

ОСОБЛИВОСТІ ФОРМУВАННЯ ПОНЯТТЯ ВЕЛИКИХ ДАНИХ У ПРОЦЕСІ ПІДГОТОВКИ ІТ-ФАХІВЦІВ

Значне збільшення обсягів інформації, що спостерігається протягом останнього десятиліття, є одним з наслідків розвитку інформаційного суспільства. Це привело до необхідності розробки нових методів і засобів роботи з великими обсягами даних – їх пошук, збір, накопичення, аналіз. Для роботи з такими даними розробляються програмні та апаратні засоби, інформаційні системи, методи і технології. З'явилося поняття, що описує всі ці явища – «Big Data». У свою чергу в системі вищої ІТ-освіти вводять нові дисципліни, пов'язані з цими явищами, у рамках яких, перш за все, необхідно розкрити поняття великих даних – «Big Data». Уявлення про великі дані формується в процесі теоретичної підготовки ІТ-фахівців. Студенти повинні розуміти, що мається на увазі під «великими даними», у чому полягає нова ідеологія обробки інформації і що призвело до її появи. Найбільш доцільно розглядати сутність, розвиток і трансформацію поняття великих даних в історичному контексті. Таким чином, у статті зроблено історичний екскурс розвитку дефініції великих даних, описані різні погляди на визначення сутності цього поняття: крізь кількісне вимірювання обсягу інформації, крізь опис якісних ознак «певну кількість V», крізь ознаки технологій класу Big Data, крізь особливу ідеологію обробки даних. Визначено сучасний стан та подальший можливий напрям розвитку цього поняття.

Ключові слова: професійні знання, майбутні фахівці з інформаційних технологій, великі дані, Big Data, інформаційні технології, зміст навчання.

Вступ. Одним із наслідків розвитку інформаційного суспільства є значне збільшення обсягів інформації. Так за даними аналітичної компанії IDC, опублікованому у звіті «Ера даних 2025» обсяг усіх даних в світі зростає з 33 Зеттабайт у 2018 р. до 175 Зеттабайт до 2025 р. [11]. При цьому значне збільшення Інформації приходить на долю Інтернету. За прогнозами компанії Cisco, що опубліковані у звіті «Наочний індекс розвитку мережевих технологій» (Cisco Visual Networking Index Complete Forecast, Cisco VNI) за період 2016–2021 рр. світовий обсяг IP-трафіку зростає триразово. До 2021 р. світовий обсяг IP-трафіку досягне 278 Ексабайт на місяць. Річний обсяг IP-трафіку до 2021 р. може досягти 3,3 Зеттабайт. Пристрої, підключені до мереж Wi-Fi і мобільного зв'язку, до 2021 р. будуть генерувати 73 % інтернет-трафіку. Це призвело до зміни парадигми накопичення даних. Самі дані стали цінним ресурсом, що можна застосовувати в різних галузях. Розробляються програмні й апаратні засоби, інформаційні системи, методи та технології роботи з великими обсягами даних. З'явилося поняття, що описує усі ці явища – «Big Data». У свою чергу в системі вищої ІТ-освіти вводять нові дисципліни, пов'язані з цими явищами в рамках яких, перш за все, необхідно розкрити поняття великих даних – «Big Data».

Аналіз досліджень і публікацій. Уже з'явилося багато ресурсів, що переймаються проблемами дослідження,

використання та стандартизації великих даних. Установи, що займаються стандартизацією також працюють над проблемами ефективного спільного використання джерел даних. Рішення з відкритими вихідними кодами дозволяють досліджувати, вивчати та експериментувати з даними великого розміру. Таким даним вже присвячено багато наукових розробок, літератури, конференцій, форумів тощо. Серед всього цього варто звернути увагу на роботи присвячені тематиці великих даних таких авторів, як Дж. Гурвіц, А. Ньюджент, Ф. Халпер, М. Кауфман [10]. Провідними організаціями, що займаються стандартизацією великих даних є Фонд відкритих даних (The Open Data Foundation, ODAF (www.opendatafoundation.org)), Альянс інформаційної безпеки (Cloud Security Alliance (cloudsecurityalliance.org)), Національний інститут стандартів ті технологій (National Institute of Standards and Technology, NIST (www.nist.gov)), Фонд програмного забезпечення Apache (Apache Software Foundation, ASF (www.apache.org)), Організація, щодо розвитку стандартів структурованої інформації OASIS (Organization for the Advancement of Structured Information Standards (www.oasis-open.org)).

