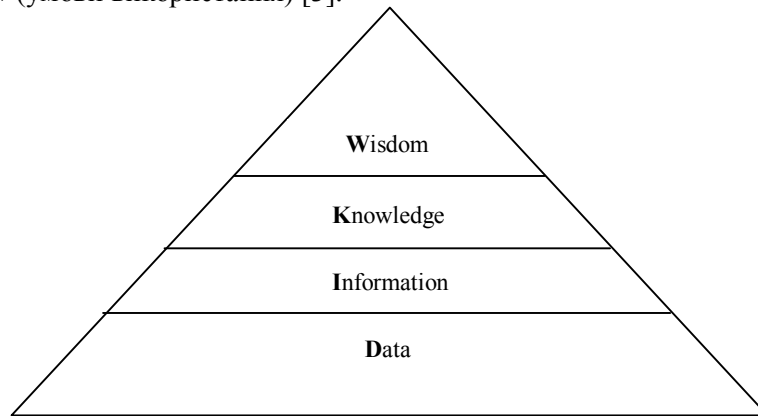


Рис. 1. Інформаційна ієрархія DIKW.

Саме поява певного контексту, у якому розуміють дані, і робить їх інформацією. Великий тлумачний словник сучасної української мови визначає контекст як «закінчений за змістом уривок тексту, що дає змогу встановити значення (виділено нами – *Авт.*) слова або речення, які входять до його складу» [4, с. 575]. Підкреслимо, що текст є лише одним з видів існування даних. Останні за своєю сутністю є «... зареєстрованими сигналами» [5, с. 21], тобто даними можуть бути будь-який символ, текст, цифри, зображення, жест, звук чи відео, фізичні методи реєстрації яких, звичайно, будуть відрізнятися. Технічні пристрої і навіть тіло людини генерують дані у вигляді різноманітних імпульсів. За збереження за допомогою комп'ютера вони представлятимуть набір чисел у вигляді байтів, котрі, у свою чергу, складаються з бітів, що набувають значення «1» чи «0». Зрозуміло, що сам факт наявності таких даних є недостатнім для їх ефективного використання, оскільки спочатку слід досягнути, що ж означають ці дані, тобто яку інформацію вони в собі несуть. Для цього потрібно застосувати відповідні методи обробки даних.

Усі можливі методи обробки даних поділяються на природні й технічні. Природні – це методи, засновані на органах чуттів людини (зір, дотик, нюх, слух, смак, спостереження, читання, логічне мислення, уява, порівняння, аналіз тощо); технічні методи включають апаратні (обробка здійснюється за допомогою спеціальних пристроїв – телефонів, магнітофонів, рентгенівських апаратів, мікроскопів тощо) та програмні (за допомогою комп'ютерного коду).

Без змістовного тлумачення (поза контекстом) дані так і залишаться просто зафіксованими сигналами. Наприклад, маючи результати лабораторного аналізу, неможливо зробити висновок, чи здоровою є людина або ж якого лікування вона потребує (для цього мають бути застосовані такі методи, як зір, читання, логічне мислення, аналіз).

Важливим для нашого дослідження є поділ отриманої з даних інформації на аналогову та цифрову. Перша отримується завдяки застосуванню природних методів обробки даних і

сприймається виключно людиною. Вона має неперервний характер і поділяється відповідно до п'яти органів людського чуття на візуальну, аудіальну, тактильну, нюхову, смакову. Цифрова інформація є дискретною і сприймається обчислювальною технікою.

Таким чином, інформація – це «... смисловий продукт взаємодії даних та адекватних їм методів» [5, с. 16]. З одних і тих же даних, застосовуючи різні методи обробки, можна отримати різну інформацію. Або ж, як зауважує Н. В. Ковтун, якщо дані вже відомі досліднику або не належать до розв'язуваної задачі, вони не містять для нього інформації [6, с. 36].

Термінологічний стандарт ISO/IEC 2382-1 «Information technology – Vocabulary» (перегляд ISO/IEC 2382:2015 [7]), оперуючи поняттями трьох нижчих рівнів інформаційної ієрархії DIKW, дає наступні визначення даних, інформації та знання:

- дані – це подання інформації в певному формалізованому вигляді, придатному для передачі, інтерпретації чи обробки;
- інформація (у процесах її обробки) – це будь-який факт, поняття чи значення, отримані з даних, а також контекст, вибраний зі знань, чи контекст, асоційований зі знаннями;
- знання – це організоване, інтегроване зібрання фактів та узагальнень.

Дані, інформація та знання можуть існувати як до, так і після процесу їх обробки. Як видно з рисунка 2, дані можуть бути отримані двома шляхами: або за допомогою здійснення вимірювань в дослідженнях, або ж «низхідним» шляхом (у контексті моделі DIKW), виокремивши певну інформацію з уже наявних знань і певним чином її представивши. Так, в статистичній практиці розмежовуються поняття первинних (до процедури їх обробки) та статистичних (після обробки) даних. Закон України «Про державну статистику» [8] визначає первинні дані як інформацію щодо кількісної та якісної характеристики явищ і процесів, яка подана респондентами під час статистичних спостережень. Статистичні дані згідно з цим законом – це інформація, отримана на підставі проведених статистичних спостережень, що опрацьована й подана у формалізованому вигляді відповідно до загальноприйнятих принципів і методології.

Зауважимо, що обробка знань, на відміну від обробки даних та інформації, може здійснюватися лише завдяки свідомості людини. Як вказує Л. І. Донець, зазвичай знання існують не в документах, а в індивідуальній чи колективній свідомості [9, с. 293], з чого автор робить висновок, що знання – «... це люди плюс інформація» [9, с. 295].

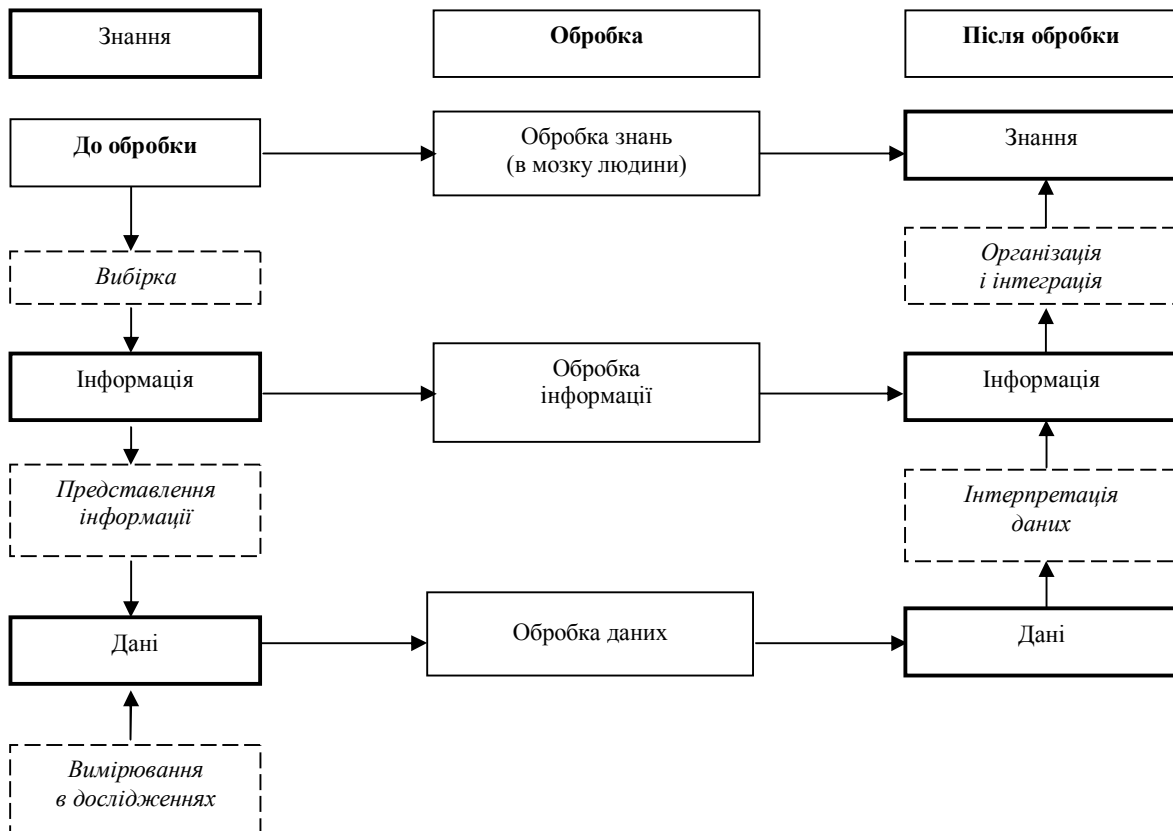


Рис. 2. Співвідношення між даними, інформацією та знаннями.

