

КОМПЬЮТЕРНАЯ ИНЖЕНЕРИЯ И ТЕХНИЧЕСКАЯ ДИАГНОСТИКА



УДК004.891

КОМБІНОВАНИЙ МЕТОД ЗБОРУ, ДИНАМІЧНОГО АНАЛІЗУ ТА ВІЗУАЛІЗАЦІЇ ЧИСЛОВОЇ ІНФОРМАЦІЇ

ФАРТУШОК І.М.

Розглядається типова проблема збору, збереження та аналізу числової інформації, будується алгоритм її попереднього динамічного аналізу і показується методика візуалізації такої інформації за допомогою табличного процесора *Microsoft Excel* з програмного пакета *Microsoft Office®*

1. Актуальність проблеми

У науці, техніці чи будь-якій іншій сфері діяльності часто доводиться вимірювати деяку неперервну змінну величину, наприклад, тиск у газо- чи нафтопроводі, температуру робочого середовища, силу струму в електричному колі тощо. Для цього, як правило, використовують самописець, який неперервно фіксує значення такої величини у вигляді нескінченної кривої на паперовій стрічці чи паперовому диску.

Проте стрімке поширення персональних комп'ютерів, які дозволяють повністю автоматизувати збір, зберігання та аналіз інформації, вимагає розробки нових методик її обробки та візуалізації.

Для збору та оцифровування даних датчики і прилади, які працюють під управлінням комп'ютера, у певні, визначені користувачем моменти часу (щосекунди чи долі секунди, кожної години, дня, місяця або року) вимірюють та передають до комп'ютера значення параметра. Так неперервна величина перетворюється у дискретний числовий ряд. Далі до неї можна застосовувати статистичні способи обробки даних, будувати математичні моделі, прогнозувати тенденції тощо.

Недоліком такого методу реєстрації та запису інформації є його внутрішня суперечливість між об'ємом збережених даних та їхньою інформативністю. Збільшення частоти вимірювань супроводжується зростанням об'єму даних, а зменшення частоти знижує їх

інформативність. При цьому неминуче втрачається значна частина інформації, яка міститься між двома сусідніми вимірюваннями. Наприклад, потужний короткочасний імпульс, який іноді свідчить про нештатний режим функціонування обладнання чи несанкціоноване втручання в його роботу, можна просто не помітити, якщо він відбудеться між двома вимірюваннями. І чим більший інтервал дискретизації, тим більша ймовірність пропуску різких коливань параметра та більші втрати інформації.

Продемонструємо це на прикладі набору даних про значення тиску в газопроводі протягом доби.

На рис. 1 зображено добовий графік зміни тиску в газопроводі при щохвилинній фіксації показів датчика тиску. При цьому до бази даних внесено 1441 трійку чисел – номер запису, часову мітку і відповідне їй значення тиску (табл. 1). CSV-файл з цими даними займає 28,5 кБ.

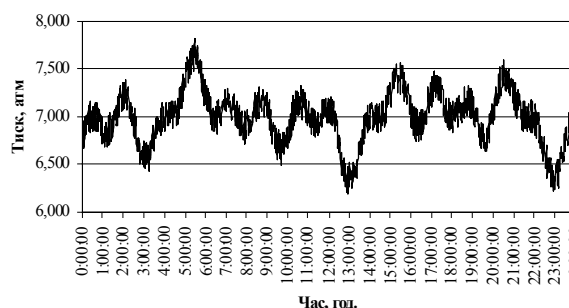


Рис. 1. Добовий графік зміни тиску в газопроводі при щохвилинній фіксації показів датчика тиску

Як бачимо з рис. 1, такий спосіб запису даних є надлишковим і суттєво збільшує об'єм бази даних.