Багато провідних світових ІТ-компаній уже мають програмні продукти пов'язані з великими даними, серед яких можна виокремити розробки таких компаній, як Google (Google AI (ai.google/research)), Amazon (Сервіси хмарних технологій (aws.amazon.com/))

ru/?nc2=h_lg)), IBM (www.ibm.com/analytics/products, cognitiveclass.ai/learn/big-data/), Oracle (Big Data, Big Data Appliance, Big Data Cloud Service, Big Data Connectors, Big Data Discovery, Big Data Platform, Big Data Preparation Cloud Service, Big Data SQL, Big Data Spatial and Graph (www.oracle.com/ru/products/oracle-a-z.html#b), Microsoft (www.microsoft.com/en-us/sql-server/big-data) тощо. Аналітичні компанії також не обійшли стороною це явище. Так аналітична компанія Gartner вперше згадує про великі дані у циклі зрілості технологій (Hype Cycle Emerging Technologies) за 2011 рік під назвою ««Big Data» and Extreme

Information Processing and Management» [12]. Google Trends (trends.google.com) показує активний ріст вживання терміна «Big Data» у пошукових запитах, починаючи саме з 2011 р. (див. рис. 1). Така сама тенденція притаманна і нашій країні (див. рис. 2).

Проте виявилось, що це поняття вельми розпливчате, неоднозначне і викликає багато суперечностей серед фахівців. Отже метою цієї статті є дослідження сутності поняття «Big Data» (Великі дані) та окреслення особливостей його формування в навчальному процесі підготовки ІТ-фахівців.

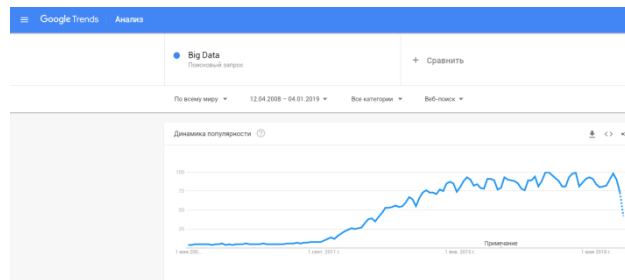


Рис. 1. Статистика використання пошукового запиту «Bag Data» в світі за даними аналітичної платформи Google Trends

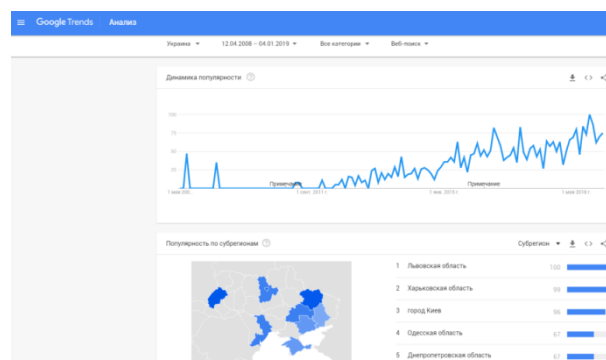


Рис. 2. Статистика використання пошукового запиту «Bag Data» в регіонах України за даними аналітичної платформи Google Trends