Поняття інформації як одне з найбільш універсальних у прихованій формі досліджувалося вченими починаючи з античних філософів. Проте, як зазначає Стенфордська енциклопедія філософії, до другої половини двадцятого сторіччя практично жоден із сучасних філософів не розглядав поняття «інформація» як важливий філософський концепт [10]. Поштовхом до виникнення зацікавленості в дослідженні проблем інформації і даних як її першоджерела стала комп'ютеризація різноманітних сфер життєдіяльності людини.

Н. Вінер, основоположник кібернетики, визначав інформацію дещо тавтологічно: «Інформація це інформація, а не матерія чи енергія (Information is information, not matter or energy)» [11, с. 132]. Тим не менш, підкреслюючи в такий спосіб якісну самобутність феномену інформації, ця дефініція визначає важливий логічний статус поняття інформації – на одному рівні з поняттями матерії та енергії. Відповідно до іншого тлумачення, що також належить Н. Вінеру, інформація – це визначення змісту, отриманого із зовнішнього світу в процесі пристосування до нього нас самих і наших відчуттів [12, с. 31].

Фізична енциклопедія розуміє під інформацією будь-які відомості та дані, що відображають властивості об'єктів у природних (біологічних, фізичних тощо), соціальних і технічних системах і передаються звуковим, графічним (у т. ч. письмовим) чи іншим способом без застосування чи із застосуванням технічних засобів [13].

Розвиток саме технічних засобів надав людству нові можливості щодо фіксування, обробки, збереження та аналізу дедалі більших за обсягом і різноманітніших за природою і структурою масивів даних. З іншого боку, еволюція технологій спричинила і збільшення виробництва даних – їх нині продукують найрізноманітніші пристрої – від мобільних телефонів і подібних «дрібних» гаджетів до приладів рівня Слоанівського телескопа чи адронного колайдера.

«Цунами» цифрових даних, що зародилося в середині 2000-х років (див. рис. 3) отримало назву Big Data – «великі дані». Однією з особливостей феномену великих даних є те, що, належачи за своєю сутністю до предметної сфери фахівців з комп'ютерних технологій, цей концепт став не лише широко відомим в ІТ-спільноті, а й викликав велику зацікавленість серед представників бізнесу, мас-медіа та науковців.

Сьогодні термін Big Data вживається щонайменше у двох значеннях:

- як власне дані, що характеризуються великим обсягом і низкою інших властивостей;
- як технології роботи з такими даними.

Інтерес із часом ?

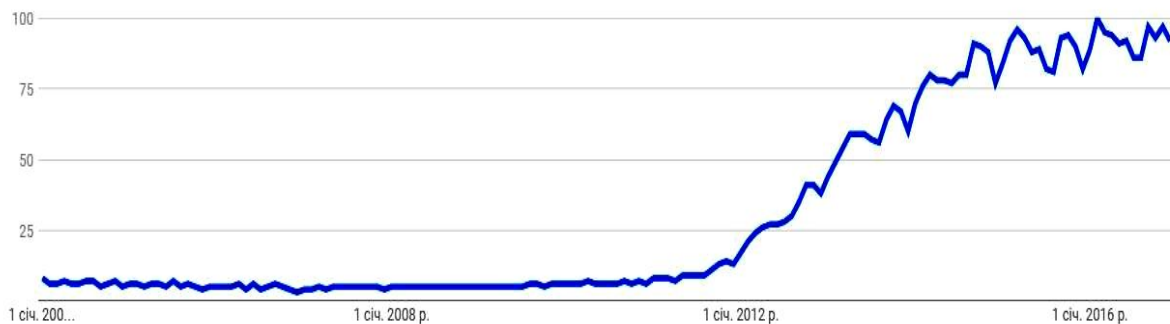


Рис. 3. Динаміка пошукових запитів «Big Data» в Google за 2004–2016 рр. [14].

Оксфордський словник визначає Big Data як надзвичайно великі набори даних, які можуть бути проаналізовані за допомогою комп'ютерів з метою виявлення певних шаблонів (зразків, паттернів), тенденцій та зв'язків, особливо стосовно поведінки та взаємодії людей [15].

