

В.І. В'ЮН

ІНТЕЛЕКТУАЛІЗАЦІЯ ІНФОРМАЦІЙНИХ СИСТЕМ – МЕХАНІЗМИ ТА ІНСТРУМЕНТАРІЇ ІНТЕРАКТИВНОГО СИТУАТИВНОГО АНАЛІЗУ

Abstract: In the article there has been considering the problems of expanding the practically-functional structure of the problem-oriented information systems with the technological means of the ternary DW & Olap & IAD (of the Data Warehouse & ON-line analytical processing & the Data Intellectual Analysis). Complecting to the information systems means traditional arsenal adds it the traits of the intelligent system to it that is capable of foreseeing undesirable development of the events in the activity accumulated facts DW interactive analysis processes and of prohosing the forestalling fctions of self-organizing (reingeneering).

Key words: information system, intellectual analysis, data Intellectual analysis, Data warehouse.

Анотація: У статті розглядається питання розширення організаційно-функціональної структури проблемно-орієнтованих інформаційних систем технологічними засобами триади СхД & Olap & ІАД (Сховища даних & On-line analytical processing & Інтелектуального аналізу даних). Доповнення до традиційного арсеналу засобів інформаційної системи надає їй рис "розумної" системи, яка на основі людино – машинного аналізу (Olap) накопичених фактів своєї діяльності (СхД) здатна передбачати небажаний розвиток подій та пропонувати запобіжні заходи самоорганізації (реінжиніринг).

Ключові слова: інформаційні системи, інтелектуалізація, методи Інтелектуального аналізу даних, Сховище даних.

Аннотация: В статье рассматриваются вопросы расширения организационно-функциональной структуры проблемно-ориентированных информационных систем технологическими средствами триады ХрД & Olap & ИАД (Хранилище данных & On-line analytical processing & Интеллектуального анализа данных). Дополнения к традиционному арсеналу средств системы придают ей черты "разумной" системы, которая в процессах интерактивного анализа (Olap) накопленных фактов деятельности (ХрД) способна предвидеть нежелательное развитие событий и предложить упреждающие действия самоорганизации (реинжиниринга).

Ключевые слова: информационные системы, интеллектуализация, методы Интеллектуального анализа данных, Хранилище данных.

“...майбутнє належить тому, хто зможе його передбачити та першим до нього наблизитися”
(К. Парсей)

1. Вступ

У порівнянні з бумом 70 – 90-х рр. сьогодні, на початку ХХІ ст., змінюється парадигма досліджень з проблем інтелектуалізації комп'ютерів та інформаційних систем на їх основі, а саме: не вважаються фундаментальними проблеми пошуку антропоморфних методів та алгоритмів виводу знань, оскільки їх місце займають дослідження методів представлення існуючих знань про об'єкт (предметну область використання комп'ютерів) та результатів інтелектуального аналізу даних його фактичної діяльності та інтерактивного виводу нових знань [1, 2, 3, 4, 5, 6].

Тому довгостроковою метою сучасних досліджень проблеми інтелектуалізації інформаційних систем є створення такого інструментарію їх використання, який був би здатним сприймати „вказівки” користувача стосовно того, які дані, в яких вузлах збирати та зберігати, як забезпечувати та підтримувати їх високу якість.

Саме тому в програмах розробки сучасних предметно (проблемно)-орієнтованих інформаційних систем все більше уваги, поряд зі створенням засобів підтримки та забезпечення виконання формалізованих задач повсякденної діяльності, приділяється концепціям та засобам, орієнтованим на інтерактивний ситуативний аналіз реальних можливостей системи по забезпеченню та підтримці достатнього рівня своєї життєздатності в умовах високої динаміки змін

внутрішніх та зовнішніх вимог до її існування. Розширений таким чином арсенал засобів надає системі рис "розумної" (інтелектуалізованої) системи, яка здатна на основі аналізу накопичених фактів своєї діяльності передбачати небажаний розвиток подій та розробляти запобіжні заходи.

Метою даної роботи є концептуальне формування можливого варіанту операційного розширення організаційно-функціональної структури інформаційної системи необхідними засобами такого типу.

2. Постановка проблеми

Почнемо з постулату інформаційної взаємодії [5], у відповідності з яким **Інформація** є "третім китом підтримки Всесвіту" (два перших – це **Матерія** та **Енергія**). Згідно з постулатами, інформація є творчою, стимулюючою процеси пізнання світу, управляючою силою розвитку та занепаду природних та штучних систем діяльності людини. Інформація, яка запам'ятовується в цих системах, здатна відтворювати накопичений досвід, тенденції та закономірності становлення нових залежностей між показниками-параметрами діяльності, її складовими та породженими ними знань, а отже може бути використана в процесах формування та розробки (або адаптації) засобів підтримки та забезпечення *необхідного рівня власної життєздатності* як однієї з ознак "розумної поведінки" [3]. Проблеми, які при цьому виникають, розглянемо на прикладах практики створення вузько спеціалізованих інформаційних систем типу АСУ.

Традиційні платформи таких систем – це Oltp-технології (on-line transaction processing – системи оперативної обробки транзакцій), які мають більш як 40-річну історію розвитку й становлення. Методологія періоду народження АСУ базувалася на засадах класичної теорії організації управління (Ф.Тейлор, Г.Форд), згідно з якою "виробничий" комплекс – це сукупність механізмів, ефективність функціонування яких задається в термінах продуктивності. Організаційна структура управління таким комплексом будується шляхом розбиття процесу функціонування на мережу інформаційно та технологічно зв'язаних робіт та на апіорному програмуванні задач кожного вузла, включаючи й людей. Тобто в системотвірному плані основу практики формування системи управління "виробничим" об'єктом складали засади довготривалої функціонально-організаційної та технологічної його стабільності.

Така орієнтація системи на повсякденне виконання незмінних у часі "виробничих" процесів обумовила і відповідні методичні та технологічні засади формування результатів обстеження об'єкта автоматизації (виявлення та використання існуючих на момент створення системи знань про *матеріальні, енергетичні та інформаційні* особливості існування) та їх відображення в регламентованих "мовах" інтерактивної взаємодії "користувач – система".

При цьому результати реальної життєдіяльності системи, які, безсумнівно, є найоб'єктивнішими носіями нової інформації та нових знань про об'єкт та його місце "у світі, що змінюється", не охоплювалися цими схемами взаємодії. Тобто поза можливостями традиційної практики використання комп'ютерів у складі АСУ залишалися важливі для аналізу життєстійкості системи управління пласти даних, які зберігають інформацію, що *не виявляється і не висвітлюється* діючою структурно-функціональною організацією системи. Додамо, що такі засоби

“висвітлення” і не могли бути спроектовані та реалізовані під час розробки системи, оскільки мова повинна була йти про майбутні події, пов’язані з невідомим станом середовища існування, змінами умов та вимог до існування системи.

Виходячи з того, що реально існує минуле має безпосереднє відношення до визначення проблем сьогодення, засоби додаткової детермінації системи, включаючи часткову або повну її реорганізацію (реінжиніринг), стають вкрай необхідними в інформаційних процесах забезпечення її життєздатності. Саме можливості такої самоідентифікації системи з точки зору поточних проблем існування і визначаються нами як методологічна платформа розширення “інтелекту” системи за рахунок забезпечення її механізмами та інструментарієм “інтерактивного ситуативного аналізу минулого” в контексті змін оточуючого середовища існування.

Терміном “інтерактивний ситуативний аналіз” пропонується користуватись, підкреслюючи специфічну рису такої взаємодії користувач – система як *необхідність орієнтації на аналіз всіх аспектів минулої діяльності, на “відкриття” та формування нових знань про себе в змінених умовах існування, а також на розробку та опрацювання заходів забезпечення та підтримки адекватної поведінки системи.*

3. Стан проблеми

Починаючи з 90-х років (Inmon 1990 [7], Codd 1993 [8]), методи інтеграції традиційних інформаційних систем управління „засобами інтерактивного ситуативного аналізу” розробляються на платформі побудови *багатовимірних БД (БМБД) “історії життя”* об’єкта, які отримали назву *Сховищ даних (СхД)* як “спеціалізованої БД минулої діяльності об’єкта” і відповідних їм *СУБД (СУБМБД), які орієнтовані саме на інтерактивну аналітичну роботу користувача в операціях “проявлення непрямої інформації”* [5, 9, 10].

Основні засади узгодженості інтеграції традиційно організованих оперативних процесів (ОП) та процесів Ситуативного аналізу (СИТА) розглянемо, порівнюючи їх методичні та системотехнічні особливості:

1. *Основні джерела інформації ОП – діяльність організації.* Для підтримки СИТА потрібно залучати також і відповідні (у часі) дані зовнішніх джерел, які визначають поточні значення зовнішніх факторів впливу.