Виклад основного матеріалу. У процесі теоретичної підготовки ІТ-фахівців необхідно сформувати уявлення про великі дані. Студенти мають розуміти, що мається на увазі під «Великими даними» і що призвело до виникнення нової ідеології обробки інформації. Якщо звернутися до історії, то поняття «Big Data» має дату появи – 4 вересня 2008 р., коли вийшов спеціальний номер британського журналу Nature, присвячений проблематиці бурхливого зростання глобальних даних та їхньої ролі в науці та суспільстві [6]. І саме тут уперше звертається увага на зміну відношення до даних. Проте поняття «Big Data» належить до понять, що не мають чіткого трактування. Так, спочатку фахівці стверджували, що критерієм відношення до категорії Big Data, перш за все, є потік даних більший за 100 Гбайт на день. Проте, виявилось, що цього недостатньо, щоб однозначно віднести продукт до категорії Big Data. Згодом аналітична компанія Gartner вводить визначення поняття «Big Data» крізь «три V» [8]:

– Volume (об’єм) – це збільшення обсягу даних у корпоративних системах, за рахунок збільшення обсягів транзакцій та використання традиційних і нових типів даних. Спеціалісти компанії Gartner наголошу-

ють на тому, що занадто великий обсяг даних викликає проблему їхнього зберігання, проте це також породжує проблему аналізу даних.

– Variety (різноманіття та неструктурованість даних) полягає в різноманітному форматі наявних даних, це можуть бути табличні дані (бази даних), ієрархічні дані, документи, дані електронної пошти, дані вимірювання, відео, нерухомі зображення, аудіо, дані біржових цінних паперів, дані фінансових операцій тощо. Це породжує проблему перекладу великих обсягів транзакційної інформації в рішення.

– Velocity (швидкість оброблення) означає як швидкість отримання, так і швидкість обробки даних для задоволення запиту.

Такий підхід підхопили інші фахівці й розширили ознаки. Так компанія Forrester визначає поняття «Big Data» як технологію в області апаратного та програмного забезпечення, яка об’єднує, організовує, керує і аналізує дані, що характеризуються «чотирима V»: об’ємом (Volume), різноманітністю (Variety), мінливістю (Variability) та швидкістю (Velocity) [9, с. 4–5]. Тут під об’ємом розуміється дуже великий обсяг інформації, накопичений у базах даних, що складно обробляти та зберігати традиційними засобами СУБД.

Аналітики компанії Forrester роблять акцент на необхідності використання нових підходів і вдосконаленні наявних інструментів обробки та збереження даних. Розуміння ознаки різноманітності даних аналітиками компанії Forrester і Gartner є схожим. Швидкість тут розуміється як швидкість не тільки обробки даних, а і їх накопичення. Аналітики звертають увагу на те, що стали більш затребувані технології обробки даних у

реальному часі. Також була додана четверта ознака Variability – мінливість інформації. Наприклад, такою є інформація, що безперервно надходить з датчиків пристроїв або з Інтернету і має важливе значення для аналізу, прогнозування та прийняття рішень. На рис. 3 наведено діаграму співвідношення цих чотирьох ознак великих даних.

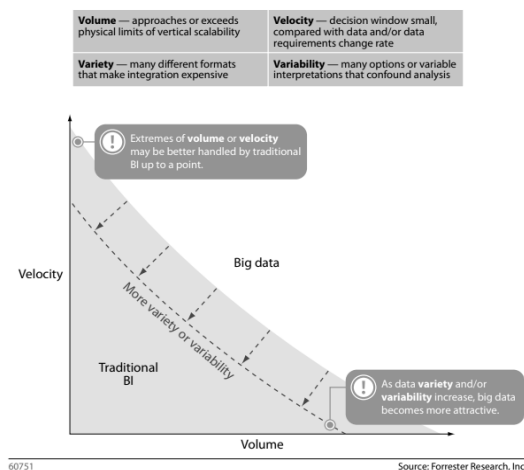


Рис. 3. Діаграми чотирьох ознак Великих даних за даними компанії Forrester [9, с. 5]

Згодом цей список ознак було поповнено: «п'ять V» (додано Viability – життєздатність, Value – цінність), «сім V» (додано Veracity – достовірність, Visualization – візуалізація) тощо. Так, наприклад, спеціалістами вітчизняного освітнього ресурсу про новітні технології «The Future» було виокремлено «п'ять V» [3]: Volume (об'єм), Velocity (швидкість), Variety (різноманітність), Variability (мінливість) та Veracity (достовірність). Під достовірністю тут розуміється саме виокремлення достовірних даних. Необхідність п'ятої ознаки пояснюється тим, що якість зафіксованих даних може сильно відрізнятись, що має істотний вплив на результати аналізу.

