

Tato práce se zabývá problematikou vyhledávání dokumentů, které jsou si natolik podobné, že je můžeme považovat za (téměř) stejné, a to v kolekcích čítajících až miliony dokumentů. Největší důraz práce je kladen na porovnání nových, rychlých algoritmů řešících danou úlohu s algoritmy stávajícími, které jsou díky své složitosti pro obrovské kolekce nepoužitelné. Práce obsahuje implementaci obou metod přístupu k dané problematice spolu s aplikacemi umožňujícími experimentální porovnání obou těchto metod.