УДК 330.88.
JEL B49
ORCID ID 0000-0002-4910-5837
ORCID ID 0000-0001-5104-1070
DOI https://doi.org/10.17721/tppe.2024.48.13

Маслов А.О., д.е.н., професор,
КНУ імені Тараса Шевченка
E-mail: a_maslov@ukr.net

Осецький В.Л., д.е.н., професор,
КНУ імені Тараса Шевченка
E-mail: val_osetski@ukr.net

ВЕЛИКІ ДАНІ ТА ОЗЕРА ДАНИХ ЯК СКЛАДОВІ ІНФОРМАЦІЙНОЇ ЕКОНОМІКИ

Статтю присвячено розгляду значення таких складових інформаційної економіки, як великі дані та озера даних. Показано, що в процесі розгортання сучасної інформаційно-технологічної революції, все зростаючого значення в економіці та суспільному житті набувають інформація та інформаційні технології. Особливо важливо це в контексті становлення та розвитку інформаційної економіки. Зазначено, що характерними ознаками інформаційної економіки є її віртуалізація, опрацювання великих обсягів інформації на основі цифровізації даних, використання квантових та нейротехнологій, штучного інтелекту, бездротового зв'язку, промислового Інтернету, сенсорики й робототехніки, цифрових платформ тощо, а ключовими технологіями цифрової трансформації економіки та суспільства є Інтернет речей, обробка структурованих і неструктурованих даних величезних обсягів, штучний інтелект і блокчейн. Показано спільне та відмінне в великих даних та озерах даних, а також їхня роль в зберіганні, обробці та використанні великих масивів інформації.

Ключові слова: *інформаційна революція, інформаційне суспільство, інформаційно-комунікаційні технології, Великі Дані, сховища даних, озера даних, «хмарні» платформи, IT-бізнес, метатехнології.*

Постановка проблеми. В процесі розгортання сучасної інформаційно-технологічної революції все більшу роль в економіці та суспільному житті вцілому відіграє інформація та інформаційні технології. Значення інформації як нового й особливого ресурсу зростає настільки швидкими темпами, що дослідники не встигають з тим, як краще назвати галузь її дослідження – інформаційна економіка, економіка знань, інформаційно-знаннєва, цифрова, або навіть ексабайтова економіка. Не може бути ексабайтової економіки (як інформаційної) без різноманітних інформаційних продуктів, а також без тих, хто їх виробляє, надає та споживає, без відповідних ринків інформаційних благ. Однак усе це не зможе функціонувати, якщо не буде відповідних інформаційних та інформаційно-комунікаційних технологій.

Інформаційна економіка – це новий тип економіки, що використовує новітні інформаційні технології у формуванні якісних державних інформаційних фондів, підприємницьких структур, наукових інститутів, комерційних підприємств, необхідних для формування і підтримки якісного рівня здійснюваних видів діяльності та планових заходів. Завдання створення загальнодоступного масиву інформації в системі господарської,

науково-технічної, освітньої, соціокультурної й інших видів діяльності у суспільстві пов'язане зі створенням сприятливих умов для інноваційного розвитку країни і регіонів.

Для інформаційної економіки характерні такі особливості: підвищення оперативності прийняття рішень; посилення конкуренції на ринках; прискорення динаміки, зростання обсягу бізнес-процесів; підвищення ролі інтелектуальної праці; розширення можливостей підприємств; впровадження електронних платіжних систем і систем електронного документообігу; рух ресурсів через телекомунікаційні мережі; подання товарів/послуг у цифровому вигляді; поява нового типу підприємств – віртуальних (у ряді країн нині немає нормативно-правового документа, який визначає статус віртуального підприємства); розвиток управління на мережевій основі.

Процеси становлення інформаційної економіки та виробництва інформаційних ресурсів ґрунтуються на розвитку інформаційних технологій і супроводжуються:

1) широким впровадження інформаційно-комунікаційних технологій, що дає можливість підприємствам подавати свої товари/послуги у зручному електронно-цифровому форматі, аналізувати діяльність конкурентів, ринкові ситуації і потреби споживачів в он-лайн режимі;

2) зростанням масштабу економічної діяльності, що досягається розміщенням у різних інформаційно-економічних просторах усіх видів економічної діяльності.

Центральне місце в інформаційній економіці посідають *інформаційні технології* (ІТ), або ж *інформаційно-комунікаційні технології (ІКТ)*, які становлять цілеспрямовану сукупність методів, процесів, комунікацій, мереж і програмно-технічних засобів, об'єднаних у технологічний ланцюг, що забезпечує збирання, зберігання, обробку та передачу інформації з метою підвищення ефективності діяльності людей.

Загальновизнано, що характерними ознаками інформаційної економіки є її віртуалізація, опрацювання великих обсягів інформації на основі цифровізації даних, використання квантових та нейротехнологій, штучного інтелекту, бездротового зв'язку, промислового Інтернету, сенсорики й робототехніки, цифрових платформ тощо. При цьому ключовими технологіями цифрової трансформації економіки та суспільства, на думку експертів Організація економічного співробітництва та розвитку (ОЕСР) є Інтернет речей, обробка структурованих і неструктурованих даних величезних обсягів, штучний інтелект і блокчейн.