Надалі цей список «V» можна продовжувати, конкретизуючи його для кожної задачі або бізнес-процесу, проте такий підхід не розкриває це поняття повною мірою, а лише торкається окремих його аспектів. Як справедливо зазначає Джудіт Гурвіц: «Вищевказані ознаки не завжди виражені у великих даних однаковою мірою. Наприклад, методи роботи з великими даними можна застосувати для управління порівняно невеликим об'ємом різноманітних і складних даних або при обробці дуже простих даних, але у великій кількості» [2, с. 27]. Усі ці ознаки є якісними. Постають питання, коли, наприклад, обсяг даних великий, а коли він ще недостатньо великий, чим це виміряти, що взяти за точку відліку. Яку швидкість оброблення вважати швидкою тощо.

Розглядаючи «Big Data» як технологію, тут можна виокремити три напрями за завданнями, що вирішуються:

- збір, первісне оброблення та зберігання даних для їхнього подальшого використання;
- структурування розрізненого контенту: текстового, графічного, відео, аудіо, звукового тощо;
- бізнес-аналітика на великих обсягах даних.

Для кожного напрямку існують свої вимоги та ознаки технологій цього класу. Отже неможливо і

немає потреби поєднувати в одному понятті всі ці напрями. Так у статті про Великі дані з Вікіпедії надаються альтернативні визначення поняття Великих даних крізь зазначені ознаки і звертається увага на різний контекст розгляду цього поняття: «Великі дані (англ. Big Data) в інформаційних технологіях – набори інформації (як структурованої, так і неструктурованої) настільки великих розмірів, що традиційні способи та підходи (здебільшого засновані на рішеннях класу бізнесової аналітики та системах управління базами даних) не можуть бути застосовані до них. Альтернативне визначення називає великими даними феноменальне прискорення нагромадження даних та їх ускладнення. Важливо також відзначити те, що часто під цим поняттям у різних контекстах можуть мати на увазі як дані великого об'єму, так і набір інструментів та методів (наприклад, засоби масово-паралельної обробки даних системами категорії NoSQL, алгоритмами MapReduce, чи програмними каркасами проекту Hadoop)» [1].

Фахівці сходяться на думці, що «Big Data» – це зонтичний термін, синергія величезної кількості технологій, багато з яких самі по собі гідні найпильнішого розгляду. Проте, з'явилася нова галузь і сформовані підходи щодо роботи з даними. З'явилися засоби, що дозволяють працювати з конкретними бізнес-проблемами і знаходити розумні рішення задач нового рівня [7]. Так аналітична компанія Forrester у своєму дослідженні TechRadar: Big Data, Q1 2016 описала поточний стан та перспективи розвитку 22 провідних технологій класу Big Data (див. рис. 4). Кількість і розмір стрілок у кружечках означають швидкість просування даної технології по кривій розвитку TechRadar. Квадратик або вертикальні лінії показують те, що технологія практично не розвивається, або досягла максимуму своїх можливостей.

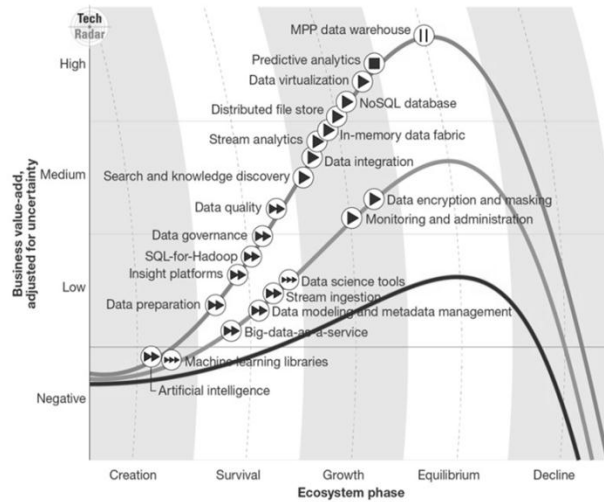


Рис. 4. Поточний стан та перспективи розвитку 22 провідних технологій Big Data від компанії Forrester