Наразі загальноновизнаною є концепція великих даних, запропонована Д. Лані (Doug Laney), згідно з якою основними їх характеристиками є обсяг (volume), швидкість (velocity), різноманітність (variety) [16].

Коли йдеться про обсяги великих даних, дослідники зазвичай оперують такими незвичними для пересічного користувача одиницями виміру інформації, як терабайти, петабайти тощо [17, с. 13–15]. Відтак серед дефініцій Big Data зустрічаються такі, у яких йдеться про «незручно великі» обсяги: М. Алексєєв визначає Big Data як набори даних такого обсягу, що традиційні інструменти не здатні виконувати їх захват, управління та обробку за прийнятний для практики час [18].

Швидкість як характеристика великих даних означає, що дані стрімко накопичуються та потребують якомога швидшого реагування у вигляді обробки потоків даних у режимі реального часу.

Різноманітність Big Data полягає в тому, що це може бути як структурована, так і неструктурована інформація. До останньої належать текстові файли різноманітних документів, електронні листи, sms-повідомлення, аудіофайли, цифрові зображення, відеокліпи тощо. Особливістю таких неструктурованих даних є те, що часто вони створюються користувачами Інтернету. Як зазначає О. Ю. Колесниченко, «... Інтернет датафікує мовну функцію людини» [19, с. 24]. Тому профілі користувачів у соціальних мережах, блоги, коментарі під статтями і новинами, пошукові запити тощо перетворюються на дані й за умови застосування відповідних методів аналізу можуть виявитися дуже інформативними з погляду наявності в них так званих «прихованих знань», котрі можуть бути отримані за допомогою інтелектуального аналізу даних – Data Mining.

Data Mining – це дослідження та виявлення «машиною» (алгоритмами, засобами штучного інтелекту) в «сирих» даних прихованих знань, які раніше не були відомі, нетривіальні, практично корисні, доступні для інтерпретації людиною [20, с. 58].

Літературні джерела з проблематики великих даних містять безліч найрізноманітніших прикладів їх застосування – від прогнозування епідемії грипу на підставі аналізу пошукових запитів в Інтернеті до оптимізації витрат на придбання авіаквитків чи прогнозування тенденцій фондового ринку, що базуються на аналізі записів у Twitter. Сфера використання Big Data і сьогодні є надзвичайно широкою, а з плином часу вона тільки зростатиме. Тому слід враховувати, що велике значення великих даних, підкреслюване практично всіма дослідниками цієї проблематики, зумовлюється не самими даними чи їх обсягами, а результатами їх обробки й аналізу, котрі дозволяють викристалізувати нове знання. Відтак однією серед найбільш затребуваних професій найближчого майбутнього буде спеціаліст з Data Science, котрий володіє трьома групами компетенцій: IT-грамотністю, математичними та статистичними знаннями; та обізнаний у предметній сфері застосування попередніх груп навичок, тобто усвідомлює бізнес-запити своєї організації чи завдання відповідної науки.

Висновки. З розвитком технологічного укладу людства стало можливим фіксувати та зберігати гігантські обсяги структурованих і неструктурованих даних, що отримали назву Big Data.

Проте більш важливим є не набуття фізичної можливості володіти цими обсягами даних, а розуміння того, яким чином і з якою метою можна їх ефективно використати. Наразі як ніколи актуальним є перетворення даних в інформацію, а інформації – в знання, і саме з такого погляду слід розуміти відомий вислів «Хто володіє інформацією, той володіє світом». Сам факт володіння не забезпечує (як не забезпечував і раніше) конкурентних переваг без відповідного застосування набутого знання. Адекватне використання отриманих з інформації знань уможливує вирішення широкого кола завдань – від задоволення нових потреб людини, що з'являються в результаті розвитку технологій, до вирішення проблем збереження навколишнього середовища чи розв'язання завдань бізнесу щодо генерації прибутку.

