ДОСЛІДЖЕННЯ ПИТАНЬ УГРУПОВАННЯ ОБ'ЄКТІВ В АРХІТЕКТУРІ ІНТЕРФЕЙСІВ КОРИСТУВАЧА МЕТОДОМ КЛАСТЕРНОГО АНАЛІЗУ

***Шакурова Т.В.**, асистент, кафедра МСТ ХНУРЕ*
***Каспарова М.Д.,** студентка, кафедра МСТ ХНУРЕ*

Кластерний аналіз - це багатовимірна статистична процедура, яка виконує збір даних, що містять інформацію про вибірку об'єктів і потім упорядковує об'єкти в порівняно однорідні групи – кластери. Основна мета кластерного аналізу – знаходження груп схожих об'єктів у вибірці.

Різні методи кластеризації можуть породжувати різні кластерні рішення для одних і тих же даних. Це звичне явище у більшості прикладних дослідженнях, у тому слід по-перше обирати найбільш осмислене рішення, по-друге – завжди вказувати, який саме метод кластеризації було використано.

Використовуючи кластерний аналіз, дослідник має на меті виявлення структури даних. В той же час дія кластерного аналізу полягає у привнесені структури у аналізовані дані. Тобто, кластеризація може призвести до появи артефактів (виявлення структури в даних, які її не мають).

Осмислене рішення при кластерному аналізі можна обрати лише тоді, коли є базис для його осмислення – теорія. Без теоретичної моделі, без гіпотези стосовно структури даних з'являється небезпека наївного емпіризму, коли результати кластеризації приймаються на істину у кінцевій інстанції.

Серед методів угрупування об'єктів кластерного аналізу найчастіше зустрічається пошук узагальнення на підставі евклідовій відстані (Euclidian distances). Це найуживаніша міра близькості об'єктів, яка являє собою геометричну відстань між об'єктами у багатомірному просторі. Формула для обчислення евклідової відстані для двох точок на площині має такий вигляд:

$$L = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \ .$$

Евклідова відстань обчислюється по стандартизованим (нормованим) даним, всі змінні повинні бути виміряні в одному масштабі (якщо це сантиметри, то всі змінні повинні бути виміряні в сантиметрах тощо).

Процес упорядкування груп за допомогою розрахунків відстані між об'єктами містить як безпосереднє обчислення відстані від кожного об'єкта до кожного іншого об'єкта (матрицю відстаней), так і зміну розташування об'єктів для більш наочної демонстрації результатів.

Таким чином виникла необхідність ускладнити просте завдання Excel обчислення коренів з суми квадратів допоміжною можливістю сортування об'єктів без суттєвих змін в посиланнях та формулах.

Для цього обчислення біли проваджені за допомогою функцій ДВССЫЛ(АДРЕС(стовпчик; строка). Додаткова таблиця (на рис. 1 праворуч з кольоровим фоном) організовано не безпосереднім посиланням, що у випадку перетягування (сортування даних) веде до перемішування посилань, а

опосередкованим посиланням через порядкові номери рядків та стовпчиків. У прикладі це дані Об'єм та Ціна.

