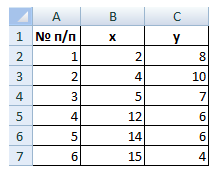
**Лабораторна практична робота**

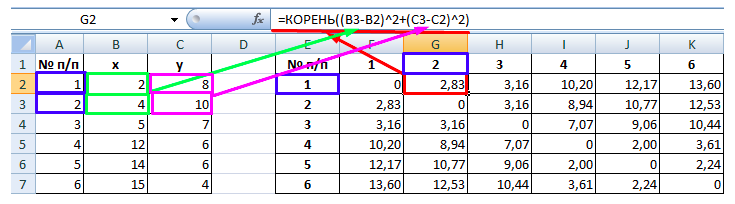
**Кластерний аналіз текстів (класифікація текстів)**

**Як зробити кластерний аналіз в EXCEL**

Для прикладу візьмемо шість об'єктів спостереження. Кожен має два параметра які його характеризують.

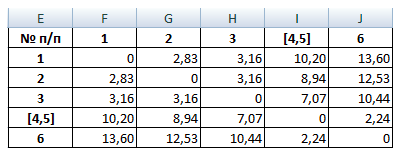


Як відстань між об'єктами візьмемо Евклідову відстань. Формула розрахунку:

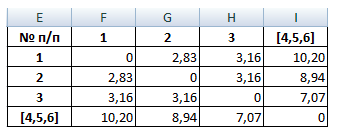


Розраховані дані розміщуємо в матриці відстаней.

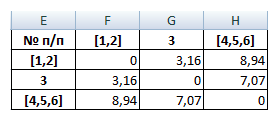
Отже, їх можна об'єднати в одну групу - при формуванні нової матриці залишаємо найменше значення.



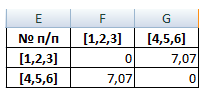
З нової матриці видно, що можна об'єднати в один кластер об'єкти [4, 5] і 6 (як найбільш близькі один до одного за значеннями). Ми залишаємо найменше значення і формуємо нову матрицю:



Об'єкти 1 і 2 можна об'єднати в один кластер (як найбільш близькі з наявних). Вибираємо найменше значення і формуємо нову матрицю відстаней. В результаті отримуємо три кластери:



Найближчі об'єкти - 1, 2 і 3. Об'єднаємо їх.



Ми провели кластерний аналіз за методом «найближчого сусіда». В результаті отримано два кластери, відстань між якими - 7,07.

Величезне значення має кластерний аналіз в економічному аналізі. Інструмент дозволяє виокремлювати з величезної сукупності періоди, де значення відповідних параметрів максимально близькі і де динаміка найбільш схожа. Для дослідження, наприклад, товарної і общехозяйственной кон'юнктури цей метод відмінно підходить.

**Завдання лабораторної роботи**

Взяти **N** текстів приблизно однакової довжини. Порахувати кількість слів які повторюються (тільки слів). У стовпчику «А» вказати вказати слова, стовпчику «B - С» номер тексту (або його назву). Заповнити таблицю отриманими даними. Класифікувати тексти за поданим в теорії алгоритмом використовуючи формулу Евклідової відстані. Отримані результати записати, зберегти, проаналізувати.