СПИСОК ЛІТЕРАТУРИ

1. Sholle D. What is Information? The Flow of Bits and the Control of Chaos / D. Sholle // *Democracy and new media*. – Cambridge, Mass.: MIT Press. – 2003. – P. 343–364. [Електронний ресурс]. – Режим доступу: <http://web.mit.edu/comm-forum/papers/sholle.html>.
2. Ackoff R. L. From Data to Wisdom / R. L. Ackoff // *Journal of Applied Systems Analysis*. – Vol. 16. – 1989. – P. 3–9.
3. DИKW [Електронний ресурс] // Вікіпедія: вільна енцикл. – Електрон. дані. – Режим доступу: <https://ru.wikipedia.org/wiki/DИKW>. – Назва з екрана. – Дата перегляду: 05.09.2016.
4. Великий тлумачний словник сучасної української мови (з дод. і допов.) / Уклад. і голов. ред. В. Т. Бусел. – К., Ірпін': ВТФ «Перун», 2005. – 1728 с.
5. Іванов В. Г. Основи інформатики та обчислювальної техніки: підруч. / В. Г. Іванов, В. В. Карасюк, М. В. Гвозденко; за заг. ред. В. Г. Іванова. – Х.: Право, 2015. – 312 с.
6. Ковтун Н. В. Теорія статистики: підруч. / Н. В. Ковтун. – К.: Знання, 2012. – 399 с.
7. ISO/IEC 2382:2015 «Information technology – Vocabulary» [Електронний ресурс]. – Режим доступу: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en>.
8. Закон України «Про державну статистику» // Відомості Верховної Ради України (ВВР), 1992, № 43, ст. 608. [Електронний ресурс] – Режим доступу: <http://zakon5.rada.gov.ua/laws/show/2614-12>.
9. Донец Л. І. Управление знаниями в контексте инновационного развития предприятия / Л. И. Донец // Теоретичні і практичні аспекти економіки та інтелектуальної власності. – 2012. – Вип. 1. – Т. 1. – С. 292–296.
10. Information [Електронний ресурс] // *Stanford Encyclopedia of Philosophy*. – Електрон. дані. – Режим доступу: <http://plato.stanford.edu/entries/information>. – Назва з екрана. – Дата перегляду: 05.09.2016.
11. Wiener N. Cybernetics: Or the Control and Communication in the Animal and the Machine. 2nd ed. / N. Wiener. – Cambridge, MA: MIT Press. – 1965. – 212 p.
12. Винер Н. Кибернетика и общество / Н. Винер. – М.: Издательство иностранной литературы, 1958. – 199 с.
13. Інформація [Електронний ресурс] // *Энциклопедия физики и техники*. – Електрон. дані. – Режим доступу: http://femto.com.ua/articles/part_1/1406.html. – Назва з екрана. – Дата перегляду: 05.09.2016.
14. Big Data [Електронний ресурс] // *Google Trends*. – Електрон. дані. – Режим доступу: <https://www.google.com.ua/trends/explore?date=all&q=big%20data>. – Назва з екрана. – Дата перегляду: 05.09.2016.
15. Big Data [Електронний ресурс] // *Oxford dictionaries*. – Електрон. дані. – Режим доступу: <http://www.oxforddictionaries.com/definition/english/big-data>. – Назва з екрана. – Дата перегляду: 05.09.2016.
16. Laney D. 3D Data Management: Controlling Data Volume, Velocity and Variety [Електронний ресурс]. – Режим доступу: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
17. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер. – М.: Манн, Иванов и Фербер, 2014. – 240 с.
18. Алексеев М. Big Data-революция в области хранения и обработки данных. [Електронний ресурс]. – Режим доступу: http://www.slideshare.net/mikhailalekseev71/big-data-40258380?qid=0b867437-14f5-4631-868a-788191c5954b&v=&b=&from_search=6.
19. «Третья волна»: многоцентровое исследование по аналитике Big Data Академического партнерства ЕМС в России и СНГ / О.Ю. Колесниченко, Г.Н.Сморозин, И.В.Ильин и др. // *Мониторинг общественного мнения: экономические и социальные перемены*. – 2015. – № 5. – С. 21–41.
20. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007. – 384 с.

