

*Р.А. Тараненко*  
*(Національний технічний університет України «Київський політехнічний*  
*інститут імені Ігоря Сікорського», Україна)*

### **Якість інформації: фундаментальні висновки в контексті великих даних та машинного навчання**

*Низька ефективність великих даних (Big Data), машинного навчання (Machine learning - ML), штучного інтелекту (Artificial Intelligence - AI), онтології (Ontology), індустрії 4.0 (Industry 4.0 - I-4.0) досягли критичного рівня. Необхідність розгляду всеохоплюючих уявлень має мету отримати цілісні аналітичні формальні підходи, основою формування яких можна розглядати якість інформації (Information Quality - IQ).*

#### **Основи проблематики.**

Сьогодні загально визнаним фактом активного розвитку інформаційних технологій і систем (ITiC) є їх комплексне і всеохоплююче поширення, значення якого, з огляду на досягнення його історично безпрецедентного стану, все ще не зрозуміле і не відображено на належному рівні, включаючи фундаментальні уявлення. Очевидно, що колізії в законодавстві; катастрофа європейської ракети Ariane 5 через 37 секунд польоту, як результат 10 річних розробок і 7 млрд. дол.; 10 млрд. фунтів стерлінгів на будівництво Євротунелю під Ла-Маншем як перевищення запланованих витрат в два рази; помилки складних військових комплексів, що ледь не спровокували ядерну війну; помилки медичного обладнання або складності копіювання, перенесення і запуску різних складних технологій та ін. не є поодинокими, а відображають закономірну складову всієї життєдіяльності суспільства, що проявляються великими втратами, тісно пов'язані з проблематикою IQ. В цій проблематиці, в широкому розумінні, можна виділити дві домінуючі тенденції - отримання конкурентних переваг і безпеку. Характеристичним прикладом катастрофічного стану IQ в будь-якій області є підприємства [1, р. 76]: 60% підприємств страждають від неякісних даних; і як правило 10-30% даних в організаційних базах даних не точні; а помилки промислових даних можуть досягати рівня 75%; 70% виробничих замовлень оцінюються з даних поганої якості; 40% даних в базах даних управління кредитними ризиками були неповними; від 50% до 80% даних про зловживання оцінені як неточні, неповні і неоднозначні. Безліч заходів щодо поліпшення IQ, зробивши ефект окремих поліпшень навіть при їх суттєвому значенні не створили якісної революції. Тому проблема визначення IQ, що явно чи приховано присутня в будь-якому інформаційному поданні, стала одним з найбільш актуальних фундаментальних наукових напрямків.

На сьогодні все ще залишається актуальною думка провідних фахівців в області якості інформації Річарда Ванга, Діана Стронг та ін. [2, 3] які вважають, що «якість інформації - неточна наука в термінах її оцінки та тестових стандартах. Хоча різні аспекти якості інформації були досліджені, все ще є

критичною потреба в методології, яка оцінить, як добре організації розвивають інформаційні продукти і надають інформаційні послуги споживачам». Автори найбільш вдало виділили чотири основні підходи у визначенні якості: 1) перевага; 2) цінність; 3) відповідність специфікації; 4) відповідність або перевищення очікувань споживача.

Наявність множини дослідницьких робіт в області IQ представлено великою кількістю публікацій з тематики, проте вони все ще викладають досить роз'єднані уявлення предметної області і констатують факт відсутності єдиного підходу, розглядаючи проблематику різноманітним раніш ізольованими вузькоспеціалізованими різними методологічними підходами [4, р. IX].

Однак, все ще не маючи єдиної думки у вирішенні проблематики IQ, виходячи з поєднання цілого комплексу нових можливостей сучасності, пов'язаних з необхідністю вирішення завдань масштабів дуже великих баз даних (Very Large Data Base – VLDB) та екстремально великих баз даних (Extremally Large Data Base - ELDB) проблематика IQ вийшла на рівень якісно нового значення і важливості, отримавши вираження великих даних  $N = "BCE"$  [4, с. 35]. Тому, нові напрямки Big Data, ML, AI, Ontology, I-4.0 і ін. розкривають нові галузі знань, особливості вирішенні цільових завдань в яких, можна висловити в головних попередніх висновках:

- нові вимоги до пошуку цільових рішень - нова міра складності, спрощено виражена в даних  $N = "BCE"$ ;

- необхідність зіставлення різних категорій і видів інформації, в отриманні їх кількісно-якісних оцінок, для вирішення цільових завдань, пред'являє якісно нові вимоги до їх розгляду з єдиних позицій;

- IQ, розглядаючи крім методів і критерії їх оцінок виявилася в стані прикордонної зони подальшого якісного розвитку не тільки в можливостях прийняття рішень, але і всіх фундаментальних уявлень і прикладних підходів, вивівши їх в область можливостей «що здатні змінити світ» [4];