Останні роки все більшої популярності набувають так звані *хмарні технології*, або *хмарні обчислення (Cloud computing)*. Цей термін став вживатися в світі інформаційних технологій з 2008 р. У свою чергу, широке впровадження хмарних технологій підготувало технічну база для нової хвилі технологій ***Великих Даних (Big Data)*** **та** *озер даних*, які матимуть ще більший вплив на всі види людської діяльності, включаючи науку, бізнес, виробництво, охорону здоров'я, соціальні мережі.

Аналіз останніх публікацій. Нові історико-економічні умови господарювання в межах сучасної, третьої науково-технічної революції і трансформації її під впливом інформаційної революції кінця 1970–80-х років в інформаційно-технологічну революцію, змусили звернутися до дослідження цієї проблематики цілої плеяди відомих науковців – М. Кастельса [1; 2], П. Друкера [5; 6], Р. Ноллана.

Так, М. Кастельс, зокрема, звернув увагу на неймовірне зростання значення нематеріальних факторів виробництва, зокрема інформації та інформаційно-знаннєвих

ресурсів, П. Друкер використовує поняття «інформаційних революцій», Р. Ноллан разробив теорію стадій развитку, що дає можливість глибше засвоїти логіку становлення інформаційної економіки. В межах сучасного інституційно-соціального напряму, з'являються численні наукові публікації визнаних економістів та соціологів Т. Стоуньєра, Ф. Вебстера, Е. Тоффлера, Д. Белла, М. Кастельса та ін.

Поряд із цим, розробці концепцій і моделей інформаційного суспільства та проблем інформаційної економіки присвячено низку фундаментальних досліджень таких видатних західних учених, що стояли у витоків теорії інформаційної економіки, як: Д. Блюменау, Р. Баран, П. Дракер, К. Ерроу, К. Келлі, К. Кларк, Дж. Лодон і К. Лодон, Й. Масуда, Р. Нельсон, Р. Нолан, Д. Норт, М. Порат, М. Портер, К. Сохней, Д. Тапскотт, К. Шеннон, Й. Шумпетер, Р. Ешбі та інші.

Зрозуміло, що питання, пов'язані зі становленням та розвитком інформаційної економіки не могли обійти своєю увагою і вітчизняні дослідники, серед яких В.Д. Базилевич, В.М. Геєць, Г.В. Жаворонкова, Т.В. Затонацька, Л.Г. Мельник, Ю.В. Ніколенко, В.Л. Плескач, В.С. Савчук, Л.І. Федулова, О.І. Черняк, А.А. Чухно та інші, в працях яких аналізуються різноманітні аспекти формування і функціонування постіндустріального, інформаційного суспільства та постіндустріальної сервісно-інформаційної економіки.

Невирішені частини проблеми. Однак, бурхливий розвиток сучасних інформаційно-комунікаційних технологій, експоненційне зростання кількості інформації, яка створюється, обробляється, зберігається та використовується у різноманітних сферах суспільного життя, у тому числі в економіці, потребує теоретичного осмислення цих процесів та явищ. Зокрема, це стосується таких нових явищ, як інтернет-речей, штучний інтелект, хмарні технології, бази даних, блокчейн тощо. На нашу думку, в літературі недостатньо висвітлено роль і значення великих даних та озер даних як складової інформаційної економіки.

Метою статті є обгрунтування теоретико-методологічних засад дослідження еволюції використання інформаційних технологій для зберігання, обробки та використання інформації в межах інформаційної економіки в умовах сучасної інформаційно-технологічної революції, показати стан, особливості та перспективи подальшого використання технологій та платформ великих даних та озер даних.

Результати дослідження. Сучасні ІКТ у вигляді багатьох різноманітних великих і малих обчислювальних пристроїв, якими наповнено сучасний світ, щосекунди генерують гігантські масиви цифрової інформації. Яким чином, ця інформація, що зібрана разом і проаналізована за допомогою сучасних високопродуктивних комп'ютерів, дає можливість отримати якісно нове розуміння того, що вона містить?

Цей перехід кількості накопиченої людством інформації в якості відповіді на поставлене вище запитання сьогодні називають феноменом *«великих даних» (big data)*, і сьогодні це одна одне із найбільш обговорюваних явищ в індустрії інформаційних технологій. Цьому феномену багато уваги приділяють спеціалісти, але, однак, ще дуже мало знають звичайні користувачі інформаційних технологій. Між іншим, ми вже живемо у новій епосі – епосі великих даних, а зміни, які несуть нові інформаційні технології, зачипають життя кожної людини.

«Великі дані» – це маса нових завдань, які стосуються суспільної безпеки, глобальних економічних моделей, недоторканості приватного життя, усталених моральних правил, правових відносин людини, бізнесу і держави. Сьогодні технології обробки «великих даних»,

незважаючи на їхню невідчутність і невагомість, стали силою, яка змінює світ, а розвиток і впровадження технологій «великих даних» може дати унікальні конкурентні переваги бізнесу, допомогти побудувати ефективнішу державу, надати нові можливості людям, і, в кінцевому підсумку – зробити наше життя зручнішим і безпечнішим.

«Великі дані» є свого роду евфемізмом, навмисно наївним скороченням для баз даних, які настільки великі, складні і динамічні, що їх не може обробляти стандартне програмне забезпечення, зокрема, таке як Microsoft Excel.