2. *У багатьох корпоративних структурах одночасно функціонує декілька проблемно-орієнтованих підсистем ОП з власними оперативними БД (ОБД, сукупністю розподілених ОБД), які можуть зберігати однорідну інформацію, але яка з різних причин відрізняється організаційною структурою – формати, часові відмітки надходження та актуалізації, домени.* Інформаційна підтримка СИТА повинна передбачати однаково форматовану, узгоджену змістовно та в часі інформацію, яка адекватно відображає фактичні результати діяльності системи в цілому. Тобто, вилучені дані ОБД перед їх включенням в БД СИТА повинні пройти процедури “очищення” та змістовного узгодження (консолідації) [10].

3. ОП розраховані на підтримку конкретних задач. Сукупність запитів до ОБД, як правило, визначається ще на етапі проектування системи. Механізми інформаційного обслуговування користувачів СИТА орієнтуються, головним чином, на непередбачені (нерегламентовані) запити.

4. ОБД за своїм призначенням повинні бути пристосованими до частих змін даних. Це і враховується існуючими СУБД (нормалізація структури, невпорядкованість записів, В-дерево індексації, транзакційність). При відсутності змін в СхД(БМБД) СИТА в їх структурній організації більш доцільним є саме не нормалізованість даних, використання апріорі впорядкованих записів, методів швидкої індексації, зберігання агрегованих даних.

5. Актуальність даних ОП з часом зменшується і тому такі дані регулярно видаляються з ОБД (в кращому випадку архівуються). СхД СИТА повинні зберігати свої записи на протязі десятиків років.

Таким чином, у складі традиційної системи об'єктивно відсутні „внутрішні носії інформації” про наростання загроз регламентованому існуванню, а, отже, і відчуття необхідності підготовки та планування необхідних запобіжних заходів стосовно можливого реінжинірингу системи. На нашу думку, ініціаторами цих процесів могли б стати доповнюючі на платформі СхД структурно-функціональну організацію інтерактивної взаємодії механізми та інструментарій – *Olap(on-line analytical processing)-систем та Інтелектуального аналізу даних (ІАД)* [9, 10, 11, 14].

Розглянемо більш детально інтелектуалізуючі можливості запропонованої нами тріади СхД&Olap&ІАД засобів інтерактивного ситуативного аналізу.

СхД – платформа аналітичних досліджень “історії життя”. Розробка СхД має свою специфіку в порівнянні з проектуванням та створенням ОБД. Це, перш за все, змістовна узгодженість даних різних джерел. Різних з точки зору визначення та інтерпретації ключових понять спільних інформаційних об'єктів, можливих структурних конфліктів та аномалій.

Ще однією особливістю, яка не проявлялася так гостро у практиці традиційних систем, є забезпечення сумісної навігації в єдиному інформаційному просторі. Відомий інструментарій та механізми метаданих системи в даному випадку повинні виступати в ролі навігаційної карти операцій доступу, яка повинна підтримувати паралельність та персоніфікованість обслуговування запитів. При цьому необхідно також забезпечувати формування однакового розуміння історичних фактів різними групами аналітиків. Адже виявлення протиріч в досягненні такого розуміння є очевидним наслідком того, що даний інформаційний об'єкт характеризується більшою сукупністю параметрів-показників, ніж ті, з якими він використовується діючими функціональними підсистемами.

Системотехнічні засади реалізації СхД визначаються відомими [9] проблемами вибору між віртуальним СхД як Вітрина даних і глобальним СхД.

Віртуальне СхД. Його основу складають: репозиторій метаданих, який описує джерела первинної інформації (ОБД, зовнішні БД та файли та ін.), SQL-форми регламентованих запитів, процедури обробки та представлення результатів. При цьому дані ОБД не вилучаються, а використовуються на місцях зберігання. Користувачі – аналітики фактично працюють з ОБД і повинні мати мережевий доступ до них, а також до зовнішних даних.

До негативних наслідків застосування такої інформаційної підтримки інтерактивних процесів аналітика відноситься те, що доступ до “живих” даних в оперативному режимі та велика інтенсивність трафіка ситуативного аналізу приводять до зменшення продуктивності регламентованої діяльності АСУ і взагалі до загрози втрати нею працездатності при невдалих діях користувачів – аналітиків.