Таблиця 1

№ запису	Час	Тиск	№ запису	Час	Тиск
1	0:00:00	6,944
2	0:01:00	6,685	1436	23:55:00	7,000
3	0:02:00	6,806	1437	23:56:00	6,867
4	0:03:00	6,680	1438	23:57:00	7,047
5	0:04:00	6,950	1439	23:58:00	7,128
6	0:05:00	6,663	1440	23:59:00	7,162
...	1441	0:00:00	7,086

Таблиця 2

№ запису	Час	Тиск	№ запису	Час	Тиск
1	0:00:00	6,944	14	13:00:00	6,271
2	1:00:00	6,832	15	14:00:00	6,886
3	2:00:00	7,080	16	15:00:00	7,289
4	3:00:00	6,550	17	16:00:00	7,170
5	4:00:00	6,947	18	17:00:00	7,411
6	5:00:00	7,314	19	18:00:00	7,113
7	6:00:00	7,374	20	19:00:00	7,038
8	7:00:00	7,068	21	20:00:00	7,120
9	8:00:00	6,758	22	21:00:00	7,243
10	9:00:00	7,257	23	22:00:00	6,953
11	10:00:00	6,712	24	23:00:00	6,364
12	11:00:00	6,916	25	0:00:00	7,086
13	12:00:00	7,032			

Для зменшення цього об'єму збільшимо інтервал вимірювань до однієї години. У результаті отримаємо набір з двадцяти п'яти трійок чисел (табл. 2). CSV-файл з цими даними займає 0,5 кБ. Графік, який побудовано за цими значеннями, зображено на рис. 2.

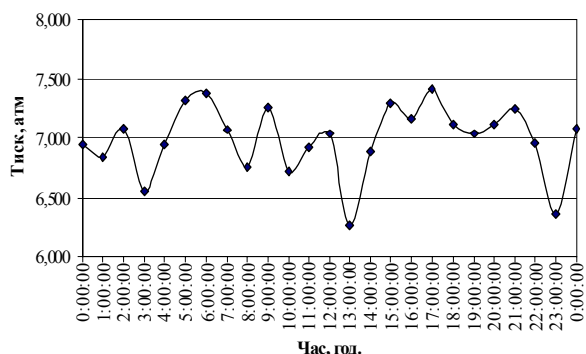


Рис. 2. Добовий графік зміни тиску в газопроводі при щогодинній фіксації показів датчика тиску

Як бачимо з рис. 2, при такому великому інтервалі дискретизації втрачено багато цінної інформації, а загальну картину поведінки параметра суттєво спотворено. Зокрема, на графіку відсутні локальні екстремуми значень тиску протягом першої, сьомої, дванадцятої, п'ятнадцятої та двадцятої годин. Таким чином, хоча кількість збережених даних зменшилася, проте їхня інформативність неприємно погіршилася.

Для оптимізації співвідношення між об'ємом збережених даних та їхньою інформативністю використаємо запропонований в [2] і дещо нами модифікований спосіб реєстрації, збереження та аналізу інформації, який називається «Біржова діаграма». Цей метод дозволяє зберегти максимум інформації про динаміку досліджуваного процесу і відкриває нові, значно ширші можливості у багатьох сферах діяльності, зокрема – у технічному аналізі.

Як свідчить назва, біржова діаграма створювалася як інструмент біржової торгівлі. Коли людина грає на біржі та ризикує власними грошми, вона стає надзвичайно вимогливою до тих засобів, які допомагають їй ці гроші зберегти. Тому біржова діаграма має відповідати декільком важливим критеріям. Вона не повинна упускати короточасні та випадкові коливання ціни, тобто має відображати усі без винятку зміни ціни

без жодного усереднення чи згладжування. Також потрібне узагальнення довгострокової динаміки торгів. І, нарешті, необхідна максимальна гнучкість у різноманітних часових масштабах, яка б дозволяла візуалізувати динаміку процесу на проміжках часу, які вимірюються як роками, так і хвилинами, зберігаючи при цьому точність та об'єктивність інформації. Тому інструмент, який має усі ці якості й принципово інакше відображає цифрову інформацію, безсумнівно буде надзвичайно корисним усім, хто оперує числовими даними.

2. Практична реалізація методу

Нехай маємо деякий змінний неперервний процес (рис. 3), який необхідно оцифрувати.