Gartner, дослідницька і консультаційна фірма в США з питань інформаційних технологій, вперше розробила модель для «великих даних» ще в далекому 2001 році. Її модель «3V» охоплювала обсяг, швидкість і різноманітність даних (англ. – *volume, velocity, variety*). Згідно цієї моделі, *обсяг даних* (*volume*) характеризують – таблиці, файли; трансакції; записи; терабайти; *швидкість даних* (*velocity*), це – потоки; реальний час; наближений час; пакетні дані; *різноманітність даних* (*variety*) характеризують – напівструктуровані; неструктуровані; та структуровані дані.

Gartner формалізувала своє визначення у 2012 р.: «Великі дані – це інформаційні активи великого обсягу, високої швидкості, та/або високої різноманітності, які вимагають нових форм обробки, щоб дозволити здійснювати покращення прийняття рішень, відкриття ідей і оптимізацію процесів» [7].

Є й інші, менш формальні визначення. «Великі дані» також стали скороченням для технологій та аналітичних інструментів, які допомагають управляти «великими даними» та запитувати їх. В останні роки спостерігається розвиток моделей програмування «великих даних» та програмного забезпечення, такого як MapReduce і Apache Hadoop. Невіддільні від «великих даних», ці системи і рішення в даний час є майже синонімами «великих даних».

Протягом наступного десятиліття обсяг, швидкість і різноманітність даних, швидше за все, буде продовжувати зростати. Коли «великі дані» стануть звичайним явищем, цілком імовірно, що з'являться нові визначальні характеристики, як тільки їхні величезні розміри стануть «нормальними».

Будь-який аналіз впливу «великих даних» необхідно розпочати з визначення того, що означають «великі дані». Термін часто використовується, але зазвичай не всі його розуміють. Він стосується, насамперед, величезної кількості даних, які постійно збираються за допомогою пристроїв і технологій, таких, як кредитні картки та картки лояльності клієнтів, Інтернету та соціальних медіа, і все частіше, через датчики WiFi і електронні мітки. Велика частина цієї інформації є неструктурованою – тобто це дані, які не відповідають певній, заздалегідь встановленій моделі даних.

Прикмети інформаційного суспільства неважко помітити всюди: в кожній кишені знайдеться мобільний телефон або смартфон, на кожному столі – комп'ютер, а в робочих кабінетах і по всьому світу – великі ІКТ-системи. Але сама інформація при цьому менше помітна. Через півсторіччя з того часу, як комп'ютери міцно увійшли в життя суспільства, накопичення даних досягло того рівня, на якому відбувається дещо нове і незвичне. Світ не просто завалений небаченою кількістю інформації – ця кількість почала зростати швидше. Зміна масштабу призвела до зміни стану. Тобто, кількісні зміни призвели до якісних. Саме в науках, які вперше зіткнулися зі сплеском даних на початку 2000-х рр., таких як астрономія і геноміка, з'явився термін «великі дані». Тепер ця концепція проникає в усі сфери людської діяльності.

Для «великих даних» немає чіткого визначення. З самого початку ідея полягала в тому, що обсяг інформації виріс настільки, що наявна кількість вже фактично не поміщалася в пам'яті комп'ютера, використовуваної для обробки, тому інженерам довелося модернізувати інструменти для аналізу всіх даних. Так з'явилися технології обробки, зокрема, модель MapReduce компанії Google і її аналог з відкритим вихідним кодом – Hadoop від компанії Yahoo. Саме вони дали можливість управляти набагато більшою кількістю даних, ніж раніше. При цьому важливо, що дані не потрібно було вибудовувати в чіткі ряди або класичні таблиці баз даних.

З'явилися також інші технології обробки даних, які обходилися без колишньої жорсткої ієрархії і однорідності. У той же час Інтернет-компанії, які мали можливість збирати величезні масиви даних і гострий фінансовий стимул для їхнього аналізу, стали провідними користувачами новітніх технологій обробки, витісняючи компанії, які деколи мали на десятки років більше досвіду, але працювали автономно.

Згідно одному з підходів до цього питання, «поняття «великі дані» відноситься до операцій, які можна виконувати виключно у великому масштабі. Це породжує нові ідеї і дозволяє створити нові форми вартості, тим самим змінюючи ринки, організації, відносини між громадянами і урядами, а також багато іншого.

І це лише початок. Епоха великих даних ставить під питання наш спосіб життя і спосіб взаємодії зі світом. Найдивовижніше за все є те, що суспільству доведеться відмовитися від розуміння причинності на користь простих кореляцій: проміняти знання *чому* на *що саме*. Це перевертає сторіччями усталений порядок речей і ставить під сумнів наші фундаментальні знання про те, як приймати рішення і пізнавати дійсність» [4, с. 14].

Беззаперечно, що великі дані знаменують початок глибоких змін. Подібно тому, як свого часу телескоп дав нам можливість осягнути Всесвіт, а мікроскоп – отримати уявлення про мікроби, так і сьогодні нові методи збору і аналізу величезного масиву даних допоможуть розібратися в навколишньому світі з використанням способів, цінність яких ми лише починаємо усвідомлювати. Але справжня революція полягає не в комп'ютерах, які обчислюють дані, а в самих даних і в тому, як ми їх використовуємо.

Сфери використання великих даних можуть бути різними – від науки – до охорони здоров'я, від банківської справи – до Інтернету, але підсумок завжди один: обсяг даних у світі швидко зростає, випереджаючи не лише наші обчислювальні машини, а й уяву.

Великі дані становлять собою *три кроки до нового способу аналізу інформації*, які трансформують нашу уяву про суспільство і його організацію. *Перший крок* – відмова від методу вибірки; *другий* – відмова від точності, і, нарешті, *третій крок* – відмова від сторічних традицій пошуку причинності.