REFERENCES

1. Sholle D. What is Information? The Flow of Bits and the Control of Chaos / D. Sholle // *Democracy and new media*. – Cambridge, Mass.: MIT Press. – 2003. – P. 343–364. [Elektronnyj resurs]. – Rezhim dostupu: <http://web.mit.edu/comm-forum/papers/sholle.html>.
2. Ackoff R. L. From Data to Wisdom / R. L. Ackoff // *Journal of Applied Systems Analysis*. – Vol. 16. – 1989. – P. 3–9.
3. DИKW [Elektronnyj resurs] // Vikipediia: vil'na entsykl. – Elektron. dani. – Rezhym dostupu: <https://ru.wikipedia.org/wiki/DИKW>. – Nazva z ekrana. – Data perehliadu: 05.09.2016.
4. Velykyj tлумachnyj slovnyk suchasnoi ukrains'koi movy (z dod. i dopov.) / Uklad. i holov. red. V. T. Busel. – K., Irpin': VTF «Perun», 2005. – 1728 s.
5. Ivanov V. H. Osnovy informatyky ta obchysliuval'noi tekhniky: pidruch. / V. H. Ivanov, V. V. Karasiuk, M. V. Hvozdenko; za zah. red. V. H. Ivanova. – Kh.: Pravo, 2015. – 312 s.
6. Kovtun N. V. Teoriia statystyky: pidruch. / N. V. Kovtun. – K.: Znannia, 2012. – 399 s.
7. ISO/IEC 2382:2015 «Information technology – Vocabulary» [Elektronnyj resurs]. – Rezhim dostupu: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en>.
8. Zakon Ukrainy «Pro derzhavnu statystyku» // Vidomosti Verkhovnoi Rady Ukrainy (VVR), 1992, № 43, st. 608. [Elektronnyj resurs] – Rezhym dostupu: <http://zakon5.rada.gov.ua/laws/show/2614-12>.
9. Donec L. I. Upravlenie znanijami v kontekste innovacionnogo razvitija predprijatija / L. I. Donec // Teoretychni i praktychni aspekty ekonomiky ta intelektual'noi vlasnosti. – 2012. – Vyp. 1. – T. 1. – S. 292-296.

10. Information [Elektronij resurs] // Stanford Encyclopedia of Philosophy. – Elektron. dani. – Rezhim dostupu: <http://plato.stanford.edu/entries/information>. – Nazva z ekrana. – Data peregljadu: 05.09.2016.

11. Wiener N. Cybernetics: Or the Control and Communication in the Animal and the Machine. 2nd ed. / N. Wiener. – Cambridge, MA: MIT Press. – 1965. – 212 p.

12. Viner N. Kibernetika i obshhestvo / N. Vinner. – M.: Izdatel'stvo inostrannoj literatury, 1958. – 199 s.

13. Informacija [Elektronij resurs] // Jenciklopedija fiziki i tehniki. – Elektron. dani. – Rezhim dostupu: http://femto.com.ua/articles/part_1/1406.html. – Nazva z ekrana. – Data peregljadu: 05.09.2016.

14. Big Data [Elektronij resurs] // Google Trends. – Elektron. dani. – Rezhim dostupu: <https://www.google.com.ua/trends/explore?date=all&q=big%20data>. – Nazva z ekrana. – Data peregljadu: 05.09.2016.

15. Big Data [Elektronij resurs] // Oxford dictionaries. – Elektron. dani. – Rezhim dostupu: <http://www.oxford-dictionaries.com/definition/english/big-data>. – Nazva z ekrana. – Data peregljadu: 05.09.2016.

16. Laney D. 3D Data Management: Controlling Data Volume, Velocity and Variety [Elektronij resurs]. – Rezhim dostupu: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

17. Majer-Shenberger V. Bol'shie dannye. Revolucija, kotoraja izmenit to, kak my zhivem, rabotaem i myslim / Viktor Majer-Shenberger, Kennet Kuk'er. – M.: Mann, Ivanov i Ferber, 2014. – 240 s.

18. Alekseev M. Big Data-revolucija v oblasti hranenija i obrabotki dannyh. [Elektronij resurs]. – Rezhim dostupu: http://www.slideshare.net/mikhailalekseev71/big-data-40258380?qid=0b867437-14f5-4631-868a-788191c5954b&v=&b=&from_search=6.

19. «Tret'ja volna»: mnogocentrovoe issledovanie po analitike Big Data Akademicheskogo partnerstva EMS v Rossii i SNG / O.Ju. Kolesnichenko, G.N. Smorodin, I.V. Il'in i dr. // Monitoring obshhestvennogo mnenija: jekonomicheskie i social'nye peremeny. – 2015. – № 5. – S. 21–41.

20. Tehnologii analiza dannyh: Data Mining, Visual Mining, Text Mining, OLAP / A. A. Barsegjan, M. S. Kuprijanov, V. V. Stepanenko, I. I. Holod. – 2-e izd., pererab. i dop. – SPb.: BHV-Peterburg, 2007. – 384 s.