В інших дослідженнях можна зустріти визначення цього поняття, як неможливість оброблення даних традиційними способами. Проте таке визначення є тимчасовим, адже способи, що нині є новітніми, завтра вже можуть стати традиційними. Так, зокрема, технології обробки великих обсягів даних експерти вже не відносять до новітніх. Отже визначення поняття «Big Data» потребує інших підходів. На противагу до цього існують точки зору, коли взагалі пропонується відмовитися від цього поняття [4]. На думку багатьох експертів, настала епоха, коли важливо не просто вміти акумулювати інформацію, а витягувати з неї бізнес-вигоду. Першими до цього висновку прийшли індустрії, які безпосередньо працюють зі споживачем: телекомунікаційна і банківська. Тепер процеси взаємодії виходять на новий рівень, дозволяючи налагодити зв'язок між різними пристроями з використанням інструментів доповненої реальності та відкривають нові можливості оптимізації бізнес-процесів компаній. Нині спостерігається втрата інтересу з боку реального бізнесу щодо поняття великих даних. Компанія Gartner у циклі зрілості технологій за 2015 рік прибрала технологію «Big Data». Аналітики компанії пояснюють це розмиванням терміна – технології, що входять у поняття великих даних стали повсякденною

реальністю бізнесу. Тобто до складу поняття «Big Data» входить велика кількість технологій, що вже частково належать до інших популярних сфер і стали повсякденним робочим інструментом. На діаграмі Gartner його місце зайняли інші більш вузько спрямовані технології. Це, у першу чергу, машинне навчання (Machine Learning) – засоби пошуку правил і зав'язків у дуже великих обсягах інформації. Такі технології дозволяють не просто перевіряти гіпотези, але й шукати невідомі раніше фактори впливу. Також крім машинного навчання, це технології для зберігання даних і паралельного доступу до них (NoSQL Database), попередній обробці потоків інформації (Marshalling); рішення для візуалізації та самостійного аналізу (Advanced Analytics with Self-Service Delivery). Крім того, зберігають своє значення засоби інтелектуального аналізу даних (Business Intelligence і Data Mining), що виходять на новий технологічний рівень. Отже в літературі починають вживатися інші поняття, що за думкою фахівців є більш влучними, наприклад, «Data Mining». Проте, якщо звернутися до Google Trends, то поки що неможливо стверджувати про беззаперечну втрату інтересу до великих даних і заміну іншими поняттями (див. рис. 5).

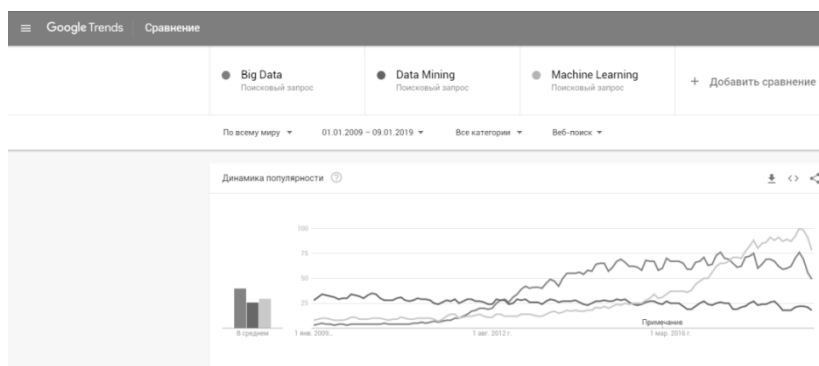


Рис. 5. Порівняльна статистика використання пошукових запитів «Big Data», «Data Mining» та «Machine Learning» у світі за даними аналітичної платформи Google Trends

