Zadanie 3c – klastrovanie

Máme 2D priestor, ktorý má rozmery X a Y, v intervaloch od -5000 do +5000. Tento 2D priestor vyplňte 20 bodmi, pričom každý bod má náhodne zvolenú polohu pomocou súradníc X a Y. Každý bod má unikátne súradnice (t.j. nemalo by byť viacej bodov na presne tom istom mieste).

Po vygenerovaní 20 náhodných bodov vygenerujte ďalších 20000 bodov, avšak tieto body nebudú generované úplne náhodne, ale nasledovným spôsobom:

1. Náhodne vyberte jeden zo **všetkých doteraz vytvorených** bodov v 2D priestore. (nie len z prvých 20)
   Ak je bod príliš blízko okraju, tak zredukujete príslušný interval, uvedený v nasledujúcich dvoch krokoch.
2. Vygenerujte náhodné číslo X_offset v intervale od -100 do +100
3. Vygenerujte náhodné číslo Y_offset v intervale od -100 do +100
4. Pridajte nový bod do 2D priestoru, ktorý bude mať súradnice ako náhodne vybraný bod v kroku 1, pričom tieto súradnice budú posunuté o X_offset a Y_offset

Vašou úlohou je naprogramovať zhlukovač pre 2D priestor, ktorý zanalyzuje 2D priestor so všetkými jeho bodmi a rozdelí tento priestor na k zhlukov (klastrov). Implementujte rôzne verzie zhlukovača, konkrétne týmito algoritmami:

- aglomeratívne zhlukovanie, kde stred je centroid
- aglomeratívne zhlukovanie, kde stred je medoid (stačí 5000 bodov)

Vyhodnocujte úspešnosť/chybovosť vášho zhlukovača. Za úspešný zhlukovač považujeme taký, v ktorom **žiaden klaster nemá priemernú vzdialenosť bodov od stredu viac ako 500**.

Vizualizácia: pre každý z týchto experimentov vykreslite výslednú 2D plochu tak, že označkujete (napr. vyfarbíte, očíslujete, zakrúžkujete) výsledné klastre.

Dokumentácia musí obsahovať opis konkrétne použitých algoritmov a reprezentácie údajov. Uveďte aj vizualizácie viacerých pokusov. V závere zhodnoťte dosiahnuté výsledky ich porovnaním.

Poznámka: Je vhodné použiť rôzne optimalizácie pre dostatočne efektívnu prácu Vášho zhlukovača:

- Použitie 2-rozmernej matice vzdialenosti dvojíc bodov. Naplnenie takejto matice má kvadratickú zložitosť. Potom sa hľadá najbližšia dvojica (najmenšie číslo v matici), to má opäť kvadratickú zložitosť, ale nenásobia sa tie časy, len sčítavajú. Po výbere najbližšej dvojice túto dvojicu treba zlúčiť a tým sa zníži veľkosť matice o 1 (lebo sa zníži počet zhlukov o 1) Pri tom sa aktualizujú len vzdialenosti pre tento nový zhluk (len jeden stĺpec/riadok, zvyšok matice ostáva nezmenený).
- PyPy je implementácia programovacieho jazyka Python. PyPy často beží rýchlejšie ako štandardná implementácia CPython, pretože PyPy používa just-in-time kompilátor. Pypy nepodporuje niektoré grafické knižnice.

Príklad vizualizácie: