V této práci studujeme úlohy odhalování duplicit v databázích v rámci datové kvality. Za duplicity považujeme ty záznamy, které se sice mohou syntakticky lišit, ale které sémanticky představují tentýž objekt reálného světa. Hlavním cílem této práce je shrnout současné deduplikační metody z hlediska jejich nároků, výsledků a využitelnosti v praxi. Detailněji se zaměříme na porovnání dvou kategorií deduplikačních metod – těch, které vyžadují detailní informace o doméně, a těch, které se bez nich naopak dokáží obejít. Praktickou částí této práce je proto implementace vlastní metody z rodiny vzdálenostních metod nevyžadující žádné znalosti, jejíž výsledky porovnáme s výsledky komerčního nástroje používaného v praxi, který naopak využívá detailních znalostí dat, ve kterých jsou hledány duplicity.