- поєднання безлічі чинників, що формують нові напрямки Big Data, ML, AI, Ontology, I-4.0 та ін. нові напрямки розкрили проблематику обмеженості традиційних класичних теоретичних підходів і розглядають їх, як наближені, спрощені, роз'єднані, фрагментарні і часткові, які можна розглядати як міждисциплінарні або інтегровані;

- необхідно констатувати факт, що поєднання безлічі чинників нових можливостей призвело до трансформації в деякий новий - цілісний стан, який завдяки новому рівню складності не було передбачено в рамках класичних теоретичних підходів, і тому склалася парадоксальна ситуація випередження теоретичних підходів над прикладними в області інформаційних технологій і систем;

- традиційні, класичні фундаментальні підходи досягли граничного рівня їх застосування, ціною якого є низька ефективність;

- складністю подальшого розвитку формальних теоретичних підходів є невизначеність, як потреба розв'язку протиріч та парадоксів;

- сучасні економічні та інші умови, які набули «всеосяжного» значення ставлячи будь-який «виробничий процес» в конкурентні умови можливої реалізації лише в умовах «економіки малих витрат», сформували запит, на

пошуки нових «організаційно-виробничих схем», що реалізують складні системи з емерджентними властивостями, що задовольняють перерахованим вище вимогам.

### **Постановка задачі вирішення проблематики IQ.**

Вимога розгляду предметної області на новому рівні складності поставила перед нами завдання пошуку нових аналітичних підходів, в рамках яких буде вирішена множина типових задач, що досягли критичного значення важливості у визначенні IQ. Наша задача, розглянути проблематику з позицій можливості створення єдиного аналітичного простору рішень, в рамках якого ми отримаємо інструменти для співставлення різних видів та категорій інформації у оцінці IQ.

### **Шляхи вирішення.**

До сьогоднішнього дня в основі побудови всіх наукових методів і підходів, включаючи математику були проєкційні принципи спрощення для отримання наближених результатів, в рамках деяких прийнятих допустимих значень, що дозволяли оцінювати їх придатність. Такий підхід дозволяв розглядати будь-яку предметну область в рамках методології трьох класичних категорій сучасної науки - **фрагментарні знання, часткові знання або фактологічні і методологія, що приймається як цілісне знання**. Індуктивний принцип сучасної фундаментальної методології дозволяв розглядати будь-яку предметну область в рамках схеми «ДО - ФАКТ - ПІСЛЯ», де «ФАКТ» - область пошуку рішень. Тому, сучасна триєдина наукова концепція в такому підході сконцентрована на **системних уявленнях (СУ), передумовою («ДО»)** яких є **елементні уявлення (ЕУ)**, а фактором розвитку («ПІСЛЯ») служать **уявлення теорія управління (УП)**. Отримано методологічна схема **ЕУ - СУ - УП**. У нових напрямках відбувся перенос основного акценту до теорії управління, про що свідчить збільшення кількості публікацій [6]. Склалася схематична ситуація **ЕУ - СУ - УП - ЦУ**, де **ЦУ - цілісні уявлення**.

Однак сучасна наука спирається на думку, що цілісність не можливо формалізувати [7]. Тому низька ефективність нових методів є наслідком відсутності чітко визначених критеріїв якості, як наслідок не визначеності. Неможливо сконцентруватися на тому, що не визначене, тим більше, якщо інтереси нових напрямків стосуються нових емпіричних областей уявлень, які з огляду на їх специфіку нерозв'язні в рамках окремих підходів. Головним критерієм досягнення успішного вирішення завдань, поставлених перед новими напрямками, став перехід від спрощених класичних методів до всеохоплюючих, що однозначно говорить про тенденцію переходу від проєкційних методів до максимально адекватного відображення явищ і процесів. На відміну від більшості публікацій, пов'язаних з предметною областю теорії управління, що все ще тягнуть до системних підходів - Кібернетика 2.0 [6], були розроблені методологічні принципи, схематично представлені на мал. 1 та мал. 2, які можна вважати підставою для розвитку формальних цілісних підходів.