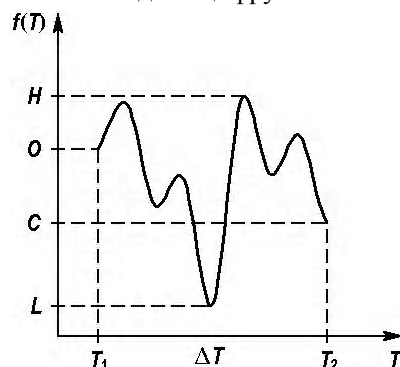


Рис. 3. Графік неперервного процесу

Позначимо часовий інтервал між T_1 і T_2 змінною ΔT (період вимірювань). При звичайному способі запису інформації значення вимірюваної величини фіксують лише у моменти часу T_1 і T_2 . У новому форматі запису фіксують чотири значення змінної величини [2]:

- у моменти часу T_1 і T_2 : $f(T_1) = O$ та $f(T_2) = C$, де O (Open) – початкове значення, на біржовому сленгу – «ціна відкриття», а C (Close) – кінцеве значення, на біржовому сленгу – «ціна закриття»;
- мінімальне L (Low) і максимальне H (High) значення, які вимірювана величина досягла на проміжку $[T_1; T_2]$.

У даному випадку датчики неперервно вимірюють вхідний параметр (правду кажучи, це вони роблять дискретно – в межах своєї роздільної здатності), а комп'ютер з усього набору одержаних значень виби-

Таблиця 3

№ запису	Час	Тиск			
		Початковий	Найвищий	Найнижчий	Кінцевий
1	1:00:00	6,944	7,157	6,663	6,874
2	2:00:00	6,832	7,356	6,675	7,285
3	3:00:00	7,080	7,378	6,473	6,739
4	4:00:00	6,550	7,145	6,432	6,959
5	5:00:00	6,947	7,443	6,860	7,361
...
20	20:00:00	7,038	7,205	6,627	6,938
21	21:00:00	7,120	7,589	6,931	7,388
22	22:00:00	7,243	7,360	6,833	6,835
23	23:00:00	6,953	7,127	6,218	6,318
24	0:00:00	6,364	7,162	6,237	7,162

рає та фіксує лише чотири: О і С – на краях часового інтервалу ΔT ; L і Н – локальні екстремуми значень параметра на інтервалі ΔT . Блок-схему процедури збору та первинного аналізу числових даних зображено на рис. 4. Протягом однієї доби роботи цієї програми до бази даних вноситься 24 записи по 6 полів у кожному: номер запису, часова мітка та 4 значення параметра (табл. 3). Розмір CSV-файлу з цими даними становить 1 кБ.

На перший погляд, запропоноване удосконалення дещо сумнівне, оскільки замість одного числа потрібно зберігати чотири. Проте насправді ця четвірка чисел описує не саму точку T_2 , а, як справедливо зазначено в [2], процес протягом періоду $[T_1; T_2]$, тривалість якого користувач (оператор) може задава-

ти довільно. При цьому будь-який інший метод опису процесу на проміжку ΔT оперуватиме наборами від декількох значень параметра до їх нескінченної кількості.

Які ж переваги даного способу запису?

По-перше, перевага кількісна. Як зазначено вище – мінімум цифр описує максимум цінної інформації про процес. При бажанні, з усього масиву даних можна вибрати однотипні значення (наприклад Open) і одержати звичайний дискретний ряд (табл. 2).

По-друге, отримані цифри якісно інакше описують досліджуваний процес і значно повніше його характеризують. Аналіз значень О, L, Н і С дає багатий матеріал для висновків і прогнозів. Зокрема, співвідно-