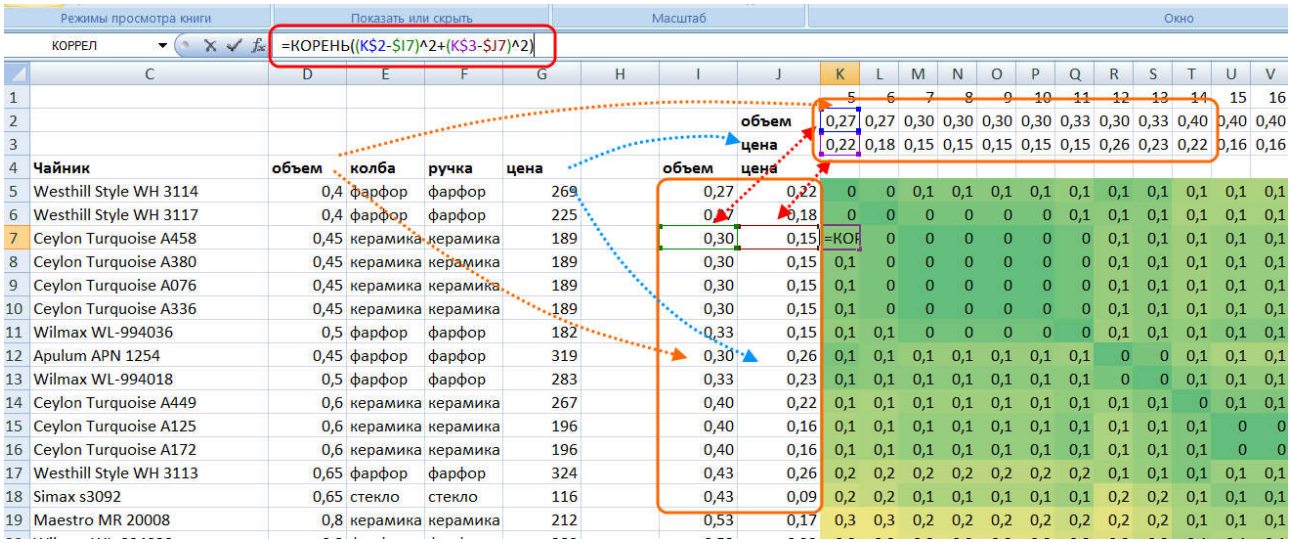

Рисунок 1 – Організація копіювання даних опосередкованим посиланням

Відстань між двома об'єктами розраховується як корінь з квадратів різниці параметрів. В прикладі параметри було нормовано шляхом ділення на максимальне можливе значення.

За рахунок такої організації таблиці стало можливе сортування даних у будь-який спосіб (вручну або автоматично) за будь-яким критерієм, що надає можливість підбирати сортування так, щоб наочно представити результати угрупувань.

Так у наведеному прикладі (рис. 2) можна виявити два або три чітких угрупування, які показані синьою рамочкою.

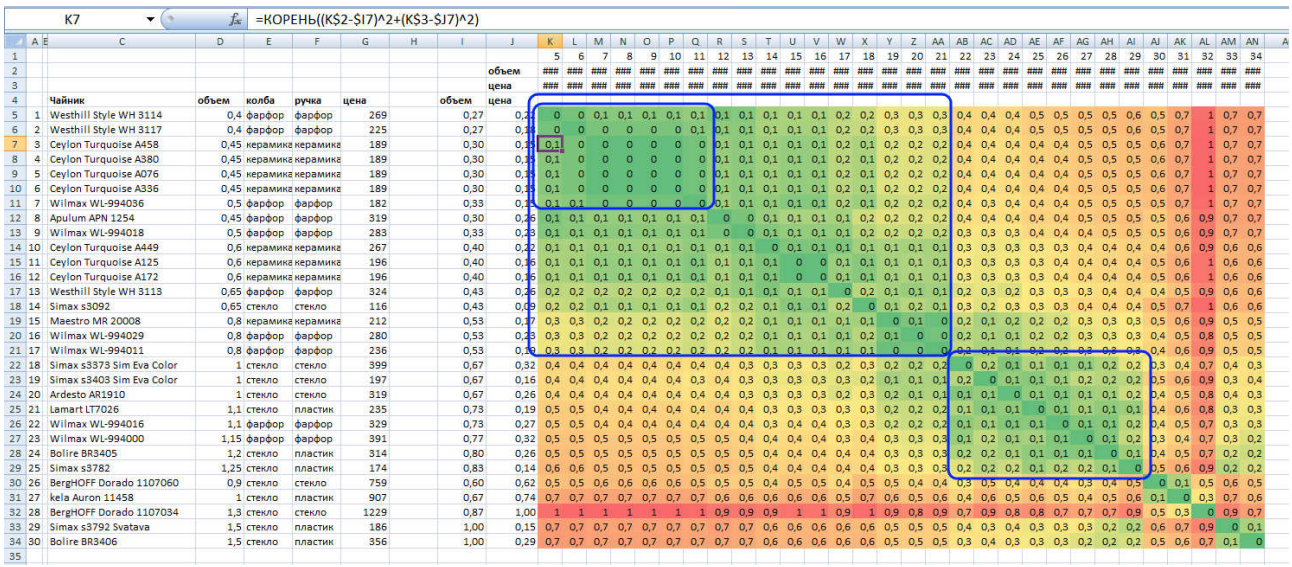

Рисунок 2 – Таблиця відстаней після сортування