Вітрина даних (Data Mart). Це набір тематично (проблемно) узгоджених вилучень з ОБД, який містить інформацію, що відноситься до конкретних задач аналізу. По суті, Вітрина даних – це вузько спеціалізоване на тематику аналізу мініСхД. Відмітимо, що Вітрина даних не обов’язково повинна бути повністю сформованою. Вона може зберігати посилання на ОБД, за допомогою яких, в разі потреби слід вилучати дані для обслуговування додаткових запитів.

Глобальне СхД. Це спроби інтегрувати всі Вітрини Даних, розглядаючи СхД як єдине джерело даних для всіх можливих тематичних напрямків аналізу.

Механізми Олар

Це інструментарій та процедури підтримки інтерактивних аналітичних процесів, пов’язаних з операціями над багатовимірними структурами даних СхД [9]. Використовуючи гнучкі механізми маніпулювання та відображення багатовимірних даних, користувач Олар-системи в умовах явних чи неявних загроз життєздатності об’єкта розглядає різні інформаційні зрізи СхД, маючи на меті визначити, що могло б бути причиною цих загроз. При цьому, у нього спочатку може навіть не бути ніякої конструктивної ідеї, як взагалі підходити до пошуку відповіді. Зацікавившись якоюсь комбінацією даних, він може розглянути їх більш поглиблено, розклавши впорядковані у часі складові, згрупувавши їх по інших вимірах СхД. Або, навпаки, ще більш узагальнити даний зріз даних, вилучаючи з нього не суттєві подробиці.

Отримані таким чином результати можуть і не привести безпосередньо до готової відповіді, але вони можуть стимулювати інтуїцію аналітика, яка здатна підказати певні асоціації та зв’язки для нових напрямків пошуку, включаючи і пропозиції стосовно змін у процесах збору первинних даних та наповнення ними СхД.

Відмітимо, що механізми та процедури Олар не намагаються моделювати природний інтелект людини, а лише підсилюють його можливості, спираючись на потужність сучасних методів візуалізації та графічного відображення результатів аналізу.

До системотвірних ознак використання Олар-технологій відносяться [8]:

– розподіл даних СхД як БМБД, які визначають історію життя об’єкта, на *показники* та *виміри*;

– логічне представлення значень *показників* у вигляді багатовимірних кубів (полікубів), впорядкованих за рівноправними *вимірами*;

– забезпечувати необмежену кількість *вимірів* та ієрархій зв’язків між ними;

– забезпечувати можливість побудови підмножини значень *показників* (зрізів) за будь-яким дискримінуючим правилом та логічних операцій над цими зрізами;

– можливості он-лайн агрегування значень зрізів, графічного та табличного їх представлення.

Методи та механізми ІАД

Виникнення терміну ІАД, а також його синонімів „Data Mining (Розкопки даних)”, „Knowledge Discovery (Відкриття нових знань)”, пов'язане з розв'язанням проблеми інтуїтивного проявлення непроявленої інформації діяльності, визначення та обґрунтування на основі її аналізу необхідності та змісту реорганізації діючої системи. Дзвоник продзвенів, коли стало ясно, що для знаходження стратегії розумної поведінки необхідно мати змогу вдумливого перегляду як єдиного цілого великих об'ємів показників історичної діяльності з метою виявлення нових залежностей між значеннями параметрів-показників, тенденцій та закономірностей розвитку та занепаду зв'язків між ними. І на цій основі відбувається опрацювання (проекування та тестування) заходів реінжинірингу.

В основі сучасних методів ІАД лежить концепція шаблонів або інформаційних згорток [2, 11, 12, 13], які відображають окремі фрагменти багатоаспектних взаємовідносин між показниками даної вибірки із СхД. Інтерактивний аналіз впорядкованої в часі послідовності таких інформаційних згорток може підштовхнути аналітика до відкриття нового знання про об'єкт. До речі, зауважимо, що класична мат статистика, як основний інструмент аналізу експериментальних даних, спираючись на парадигму визначення та оперування з середніми значеннями вибірки, виявилася неспроможною в розв'язанні цих задач. Зміну парадигми аналізу проілюструємо прикладом із [14]:

Традиційна прикладна статистика – *„Які середні показники травматизму для курців і некурців?”*.