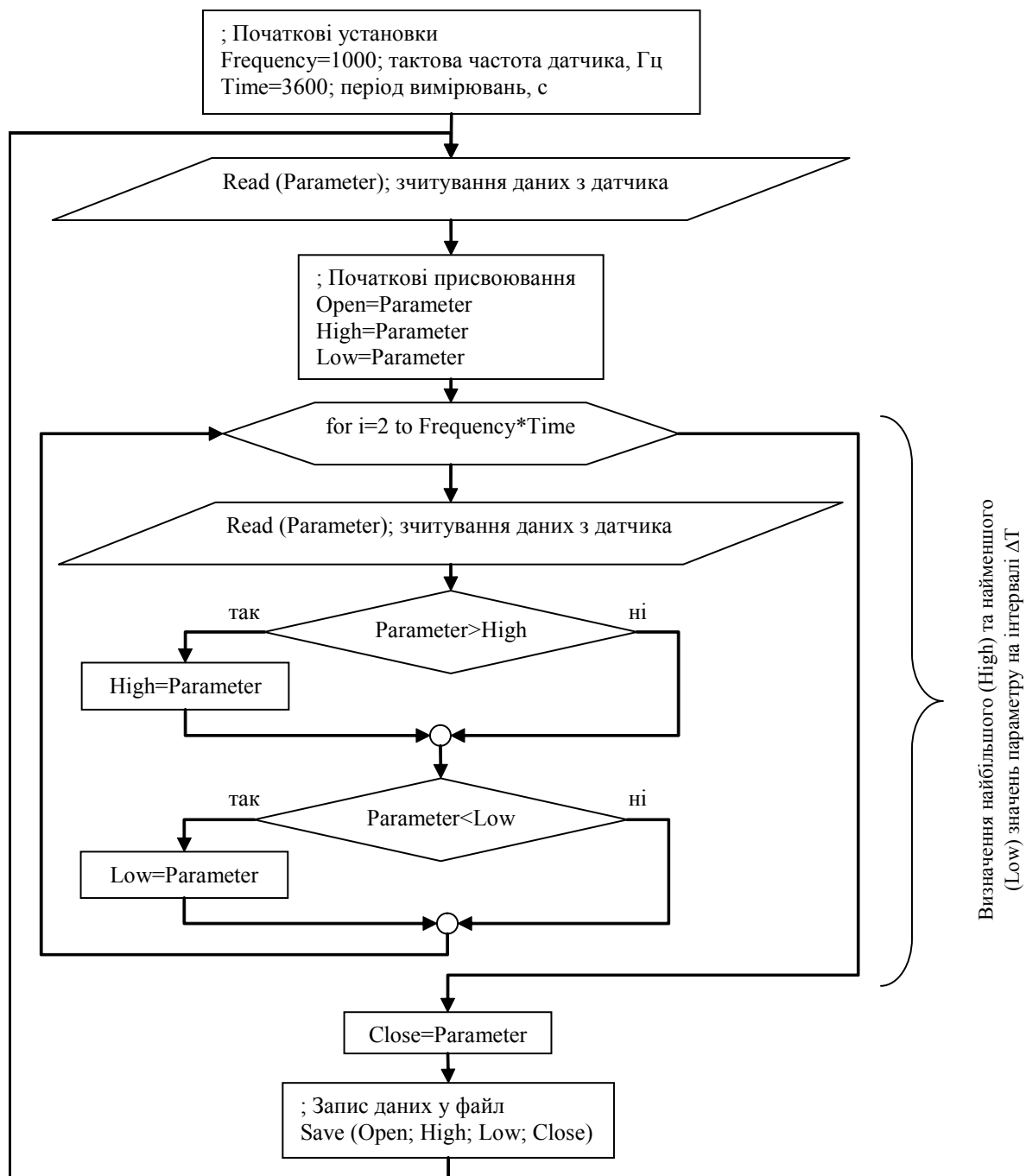


Рис. 4. Блок-схема алгоритму збору та первинного аналізу числових даних

шення L і H характеризує мінливість процесу (щось аналогічне, скажімо, середньому стандартному відхиленню), а O та C дозволяють визначити тенденцію процесу на даній ділянці кривої (зростання-спадання) тощо.

По-третє, у такому вигляді до числових даних, крім класичних методів математичного та статистичного аналізу, можна застосувати широкий спектр спеціальних аналітичних інструментів – так звані технічні індикатори, які суттєво розширюють можливості технічного аналізу.

3. Візуалізація та аналіз даних у *Microsoft Excel*

Перейдемо тепер безпосередньо до діаграми, тобто до картинки, яка відображає всю красу та багатство чисел, і розглянемо методику побудови біржової діаграми у *Microsoft Excel* [1, 2].

Існує два основні види біржової діаграми. Перший, винайдений у Європі, складається з вертикальних ліній – барів (рис. 5). Нижня точка бари відповідає значенню L, верхня – значенню H, горизонтальна риска ліворуч – O, горизонтальна риска праворуч – C. У цілому біржова діаграма є послідовною множиною барів і дозволяє одним поглядом оцінити мінливість процесу та побачити «коридор» значень параметра.

