

## Abstrakt

V tejto práci sa zaoberáme problematikou manipulácie s obrovskými dátami v kontexte výpočtovej pangénomiky. Cieľom pangénomiky je nahradiť referenčný genóm v analýzach DNA sekvencií genómami väčšieho počtu jedincov, prípadne príbuzných druhov, za účelom zvýšenia robustnosti týchto analýz. Práca s rádovo väčším počtom genómov je však limitovaná výpočtovou náročnosťou spracovania takto objemných dát. V práci predstavujeme nové efektívne algoritmy a dátové štruktúry pre manipuláciu s pangénomickými dátami. Prvou časťou našich výsledkov je skupina algoritmov pre konverziu z jedného kompresného formátu (straight-line program) do iného (LZ77) bez potreby dekompresie, v niektorých prípadoch v čase úmernom iba veľkosti komprimovaného vstupu a výstupu. V druhej časti popisujeme novú indexovú dátovú štruktúru, ktorá umožňuje rýchle exaktné vyhľadávanie vzorov alebo ich častí v komprimovanej reprezentácii pangénomického grafu bez falošných zhôd. Poslednou časťou je zlepšenie časovej zložitosti algoritmu pre vyhľadávanie maximálnych exaktných zhôd medzi vzorom a indexovaným textom, a týmto inšpirované heuristické zrýchlenie vyhľadávania dlhých zhôd v praxi.

**Kľúčové slová:** kompresia, algoritmy, pangénomika, textové indexy, repetitívne dáta