Стосовно першого кроку – відмови від методу вибірки, то якщо, починаючи з XIX ст., суспільство, зіштовхуючись з великими числами, покладалось на метод вибірки, то сьогодні від сприймається як пережиток часів дефіциту інформації, продукт природних обмежень для взаємодії з інформацією в «аналогову епоху». Зрозуміти штучність цих обмежень, які значною мірою сприймались як належне, вдалося лише після того, як високопродуктивні цифрові технології набули значного поширення.

У світі великих даних дослідники можуть проаналізувати величезну кількість даних, а в окремих випадках – обробити всі дані, які стосуються того чи іншого явища, а не покладатися на випадкові вибірки. Використовуючи всі дані, отримується точніший

результат, що дає можливість побачити нюанси, які недоступні при обмеженні невеликим обсягом даних. Великі дані дають особливо чітке уявлення про деталі категорій і сегментів, які неможливо оцінити за допомогою вибірки.

Стосовно другого кроку – відмови від точності, то абсолютизація точності як спосіб мислення характерна для середовищ «малих даних». Коли можливість вимірювання обмежена, рахуються тільки найважливіші показники, і намагання отримати точне число абсолютно доцільне. Вимірювалось так мало показників, що необхідно було якомога точніше порахувати все записане.

Чим більший масштаб, тим менше значення відіграє точність. Якщо точність вимагає детальної перевірки даних, і вона підходить для невеликих обсягів даних, а в деяких випадках навіть необхідна, то *у світі великих даних сувора точність неможлива* (!), *а деколи і не бажана*. Якщо доводиться оперувати даними, більшість яких постійно змінюється, абсолютна точність відходить на другий план. Більшість даних невпорядковані, далеко не всі однакої якості і розкидані по незчисленних серверах по всьому світу. Коли маєш справу з великими даними, то, зазвичай, доводиться вдовольнятися загальними уявленням, а не розуміти в усіх деталях. Зрозуміло, що від точності як такої ніхто не відмовляється, а лише зменшується прихильність до неї, а те, що втрачається із-за неточності на мікрорівні, дає можливість робити відкриття на макрорівні.

Стосовно третього кроку – відмови від сторічних традицій пошуку причинності (детермінованості), то у своїй діяльності люди звикли в усьому шукати причини, навіть якщо встановити їх не так просто. Але у світі великих даних ми більше не повинні надавати зумовленості того значення, як раніше. Замість цього дослідники можуть знаходити кореляції між даними, які відкривають нові неоцінимі знання. Зрозуміло, що кореляції не можуть дати точної відповіді на запитання, *чому* відбувається та чи інша подія, однак попереджують про те, *якого вона роду*, і у більшості випадків цього повністю достатньо. Питання не в тому *чому*, а в тому *що*.

У світі великих даних не завжди потрібно знати причини, які стоять за тими чи іншими явищами та подіями, а краще дати можливість даним говорити самим за себе. Дослідникам не потрібно більше обмежуватися перевіркою невеликої кількості гіпотез, старанно сформульованих задовго до збору даних. Даючи можливість даним «говорити», можна виявити кореляції, про існування яких до цього навіть не підозрювали. Зокрема, у зв'язку з цим хедж-фонди аналізують записи в Twitter, щоб прогнозувати роботу фондового ринку, а Amazon і Netflix рекомендують продукти, опираючись на величезну кількість взаємодій користувачів зі своїми сайтами. У свою чергу, Twitter, LinkedIn і Facebook вибудовують «соціальні графи» (3D моделі) відносин користувачів для вивчення їхніх уподобань.

Великим кроком на шляху від «аналогової епохи» до ефективнішого управління даними стала поява *оцифрування* – переводу аналогової інформації у доступну для читання на комп'ютерах, що спрощувало і здешевлювало її зберігання і обробку. Все це значно підвищило ефективність, а те, на що раніше витрачалися роки збору і обчислення, тепер виконувалося за декілька днів, або годин.

Великі дані вносять суттєві корективи в характер бізнесу, ринків і суспільства. Якщо в XX ст. особливого значення надавалось не фізичній інфраструктурі, не землі і заводам, а нематеріальним активам, інтелектуальній власності, то сьогодні суспільство йде до того, що новим джерелом цінності стане не потужність комп'ютерного обладнання, а отримані ним

дані та спосіб їхнього аналізу. Дані, особливо, *великі дані*, перетворюються у важливий корпоративний актив, життєво важливий економічний вклад і основу нових бізнес моделей.

Поява «великих даних» як бізнес-питання можна розглядати як більш загальне вираження переходу від одних технологій до інших. Це також свідчить, про те, що, коли вплив ІКТ на бізнес буде зрозумілішим, термін «великі дані» вийде з використання і буде замінений словами, що конкретніше описуватимуть технологічні процеси та програми.

«Великі дані» мають потенціал для перетворення майже кожного аспекту бізнесу – від досліджень і розробок до продаж і маркетингу та управління ланцюгом поставок, а також для нових можливостей у зростанні.

Як було видно вище з визначення фірми Gartner, «великі дані» мають потенціал для підвищення цінності. Компанії використовують бізнес-аналітику та інструменти інтелектуального аналізу даних для підвищення ефективності, визначення нових можливостей, надання клієнтам кращих продуктів і послуг, і для передбачення майбутніх моделей поведінки. Не дивно, що слово «цінність» тепер підноситься як нове «V» в моделі Gartner.