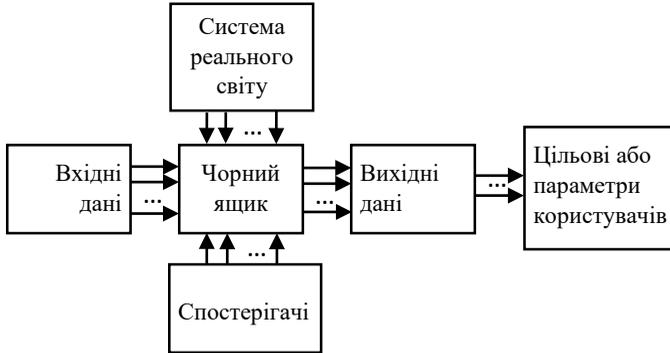
Спірні підходи «3 V або 4V або 5V» в Big Data що розглядають наступні категорії:

- **3V (Volume-Velocity-Variety) - (Обсяг – Швидкість - Різноманітність);**

- 4V (... + Value) – (... + Цінність);
- 5V (... + Veracity) – (... + Достовірність)

трансформуємо у простір виміру 4D в рамках якого проводиться пошук рішень:

- 1-ий вимір – Data (дані);
- 2-ий вимір – Volume (обсяг);
- 3-ий вимір – Velocity (швидкість);
- 4-ий вимір – Knowledge (знання) – впорядковані представлення, отримані на основі цілісного підходу.



Мал. 1. Спрощене схематичне представлення цілісного підходу



Мал. 2. Ієрархічна піраміда сучасних інформаційних технологій

### Висновки.

Фахівці стали все гірше розуміти, що відбувається в складних системах та явищах. Ні один простий метод не дозволяє отримувати «достатньо якісні результати» та не надає будь яких переваг і тому доводиться застосовувати складні. Отримання цільових рішень в складних «всеохоплюючих» представленнях явищ та процесів засноване на інформаційних даних. Однак сама трактовка інформації до сих пір неоднозначна. Тому інформаційні технології та системи не можуть ефективно працювати з інформацією - вони її не розуміють. Незважаючи на «нові можливості революції, яка змінить те, як ми живемо, працюємо та мислимо»[4] виникли складності. В нових, складних технологіях, якісні результати важко отримати, а отримання рішень пов'язане з суттєвими витратами та низькою ефективністю, що перетворює досягнення

практичних результатів в смну ручну працю, унеможливаючи автоматизацію більшості її складових. Серед множини рішень найбільш ефективними показали себе складні комбіновані, так звані ансамблі, для ефективного застосування яких потрібно обрати відповідний підхід. Застосування Big Data та інших новітніх підходів наближає нас до можливостей найбільш адекватного представлення явищ та процесів у напрямку представлення їх аналітичних цілісних моделей. Потреба у засобах оцінки IQ, що окрім методів розглядають і критерії оцінок очевидна. Практичні підходи, які засновані на методології цілісності демонструють свою ефективність. Наприклад, на основі принципів розширення поняття інформації до меж, що описують його показники, як невід'ємну його станів » поданому у роботі [5], було побудовано систему Corezoid (Corezoid - це хмарна операційна система для створення ІТ-рішень з використанням методів автоматного програмування з явним виділенням станів. Початок розробки 2005 р. Представлена суспільству 2015 р). Corezoid дозволяє візуально створювати та обробляти будь-які алгоритми та процеси, а також керувати технікою та програмним забезпеченням будь-якої складності. Приватбанк України заявляє, що автоматизована обробка СМС повідомлень системою Corezoid дозволяє економити близько 10 млн дол. на рік, зменшити час програмування завдань з 30 днів до 4 та зменшити штат програмістів з 34 до 4+8 аналітиків.

**Головний висновок.** Пошук підходу розв'язання рішень складних систем - Big Data, ML, AI, Ontology, I-4.0 і ін. розкриває перехід від технологій обробки інформації до технологій оперування знаннями – до так званих розумних, або SMART технологій.

### Перелік джерел

1. Handbook of Data Quality. Ed. by S. Sadiq. Springer-Verlag Berlin Heidelberg Publ., 2013. 438 p. DOI:10.1007/978-3-642-36257-6.
2. Pipino L. L., Lee Y. W., Wang R. Y. Data quality assessment //Communication of the ACM.-2002.-Vol.45, № 4.-P.211-218.
3. Kahn B. K., Strong D. M., Wang R. Y. Information quality benchmarks: Product and service performance //Communication of the ACM.-2002.-Vol.45, № 4.-P.184-192.
4. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим/ Виктор Майер-Шенбергер, Кеннет Кукьер; пер. с англ. Инны Гайлюк. - М.: Манн, Иванов и Фербер, 2014. - 240 с.
5. Тараненко Р.А. Качество информации: обзор представлений в контексте целостной парадигмы / Руслан Анатольевич Тараненко. // УСиМ. – 2005. – №5. – С. 3–12.
6. Новиков Д.А. Кибернетика: Навигатор. История кибернетики, современное состояние, перспективы развития. - М.: ЛЕНАНД, 2016. - 160 с.
7. Уемов А.И., Штаксер Г.В. К проблеме построения измерительной шкалы для определения степени целостности систем. // Системные исследования. Методологические проблемы. Ежегодник. 2002. – М.: Едиториал УРСС (400 с. – С. 7-33.