

Algoritmy na porovnávanie súborov a pomocné programy „diff“, „patch“ a „diff3“ sa často používajú v programovaní za účelom porovnávania kódu a v mnohých systémoch správy verzií. Napriek užitočnosti sú výsledky týchto nástrojov striktne orientované na riadky, čo môže komplikovať spracovanie rôznych formátov textov, ako je napríklad obyčajný text, značkový jazyk a rôzne ďalšie formáty, v ktorých nie sú zlomy riadkov významné. Táto práca popisuje a implementuje prispôsobiteľnú verziu týchto nástrojov, ktoré umožňujú používateľovi určiť ľubovoľnú tokenizáciu vstupu, čoho dôsledkom ľahké diferencovanie, opravovanie a zlučovanie zmien obsahu, ktorý nie je podporovaný tradičným programom „diff“. Práca sa zaoberá novovzniknutým problémom vysporadúvania sa s bielymi znakmi.