Можливості великих даних – не тільки для великого бізнесу. Наявність онлайнових і «хмарних» платформ, таких як *Google Analytics* і *Tableau*, означає, що малі та середні підприємства (МСП) можуть брати ідеї з «великих даних» без істотних капітальних вкладень. Не обтяжені великими застарілими системами, ці підприємства іноді здатні перестрибнути старі технології, і почати безпосередньо з «великих даних».

Бізнес-потенціал настільки великий, що «великі дані» тепер часто називають «новою нафтою», інформаційним еквівалентом природного ресурсу, який сформував економіку в XIX і XX сторіччях. Ця аналогія має певну цінність, але є в ній і недоліки. На відміну від нафти, «великі дані» є в майже необмеженій кількості, і до того ж, вони «поновлювані». Їх обсяг росте щороку, і на кілька порядків. Десять років тому люди говорили про гігабайти даних, тепер вони говорять про терабайти, що в тисячу разів більше.

Коли люди в бізнесі думають про «великі дані», то вони думають, повертаючись до визначення компанії Gartner, про швидкість, а також про обсяг і різноманітність. «Великі дані» – швидкі дані: збираються швидко, передаються швидко і обробляються швидко.

Навчені структурувати, збирати та аналізувати фінансову інформацію економісти, зокрема, бухгалтери та фінансові фахівці, можуть застосувати свої ключові навички для не фінансових та інших баз даних – і, найголовніше, допомагати робити «великі дані» меншими і більш структурованими.

Але, отримати вигоди від аналізу і використання великих даних буде нелегко. Бази даних можуть як знищити цінність, так і створити її, а тому – їх необхідно ефективно і вміло скеровувати, і вони вимагають значних інвестицій від компаній.

Таким чином, великі дані – важливий крок людства у постійному прагненні виміряти й пізнати навколишнє середовище. Те, що раніше не можливо було виміряти, зберігати, аналізувати і поширювати, знаходить своє вираження у вигляді даних. Використання величезних масивів даних замість їхньої малої частини і вибір кількості в ущерб точності відкривають шлях до нових способів розуміння світу. Все це підштовхує суспільство відмовитися від освяченого сторіччями пошуку причинності і у більшості випадків користуватися перевагами кореляцій.

Але, тим не менше, тенденції розвитку ІТ, що намітилися на сучасному етапі, свідчать, що ні подальше вдосконалення давно відомих технологій, ні лідерство у пропозиції найпередовіших новинок не здатні дати достатній імпульс ринковій експансії галузі в цілому та вивести її на нові рубежі, і що таким імпульсом може стати створення систем, додатків, платформ і стандартів, які охоплюватимуть усе суспільство, і провідна роль у цьому належатиме споживачу.

Якщо споживач відчує необхідність активніше оволодівати новими сферами використання ІТ, то бізнес постачальників ІТ буде розвиватися. Якщо ж споживач не захоче чи не зможе рухатися в цьому напрямі, зрозуміло, що ІТ-бізнесу майже гарантовано загрожує стагнація. Ця нова конфігурація взаємозалежності галузі та споживача і визначатиме динаміку розвитку ІТ у найближчий період.

Слід зазначити, що в епоху глобалізації світової економіки інформаційні технології починають відігравати надзвичайно важливу основоположну роль, що виникає тенденція виділяти їх як п'ятий фактор виробництва поряд із відомими працею, капіталом, природними ресурсами і підприємництвом. Виробнича функція в багатьох наукових дослідженнях починає включати в себе цей фактор виробництва, що знаходить своє відображення в моделях економічного росту і тих висновках, до яких вони приходять.

У найпростішому вигляді математичну ідею ендогенності технології можна виразити так:

$$W = PT^n, \qquad (4.1)$$

де W – багатство; P – природні ресурси (земля, робоча сила, корисні копалини); T – технологія; n – ступінь впливу технічних досягнень. Існують і складніші багатофакторні моделі, які наголошують на зростаючій значимості технологічного й інформаційного факторів у процесі виробництва. Так, американські економісти М. Боскін і Л. Лау на основі таких досліджень доходять до висновку, що технічний прогрес у розвинутих країнах забезпечує в середньому 40 % економічного зростання, а решта – 60 %, припадають на працю і капітал, при цьому під технічним прогресом розуміють, у першу чергу, упровадження нових інформаційних технологій.

Подвійність інформаційного ресурсу пов'язана з тим, що він, з одного боку, є у великій кількості, а з іншого – є обмеженим, якщо йдеться про якісь окремі види інформації. Інформації завжди дуже багато, і кожен вид економічної діяльності виробляє її більше, ніж в змозі спожити. У цьому випадку мається на увазі відома теза про величезний обсяг інформації і можливостях її копіювання та збільшення у зростаючих масштабах.

Однак поряд із цим інформація є обмеженим ресурсом, відповідно до традиційного її трактування у стандартних курсах з мікроекономіки, зокрема, коли йдеться про стратегічну й комерційну інформацію. Тому у зв'язку з поширенням нових інформаційних технологій і новими можливостями, які вони надають у використанні інформаційного простору, проблема захисту прав на інтелектуальну власність постає особливо гостро.