Висновки. Отже у процесі підготовки IT-фахівців відмовитися від поняття «Big Data» поки що неможливо, адже воно глибоко увійшло в життя. Цей процес змінив наше відношення до даних та інформації, а це

у свою чергу вплинуло і на сутність поняття «Big Data». Якщо спочатку «Big Data» розумілося як певний стек технологій, то нині, як влучно зазначають IT-фахівці, це скоріше особливий підхід – ідеологія

процесінгу інформації, що застосовується для оброблення великих масивів «сирих» даних [5]. На наш погляд, такий підхід найбільш влучно відображає сутність поняття «Big Data» на сучасному етапі розвитку. Проте не виключено, що в майбутньому це поняття

може зникнути як застаріле та неактуальне, адже великі обсяги даних і сучасні підходи їхнього оброблення можуть стати звичайними засобами і не буде потреби робити наголос на їхній інноваційній складовій.

Список використаних джерел

1. Великі дані [Електронний ресурс] : Матеріал з Вікіпедії – вільної енциклопедії, 2019. – Режим доступу : https://uk.wikipedia.org/wiki/Великі_дані Мова укр. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
2. Гурвиц Дж. Просто о больших данных / Дж. Гурвиц, Ньюджент А., Халпер Ф. и др. : [перевод с английского]. – М. : Эксмо, 2015 – 400 с.
3. Кравчук С. Що таке Big Data? [Електронний ресурс] / С. Кравчук // The Future. – Режим доступу : <http://thefuture.news/bigdata>. Мова укр. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
4. Реймер Д. Gartner : Big Data больше не существует! [Електронний ресурс]. – Режим доступу : <http://denreymer.com/gartner-end-of-big-data>. Мова рос. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
5. Савчук И. Big Data – технология, рождающая новый тип бизнеса [Електронний ресурс] / И. Савчук // БИТ. – № 3 (36). – 2014. – Режим доступу : <http://bit.samag.ru/archive/article/1352>. Мова рос. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
6. Big Data [Електронний ресурс] / Nature. – V. 455. – N. 7209. – pp 1-136. – Режим доступу : <http://www.nature.com/nature/journal/v455/n7209/index.html> Мова англ. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
7. BIG DATA 2016 : Форум умных решений [Електронний ресурс] / Издательство «Открытые системы». – 07.04.2016. – Режим доступу : <https://www.osp.ru/news/articles/2016/14/13048945>. Мова рос. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
8. Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data [Електронний ресурс] // Gartner. – STAMFORD. – Conn. – June 27, 2011. – Режим доступу : <http://www.gartner.com/newsroom/id/1731916> Мова англ. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
9. Hopkins B., Evelson B. Expand your digital horizon with Big Data / B. Hopkins, B. Evelson. – Forrester Research, Inc. Reproduction Prohibited. – September 30, 2011. – 16 p.
10. Hurwitz & Associates [Електронний ресурс]. – Режим доступу : <https://hurwitz.com> Мова англ. – Дата останнього доступу : 15.01.2019. – Назва з екрана.
11. Reinsel D. The Digitization of the World From Edge to Core [Електронний ресурс] / D. Reinsel, J. Gantz, J. Rydning // Data Age 2025. – November 2018. – Режим доступу : <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>
12. Smith D. Hype Cycle for Cloud Computing, 2011 [Електронний ресурс] / D. Smith // Gartner. – 27 July 2011. – Режим доступу : <https://www.gartner.com/doc/1753115/hype-cycle-cloud-computing-> Мова англ. – Дата останнього доступу : 15.01.2019. – Назва з екрана.

Е. А. Кирей,

ЧНУ ім. Петра Могили, г. Николаев, Україна

ОСОБЕННОСТИ ФОРМИРОВАНИЯ ПОНЯТИЯ БОЛЬШИХ ДАННЫХ В ПРОЦЕССЕ ПОДГОТОВКИ ИТ-СПЕЦИАЛИСТОВ

Значительное увеличение объемов информации, которое наблюдается в течение последнего десятилетия, является одним из следствий развития информационного общества. Это привело к необходимости разработки новых методов и средств работы с данными – их поиску, сбору, накоплению, анализу. Для работы с такими данными разрабатываются программные и аппаратные средства, информационные системы, методы и технологии. Появилось понятие, описывающее все эти явления – «Big Data». В свою очередь в системе высшего ИТ-образования вводят новые дисциплины, связанные с этими явлениями, в рамках которых, прежде всего, необходимо раскрыть понятие больших данных – «Big Data». Представление о больших данных формируется в процессе теоретической подготовки ИТ-специалистов. Студенты должны понимать, что подразумевается под «Большими данными», в чем заключается новая идеология обработки информации и что привело к ее появлению. Наиболее целесообразным здесь является рассмотрение сущности, развития и трансформации понятия больших данных в историческом контексте. Таким образом, в статье сделан исторический экскурс развития дефиниции больших данных, описаны разные взгляды на определение сущности этого понятия: через количественное измерение объема информации, через описание качественных признаков «определенное количество V», через признаки технологий класса Big Data, через особую идеологию обработки данных. Определено современное состояние и дальнейшее возможное направление развития этого понятия.

Ключевые слова: профессиональные знания; будущие специалисты информационных технологий; информационные технологии; большие данные; Big Data.

Е. А. Kirey,

Petro Mohyla Black Sea National University, Mykolaiv, Ukraine

FEATURES OF FORMATION OF THE CONCEPT OF THE BIG DATA IN THE COURSE OF TRAINING OF IT SPECIALISTS

Significant increase in volumes of information which is observed within the last decade is one of consequences of development of information society. It led to develop new methods and means (tools) of working with large volumes of data, their search, collecting,

accumulation, the analysis. Software and hardware, information systems, methods and technologies are developed for work with such data program. The term appeared describing all these phenomena – «Big Data». In turn the new disciplines are entered in the system of the higher IT education connected with these phenomena within which, first of all, it is necessary to open a concept of big data – «Big Data». Understanding about big data is formed in the course of theoretical training of IT specialists. Students have to understand that is meant «Big data» what the new ideology of information processing is in and which led to its emergence. The concept «Big Data» belongs to the concepts which do not have clear treatment. Its development and transformation are appropriate to consider in a historical context. Initially was considered that criterion of reference to category Big Data, first of all, is the data flow of bigger 100 Gb a day. However it turned out that it is not enough to refer a product to category Big Data unambiguously. Subsequently as such criterion were used qualitative signs – «a certain number of V»: Volume is significant increase in volume of data in corporate systems; Variety is a variety of formats and structures of the available data; Velocity is the speed of receiving and data processing for satisfaction of inquiry, etc. However such approach does not open this concept fully, and only concerns its separate aspects. In other researches it is possible to meet definition of this concept as impossibility of data processing in the traditional ways. However such definition can already be considered outdated as experts do not refer technology of processing of large volumes of data to the latest any more. Some experts suggest refusing this concept in general. However in the course of training of IT specialists it is not possible to refuse the concept «Big Data» meanwhile. Development of information technology changed our attitude to data and information, and in turn it affected an essence of the concept «Big Data». Now it is understood as special approach rather – the ideology of processing of information applied to processing of big arrays of «crude» data. However, it is possible that in the future this concept can disappear as outdated and irrelevant, large volumes of data and modern approaches of their processing can become usual means and there will be no need to place emphasis on their innovative component.

Key words: professional knowledge; future information technology specialist; information technologies; Bid Data.

Рецензенти: Мещанінов О. П., д-р пед. наук, професор ЧНУ ім. Петра Могили, м. Миколаїв, Україна;
Гришкова Р. О., д-р пед. наук, професор, ЧНУ ім. Петра Могили, м. Миколаїв, Україна.