

Práca sa zaoberá možnosťami klastrovania inzercie so zameraním na realitnú inzerciu. V prvej časti práce definujeme čo to je klastrovanie, kde sa používa a aké sú typické požiadavky na klastrovacie algoritmy. Popíšeme existujúce klastrovacie metódy, ich vlastnosti a použitie. Posúdime ich vhodnosť pre oblasť inzercie a vyberieme najvhodnejší algoritmus pre klastrovanie rádovo miliónov inzerátov. V ďalšej časti detailne popíšeme interpretáciu inzerátu ako prvku vektorového priestoru s vysokou dimenziou a algoritmus klastrujúci prvky takéhoto vektorového priestoru založený na rodinách lokálnych hašovacích funkcií. Popíšeme jeho vlastnosti, časovú a pamäťovú zložitosť, jeho parametre a očakávané výsledky behu algoritmu. V implementačnej časti rozoberieme detaily implementácie v programovacom jazyku Java a navrhujeme vhodné uloženie dát v relačnej databázi. V časti venovanej testom potom zhodnotíme výsledky behu algoritmu na reálnych dátach a porovnáme ich s očakávaným výstupom algoritmu. V závere práce posúdime možnosti ďalšieho rozšírenia použitej klastrovacej metódy.