Ще один вид біржової діаграми створено в Японії. Він з'явився незалежно від європейських барів ще в середині минулого століття. За його допомогою японські торговці рисом відслідковували ситуацію на ринку. Разюча кореляція європейського та азійського підходів переконливо свідчить про об'єктивність розвитку наукової думки. Умовно цей вид діаграми називається «японськими свічками» (правильна назва – «японські підсвічники», але частіше вживають саме слово «свічки»). Можливо тому, що воно коротше й легше вимовляється). Японські свічки відображають ту ж інформацію, що й бари, але своїм, особливим східним чином. Свічка є прямокутником, верхній та нижній краї якого відповідають значенням O і C. Догори й донизу від свічки проводять вертикальні лінії до значень L і H. Саму свічку зображають світлою, якщо значення C більше за O – параметр зріс, і темною (кольоровою), якщо C менше за O – параметр зменшився.

Японські свічки наочніше відображають характер процесу: ділянка зі світлих свічок – тенденція зростання, ділянка з темних свічок – тенденція спадання. Крім того, завдяки притаманній східному менталітету увазі до найменших дрібниць, „рисові” аналітики визначили цілий ряд фігур – особливих комбінацій світлих і темних свічок різної форми та дали їм поетичні назви: «молот», «зірка, яка заходить», «сонце серед хмар» тощо. Кожна така фігура, появившись на діаграмі, сигналізує про певну ситуацію: сповільнення або прискорення тенденції, зміну зростання зниженням чи навпаки. Існують спеціальні методики аналізу складних фігур японських свічок на діаграмах, які дозволяють глибше зрозуміти досліджуваний процес, проте їх аналіз виходить за рамки даної статті.

Для побудови біржової діаграми у табличному редакторі *Microsoft Excel* необхідно:

Ш виділити курсором діапазон числових даних;

Ш на панелі інструментів натиснути кнопку «Майстер діаграм»;

Ш у вікні, яке відкрилося, вибрати пункт «Біржова діаграма».

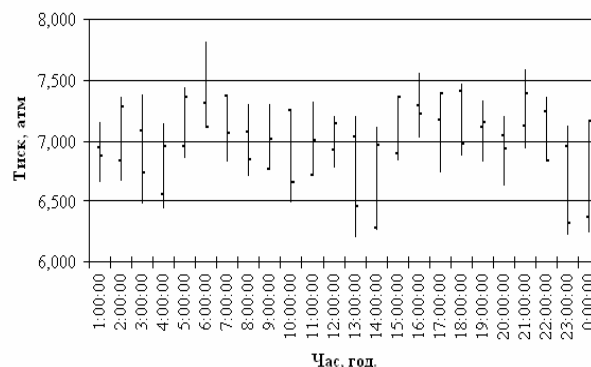


Рис. 5. Чотиризначна барова діаграма

Можливі чотири варіанти біржових діаграм:

1. Барова діаграма за трьома значеннями – H, L і C (рис. 6). У цьому варіанті не використовується значення Open. Очевидно передбачається, що Open дорівнює значенню Close попередньої бари. Така рівність буває часто, проте не завжди.

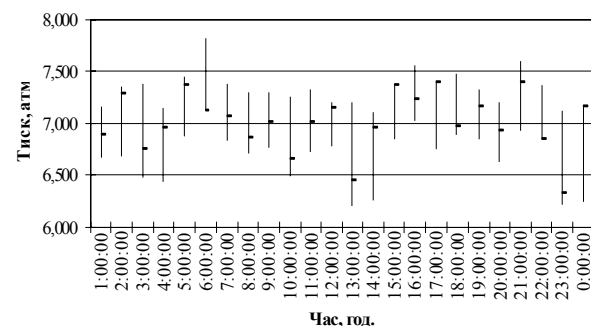


Рис. 6. Тризначна барова діаграма

2. Діаграма «Японські свічки» за чотирма значеннями – O, H, L і C (рис. 7). Це класична біржова діаграма. Як бачимо, вона достатньо інформативна, наглядно відображає тенденції поведінки параметру (білі та чорні бари) і дозволяє визначити його локальні екстремуми на заданих часових проміжках.