Інтелектуальний аналіз даних – *„Чи зустрічаються точні шаблони описів людей, що характеризують їх здатність до підвищеного травматизму?”*.

Легко побачити, що інтелектуалізуючою ознакою нової парадигми є нетривіальність вибірки. В даному випадку ми маємо на увазі саме несподіваність, неочікуваність даного групування показників, замаскований характер нових фактів, які стимулюють процеси відкриття нових знань.

4. Висновки

Запропонований в даній роботі підхід до інтеграції системотвірних засобів традиційних інформаційних систем управління та інтелектуалізуючих їх засобів СхД&ОІар&ІАД дає підстави стверджувати, що таке поєднання забезпечує систему повнофункціональними механізмами, інструментами підтримки процесів ситуативного визначення можливих загроз її існуванню та шляхів їх попередження.

Підсумовуючи вищесказане, відмітимо, що цей висновок обумовлений таким:

– СхД вилучають і окремо зберігають для подальшого аналізу (ОІар&ІАД) інформацію змістовно зв'язаних різнорідних джерел первинних даних, які в сукупності характеризують історію діяльності об'єкта;

– механізми ОІар дозволяють виконувати інтерактивні операції над багатовимірними зрізами СхД в пошуках гіпотез-припущень щодо виявлення та опрацювання нових залежностей між параметрами-показниками діяльності, тенденцій (як негативних, так і позитивних) їх розвитку;

– методи IAD допомагають виявляти та формувати нові знання на базі результатів інтерактивного ситуативного аналізу, тестувати їх на різних часових зрізах СхД.

СПИСОК ЛІТЕРАТУРИ

1. Капитонова Ю.В., Скурихин В.И. О некоторых тенденциях развития и проблемах Искусственного Интеллекта. // Кибернетика и Системный Анализ. – 1999. – № 1.
2. Морозов А.А. Базы знаний в системах ситуационного управления коллективного пользования // УсиМ. – 1995. – № 5. – С. 1 – 5.
3. Морозов А.О., В'юн В.І., Кузьменко Г.Є. Інтелектуалізація АСУ: проблеми, напрямки досліджень // Математичні машини і системи. – 2001. – № 1–2. – С. 21 – 25.
4. Вейценбаум Дж. Возможности ЭВМ и человеческий разум. – М.: Радио и связь, 1982. – С. 367.
5. Теслер Г.С. Информация – феномен природы: роль информации в естественной и искусственной природе // Математичні машини і системи. – 2003. – № 1. – С.152 – 165.
6. В'юн В.І., Довгополий А.С., Кузьменко Г.Є. Багатоагентні риси інтелектуалізації систем управління // Математичні машини і системи. – 2002. – № 4. – С. 52 – 56.
7. В'юн В.І., Кузьменко Г.Є. Багатоагентні риси інтелектуалізації систем управління // Математичні машини і системи. – 2002. – № 4. – С. 52 – 56.
8. Inmon W.H. Building the Data Warehouse (second edition). – NY, NY:John Wiley, 1993.Codd E.F., Codd S.B., Salley C.D. Providing Olap to User-Analysts: An IT Mandate – E.F// Codd&Associates, 1993.
9. Codd E.F., Codd S.B., Salley C.D. Providing Olap to User-Analysts: An IT Mandate – E.F// Codd&Associates. – 1993.
10. Сахаров А.А. Концепция построения и реализации информационных систем, ориентированных на анализ данных // СУБД. – 1996. – № 3. – С. 44 – 53.
11. Parsaye K. Surveying Decission Support: New Realms of Analysis // Database Programming and Design. – 1995. – N 4. – С.15 – 20.
12. Raden N. Star Schema / Santa Barbara, CA: Archer Decissions Sciences, Inc., 1995–1996. (<http://members.aol.com/raden/str/oi.htm>).
13. Андон Ф.И., Балабанов А.С. Выявление знаний и изыскания в БД: подходы, модели, методы и системы (обзор) // Проблемы программирования. – 2000. – № 1–2. – С. 513 – 525.
14. Балабанов А.С. Проблема вывода знаний о структуре зависимостей между переменными из данных Больших объемов в условиях помех // Проблемы программирования. – 2000. – № 1 – 2. – С. 527 – 535.
15. Дюк В., Самойленко А. Data Mining – интеллектуальный анализ данных. – Санкт-Петербург: Питер, 2001. – С. 365.