Big Data – новый информационный феномен цифровой эпохи

Р. П. Задорожная

Исследованы сущность и взаимосвязи понятий «данные», «информация», «знания» в рамках концепции информационной иерархии DIKW и в соответствии с терминологическим стандартом ISO/IEC 2382:2015. На высшем уровне обобщения раскрыто сущность данных как зарегистрированных сигналов. Проанализированы существующие подходы к толкованию понятия информации; значение контекста и необходимость применения соответствующих методов обработки данных для превращения их в информацию. Рассмотрено деление полученной из данных информации на аналоговую и цифровую.

Охарактеризованы сущность и особенности Big Data как разновидности данных, которая возникла вследствие развития компьютерных технологий. Рассмотрены основные особенности (3V) больших данных: их объём (volume), скорость (velocity) и разнообразие (variety). Рассмотрен потенциал интеллектуального анализа данных (Data Mining) как средства получения скрытых знаний из структурированных и неструктурированных массивов больших данных.

Ключевые слова: данные, информация, знания, DIKW, Big Data, Data Mining, неструктурированная информация.

Big Data – a new information phenomenon of digital epoch

R. Zadorozhna

The essence of concepts «data», «information», «knowledge» is investigated in the article from positions of DIKW (Information Hierarchy) of American organizational theorist Russell Ackoff. DIKW belongs to a class of models for representing relationships between data, information, knowledge and wisdom. In this model information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge.

Also concepts «data», «information», «knowledge» are described according to ISO/IEC 2382:2015 «Information technology – Vocabulary».

Data is the representation of information in a formalized manner suitable for communication, interpretation or processing. Information (in information processing) is any fact, concept or meaning derived from data and associated context, or selected from knowledge. Knowledge can be described as organized, integrated collection of facts and generalization.

It is important for our research to divide information obtained from data on analog and digital. The first is obtained through the use of natural methods of data processing and is perceived exclusively by human. It has a continuous character and is divided according to the five human senses to visual, auditory, tactile, olfactory and gustatory. Digital information is discrete and is perceived by computers.

Development of technical facilities gave new possibilities for recording, processing, storage and analysis of increasingly larger and more varied in nature and structure data sets. On the other hand, the evolution of technology increased data production.

Big Data appeared due to development of computer technologies. During the last years Big Data has attracted attention of more and more researchers. Year by year, data collected from different spheres of activity is getting larger and more complicated.

Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze. That's why some investigators called Big Data as “inconveniently large data”.

Oxford dictionary defines Big Data as extremely large data sets that may be analyzed computationally to reveal patterns, trends and associations, especially relating to human behavior and interactions.

Big data is a term used to describe a large volume of both structured and unstructured data. But a huge volume is not the only one feature of Big Data.

The other features are:

velocity (data in motion) – data streams with an unprecedented speed and must be processed in time. Speed as a characteristic of Big Data means that data is collected rapidly and requires response as soon as possible as processing data streams in real time;

variety (data in many forms) – data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock ticker data and financial transactions. Such unstructured data is often created by Internet users. In addition, companies collect data from a wide range of sources, including different financial and business transactions, social media and information from sensor or machine-to-machine data.

SAS, one of the leading producers of software for statistical analysis, nominates two additional features of Big Data:

variability – data flows can be highly inconsistent with periodic peaks (daily, seasonal etc.);

complexity – data comes from multiple sources, which makes it difficult to link, match, cleanse and transform data across systems.

Unstructured arrays of data reflect the actual processes in the Internet information environment and could be used for obtaining information through Data mining. Data mining is the process that attempts to discover useful information (or patterns) in large data repositories. Data mining enables to analyze data regardless of source, type, size, or format.

A scope of Big Data using is extremely broad now and it will grow in future. Today it is a business priority, because the effective use of data is becoming the basis of competition. Big Data can improve the profitability, business value and provide a success of many enterprises. So, Big Data will fundamentally change the way businesses compete and operate.

It should be noted that the huge importance of Big Data is predefined not by data itself or by its size, but by results of processing and analysis that allow to provide new knowledge. Thus one of the most popular professions of the near future will be a specialist in Data Science, who has three groups of competencies: IT literacy, mathematical and statistical knowledge and awareness in practical application of the previous groups of skills, who understands business needs of his organization or tasks of relevant science area.

Key words: data, information, knowledge, DIKW, Big Data, Data Mining, unstructured information.

Надійшла 3.08.2016 р.