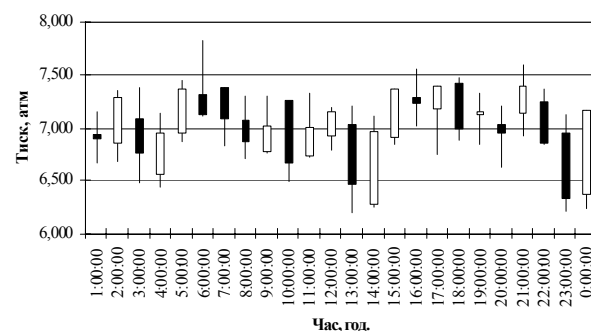


Рис. 7. Діаграма «Японські свічки»

3. Барова діаграма за чотирма значеннями – Н, L, C і значенням «Об'єм» («Volume»). У практиці біржових торгів таку діаграму іноді доповнюють гістограмою (стовпчиковою діаграмою), яка відображає об'єм придбаних чи проданих товарів (у нашому випадку – об'єм перекачаного газу). Це не частина біржової діаграми, а окрема діаграма, поєднана з біржовою.

4. Діаграма «Японські свічки» за п'ятьма значеннями – О, Н, L, C і значенням «Об'єм» («Volume»).

На біржовій діаграмі можна зобразити поведінку не лише неперервної величини, а й дискретної. Наприклад, маючи дискретний числовий ряд (табл. 4), його можна розбити на періоди і, представивши кожен період окремою барою, побудувати біржову діаграму.

1) Для цього необхідно [2]:

1. У редакторі Microsoft Excel сформувати таблицю даних (табл. 1). Зокрема:

1.1. У комірку A1 записати заголовок стовпчика А – «Номер запису» або «Час». У решту комірок стовпчика А, починаючи з комірки A2, занести відповідні мітки даних (номер вимірювання чи його дату, час тощо).

Таблиця 4

Комірка	Формула	Коментар
C61	=B2	Значення Open дорівнює першому значенню у стовпчику В
D61	=МАКС(B2:B61)	Значення High дорівнює максимальному серед шестидесяти значень у стовпчику В
E61	=МИН(B2:B61)	Значення Low дорівнює мінімальному серед шестидесяти значень у стовпчику В
F61	=B61	Значення Close дорівнює шістдесятому значенню у стовпчику В

1.2. У комірку B1 записати заголовок стовпчика В – «Дані», «Тиск» тощо. У решту комірок стовпчика В, починаючи з комірки B2, занести дані.

2) Згрупувати дані (скажімо, для погодинного аналізу – по шістдесят значень) і розрахувати параметри для побудови біржової діаграми. Результати записати у стовпчики C, D, E та F. Для цього:

2.1. У перші комірки цих стовпчиків записати їхні заголовки: C1 – «Open»; D1 – «High»; E1 – «Low»; F1 – «Close».

2.2. В інші комірки записати формули.

Примітка: Це саме той випадок, коли значення Open поточної бари чи свічки не дорівнює значенню Close попередньої.

2.3. Виділити діапазон C2:F61 і скопіювати його донизу листа Excel до закінчення даних у стовпчику В (обов'язково копіювати разом із порожніми комірками діапазону C2:F60).

Тепер стовпчики C, D, E та F міститимуть усі необхідні дані для побудови біржової діаграми (табл. 3).

4. Висновки

Розглянутий метод збору, попереднього динамічного аналізу та збереження великого масиву числової інформації є гнучким у користуванні, дозволяє оперативно реагувати на зміни в поведінці параметра і коректно регулювати частоту його зчитування з метою зменшення об'єму збережених даних без втрати їх інформативності. Візуалізація таких даних за допомогою біржової діаграми робить їх максимально наглядними, а аналіз – ефективним.

Література: 1. Долженков В., Колесников Ю. Microsoft® Excel 2000. Полное руководство. СПб.: БХВ – Санкт-Петербург, 1999. 1088 с. 2. Шкуренко В. Excel: биржевая диаграмма // CD «Компьютеры + программы», 2005, № 9.

Надійшла до редколегії 05.06.2011

Рецензент: д-р техн. наук, проф. Русин Б.П.

Фартушок Ігор Михайлович, канд. техн. наук, доцент каф. машинознавства та матеріалознавства Дрогобицького державного педагогічного університету імені Івана Франка. Наукові інтереси: технічна діагностика. Адреса: Україна, 82100, Львівська обл., Дрогобич, вул. Стрийська, 5, тел. 067-755-00-22.