Сьогодні величезного значення для всієї світової економіки набуває найрізноманітніша інформація: політична, економічна, фінансова, науково-технічна. Зростає масштаб економічної діяльності, розширюються межі світових ринків, зростає значення реального і грошового секторів, і для того, щоб правильно приймати рішення в короткостроковому і в довгостроковому плані, необхідно мати постійний доступ до зростаючого обсягу різноманітної інформації. Виникає не лише світовий ринок інформаційних технологій, що

швидко розвивається, а й ринок міжнародної інформації у вигляді досить широкого спектру баз даних, у тому числі, *великих даних (big data)*, про що йшлося вище, а інформація стає цінним товаром, який до того ж потребує постійного оновлення.

Інформаційні технології стали потужною галуззю світової економіки, що інтенсивно розвивається. Темпи зростання частки галузей, що безпосередньо пов'язані з виробництвом і використанням знань, у ВВП США, на кінець 90-х рр. становили не менше 60 %. Жодна розвинена країна не може відмовитися від того, щоб стати частиною інформаційного простору. Експерти вважають, що технічно Інтернет зможе поєднати у перспективі до 600 млн. комп'ютерних мереж [3, с. 330]. Обсяги витрат на нові технології, які здійснили постіндустріальні країни за останнє десятиріччя XX ст., часто перевищують витрати на придбання основних виробничих фондів. Так, у 1991 р. витрати американських компаній на придбання інформації й інформаційних технологій, що досягли 112 млрд. $, перевищили витрати на придбання основних виробничих фондів, які становили в тому ж році 107 млрд $, а вже до 1996 р. інвестиції в інформацію й інформаційні технології зросли майже вдвічі, тоді як витрати на основні виробничі фонди майже не змінились.

Можна стверджувати, що формуються так звані метатехнології – мережеві комп'ютери, організаційні технології, технології комунікаційного плану, які є технологічною основою сучасного інформаційного суспільства. Наявність таких технологій визначає ступінь конкурентоспроможності економіки в цілому, а їхнє використання ставить країну в пріоритетне становище.

Про масштаб технологічних змін, що відбуваються в межах інформаційної революції свідчить те, що ринкова капіталізація компаній, акції яких не можна придбати в системі NASDAG (Nation Association of Securities Dealers Automated Quotation – Автоматизовані котирування Національної асоціації дилерів із цінних паперів), зросли від 386 млрд $ до більше ніж 5 трлн $ за період із 1989 р. по 1999 р. [8]. Ринкова капіталізація шести провідних американських компаній, які відіграють велику роль у розвитку і поширенні Інтернет (*Microsoft*, *Cisco*, *Sun Microsystems*, *Dell*, *MCWorldcom*, *Charles Schwab*) на початок XXI ст. становила понад 1,6 трлн $.

Швидкими темпами зростає кількість користувачів Інтернет, а це величезний потенціал для розвитку електронного бізнесу й найрізноманітніших форм електронного зв'язку, де починає домінувати саме споживач інформаційних технологій над його виробником.

Не може бути інформаційної економіки без різноманітних інформаційних продуктів, без тих, хто їх виробляє, надає та споживає, без відповідних ринків інформаційних благ. Однак усе це не в змозі буде функціонувати, якщо не буде необхідних інформаційних та інформаційно-комунікаційних технологій.

Поряд з *Великими Даними (Big Data)* є також *Озера даних*. *Озеро даних* – це система, або централізоване сховище даних, яке дає можливість зберігати всі структуровані, напівструктуровані, неструктуровані та двійкові дані у своєму природному, тобто, власному необробленому форматі.

Зокрема, *структуровані дані* охоплюють таблиці із системи управління базами даних (СУБД); *напівструктуровані дані* включають файли CSV, XML, журнали, JSON тощо; *неструктуровані дані* включають PDF-файли, текстові документи, текстові файли, електронні листи тощо; і *двійкові дані* можуть включати аудіо, відео, файли зображень.

Термін *озеро даних* придумав Джеймс Діксон, технічний директор у Пентахо. Він визначає маркет даних (підмножина сховища даних) як подібний до пляшки з водою, наповненої очищеною, дистильованою водою, упакованої та структурованої для прямого та простого використання. Інакше кажучи, це аналог водойми в її природному вигляді. Дані надходять з потоків (різні бізнес-функції та/або системи джерел) до озера.

Користувачі озер даних мають доступ до озера для різноманітних цілей: аналізу, дослідження, збору зразків та занурення. Так, як вода в озері задовольняє різні потреби людей (риболовля, катання на човнах, забезпечення питною водою тощо), так само і архітектура даних озера буде служити різноманітним цілям. Зокрема, вчений-дослідник може використовувати озеро даних для дослідження даних та створення гіпотези, що дасть можливість аналітикам даних аналізувати дані та виявляти закономірності. Він надає можливість діловим користувачам та зацікавленим сторонам досліджувати дані.

За допомогою озера даних суб'єкт господарювання може зберігати все своє підприємство в одному місці без потреби попередньо структурувати дані, безпосередньо виконувати тут різні типи аналітики, зокрема, машинне навчання, аналітику в реальному часі, локальний рух даних, переміщення даних у реальному часі, інформаційні панелі та візуалізацію. Суб'єкт господарювання зберігає в озері даних всі дані в оригінальній формі і передбачає, що їх аналіз відбудеться за потреби пізніше. Це також дає можливість аналітикам розробляти звіти та представляти їх бізнесу. У сховищі ж даних навпаки – дані упаковують для чітко визначених цілей.

