

Práce se věnuje automatizovanému hledání shluků tématicky podobných textových dokumentů v rozsáhlých textových kolekcích. V práci je navržen algoritmus pro nalezení těchto shluků a metoda pro optimalizaci jeho parametrů pomocí strojového učení. Byla provedena implementace a experimentální ověření funkčnosti navrženého postupu. Pro evaluaci je využita ručně anotovaná kolekce českých dokumentů obsahující množinu vzorových shluků a dále obsáhá kolekce novinových článků. Provedené experimenty ukazují, že výstupem navrženého algoritmu jsou požadované shluky tématicky podobných textů.