Для чого ж потрібні озера даних, якщо є сховища даних? Єдина подібність між Озером даних і сховищем даних полягає в тому, що обидва вони є сховищами даних. Насправді ж вони різні і мають різні цілі та варіанти використання.

Якщо коротко, то відмінності між ними полягають в *аналітиці*, *даних*, *користувачах*, *доступності*, *схемах*, *архітектурі*, *призначенні*, *зберіганні*, *підтримці типів даних*, *безпеці* та *швидкості*.

Так, в *аналітиці*, якщо сховище даних можна використовувати для бізнес-аналітики, візуалізації та різноманітної звітності, то озеро даних можна використовувати як для машинного навчання, так і для профілювання даних для виявлення даних та прогнозного аналізу. Стосовно *даних*, якщо в сховищі даних знаходяться лише ті дані, що обробляються та вдосконалюються, тобто структуровані дані, необхідні для складання звітів та вирішення конкретних бізнес-проблем, то в озері даних зберігаються всі вихідні дані. Дані можуть бути структурованими, неструктурованими або напівструктурованими. Можливо, деякі дані в озері даних навіть ніколи не будуть використані.

Відмінність в *користувачах* полягає в тому, що, як правило, користувачами сховища даних є бізнес-професіонали, оперативні користувачі та бізнес-аналітики, а користувачами озера даних є науковці та розробники даних. *Доступність* до даних відрізняється тим, що якщо у сховищі оновлення даних є складнішою та витратнішою операцією, оскільки сховища даних структуровані за проектом, то озеро даних є доступнішим, і таким, яке легко та швидко оновлюється, оскільки не має жодної структури.

Якщо в озері даних *схема* орієнтована на запис за допомогою розробленого до провадження DW (*Data Warehouse/Mart*), то в інформаційному сховищі схема спрямована на читання того, що написано під час аналізу. Відмінність *архітектури* полягає у її пласкому стані в озері і ієрархічності у сховищі.

Стосовно *призначення* необроблених даних, які зберігаються в озерах даних, то воно не є фіксованим або невизначеним. Дані можуть надходити до озера даних, передбачаючи їх певне використання у майбутньому, або ж для того, щоб мати їх під рукою. В озері даних зберігаються менш організовані та менш відфільтровані дані.

У сховищі ж даних зберігаються оброблені дані, які мають конкретне призначення. Інформаційне сховище (DW) організувало та відфільтрувало дані. Отже, сховище вимагає менше місця для зберігання, ніж озеро даних.

Озера даних, на відміну від інформаційних сховищ, призначені для недорогого *зберігання*. Апаратне забезпечення озера даних суттєво відрізняється від апаратного забезпечення сховища даних. Використання озером даних готових серверів у поєднанні з дешевими сховищами робить його досить економічним та дуже масштабованим до терабайт та петабайт. Це робиться насамперед для збереження всіх даних в озері даних, щоб користувач міг у будь-який момент повернутися, щоб зробити необхідний аналіз. В інформаційному ж сховищі *зберігання* даних є витратним для великих обсягів даних завдяки наявності дорогого дискового сховища. І щоб зробити його високоефективним, то для економії простору модель даних спрощується, і тут будуть зберігатися лише ті дані, які дійсно необхідні для прийняття бізнес-рішень.

Стосовно відмінностей у *підтримці типів даних*, то озеро даних дуже добре підтримує нетрадиційні типи даних – журнали серверів, дані датчиків, активність у соціальних мережах, текст, зображення, мультимедіа тощо, а усі дані зберігаються незалежно від джерела та структури. Сховище ж зазвичай складається з таких даних, які отримані з транзакційних систем, і не дуже добре підтримує нетрадиційні типи даних, зберігання та споживання яких може бути дорогим та складним.

Безпека озер даних та сховищ дуже схожі і однаково знаходяться на стадії «дозрівання», до того ж, озера даних є порівняно новою концепцією, ніж сховище даних. *Швидкість* в озері даних, на відміну від сховища, є високою, що дає можливість налаштовувати та переналаштовувати дані за необхідності. Сховище даних є менш рухливим з фіксованою конфігурацією.

Висновки. Таким чином, розвиток інформаційно-комунікаційних технологій, широкі можливості та переваги хмарних технологій у поєднанні з використанням Великих даних і озер даних, відкривають подальші перспективи для цифровізації не лише економіки, а й різноманітних сфер суспільного життя.

Перспективи подальших досліджень. Подальший розвиток інформаційної економіки у міру розгортання сучасної інформаційно-технологічної революції, зростання ролі та значення інформації як нового ресурсу та різноманітних інформаційних продуктів, висуватиме все більші вимоги до інформаційних та інформаційно-комунікаційних технологій з питань зберігання, обробки та використання величезних масивів даних інформації як в економіці, так і в інших різноманітних сферах суспільного життя. Кількість інформації, яку людство виробляє і яку використовує, зростає надзвичайно швидко. Якщо нещодавно ми вимірювали інформацію в кілобайтах (10^3), мегабайтах (10^6), гігабайтах (10^9), то вже сьогодні в терабайтах (10^{12}), петабайтах (10^{15}) та ексабайтах (10^{18}) (у двійковій системі ексбібайт – 2^{60}. Так, за даними досліджень International Data Corporation (IDC) – міжнародної компанії, постачальника маркетингових досліджень, консультаційних послуг, організатор конференцій в області інформаційних технологій, телекомунікацій і споживчої електроніки, загальний

обсяг цифрової інформації, яка містилася на земній кулі у 2006 році, становила всього 161 Ексабайт, то вже у 2007 році було виявило 281 Ексабайт створеної і скопійованої інформації.

Згідно твердження аналітиків з Каліфорнійського університету людству знадобилось 300 тисяч років, щоб створити перші 12 Ексабайт інформації, проте другі 12 Ексабайт були створені всього за два роки. Згідно заяви CERN (Європейська організація з ядерних досліджень – ЦЕРН) на 29 вересня 2023 року сумарна місткість накопичувачів у її дата-центрах перевищила знакову позначку 1 Ебайт. Це свідчить про те, що ще навіть не всі звикли до назви ексабайтова (ексбібайтова) економіка, як людство вже стоїть на порозі нової епохи – зетабайтової (10^{21}) (зебібайтової – 2^{70}), йотабайтової (10^{24}) (йобібайтової – 2^{80}) чи наступних економік, зі ще на багато порядків більшої кількості інформації, коли може виникнути ситуація, що існуючі сховища та IT перестануть нас задовольняти, і виникне потреба в розробці та використанні принципово нових.

Література.

1. Кастельс М. Інтернет-ґалактика. Міркування щодо Інтернету, бізнесу і суспільства ; пер. з англ. Київ : «Видавництво «Ваклер»» у формі ТОВ, 2007. 304 с.
2. Кастельс М., Хіманен П. Інформаційне суспільство та держава добробуту. Фінська модель ; пер. з англ. Київ : «Видавництво «Ваклер»» у формі ТОВ, 2006. 256 с.
3. Castells M. The Information Age: Economy, Society and Culture. Vol. 1-3, Maiden (Ma) / Castells M. – Oxford, 1996–1998.
4. Mayer-Schönberger Viktor, Cukier Kenneth. (2013) Big Data: A Revolution that Will Transform how We Live, Work, and Think. Boston; New York: Houghton Mifflin Harcourt, 242 s.
5. Drucker P. Post-capitalist society. New York: Routledge, 2011. 216 p.
6. Drucker P.F. The New Society: The Anatomy of Industrial Society. New Jersey: Harper & Brothers, 2003. 363 p.
7. The Importance of 'Big Data': A Definition / http://www.gartner.com/id=2057415
8. http://www.nasdaq.com

References

1. Kastels M. (2007). Internet-galaktyka. Mirkuvannia shchodo Internetu, biznesu i suspilstva [Internet galaxy. Considerations regarding the Internet, business and society] ; per. z anhl. Kyiv : «Vydavnytstvo «Vakler»» u formi TOV, 304 s. (in Ukrainian).
2. Kastels M., Khimanen P. (2006) Informatsiine suspilstvo ta derzhava dobrobutu. Finska model [Information society and the welfare state. Finnish model] ; per. z anhl. Kyiv : «Vydavnytstvo «Vakler»» u formi TOV, 256 s. (in Ukrainian).
3. Castells M. (1996–1998) The Information Age: Economy, Society and Culture. Vol. 1-3, Maiden (Ma) / Castells M. – Oxford.
4. Mayer-Schönberger Viktor, Cukier Kenneth. (2013) Big Data: A Revolution that Will Transform how We Live, Work, and Think. Boston; New York: Houghton Mifflin Harcourt, 242 p.
5. Drucker P. (2011) Post-capitalist society. New York: Routledge, 216 p.
6. Drucker P.F. (2003) The New Society: The Anatomy of Industrial Society. New Jersey: Harper & Brothers, 363 p.
7. The Importance of 'Big Data': A Definition / http://www.gartner.com/id=2057415
8. http://www.nasdaq.com

Anatolii Maslov, Doktor of Sciences (Economics), Professor
Taras Shevchenko National University of Kyiv
Valerii Osetskyi, Doktor of Sciences (Economics), Professor
Taras Shevchenko National University of Kyiv

BIG DATA AND DATA LAKES AS COMPONENTS OF THE INFORMATION ECONOMY

The article is devoted to consideration of the importance of such components of the information economy as big data and data lakes. It is shown that in the process of unfolding the modern information technology revolution, information and information technologies are gaining increasing importance in the economy and social life. This is especially important in the context of the formation and development of the information economy. It is noted that the characteristic features of the information economy are its virtualization, the processing of large volumes of information based on the digitization of data, the use of quantum and neurotechnologies, artificial intelligence, wireless communication, the industrial Internet, sensors and robotics, digital platforms, etc., and the key technologies of the digital transformation of the economy and society are the Internet of Things, processing of structured and unstructured data of huge volumes, artificial intelligence and blockchain. Key commonalities and differences in big data and data lakes are shown, as well as their role in storing, processing and using large amounts of information.

In particular, it states that "Big Data" has become shorthand for the technologies and analytical tools that help manage and query "big data" and represent three steps to a new way of analyzing information that will transform how we think about society and its organization: the first step - rejection of the sampling method; the second is the rejection of precision, and, finally, the third step is the rejection of centuries-old traditions of the search for causality. It is shown that along with big data, which stores processable and structured information, there are also data lakes, which are a system or centralized data repository where all structured, semi-structured, unstructured and binary data are stored in their natural, i.e., their own raw format. The further development of the information economy and the growth of the role and importance of information will lead to new requirements for the technologies of processing, storage and use of information resources.

Key words: *information revolution, information society, information and communication technologies, Big Data, data warehouses, data lakes, "cloud" platforms, IT business, meta-technologies.*

8.04.2024 року