

XML dokumenty a technologie reprezentují široce akceptovaný standard pro správu a výměnu semistrukturovaných dat. Překvapivě vysoký počet XML dokumentů však obsahuje chyby dobré formovanosti, strukturální validity nebo nekonzistence dat. Cílem této práce je analýza existujících přístupů vedoucí k návrhu nového korekčního systému. Představený model zahrnuje opravy elementů a atributů vůči jednotypovým stromovým gramatikám. Průchodem stavového prostoru automatu na rozpoznávání regulárních výrazů jsme vždy schopni nalézt všechny minimální opravy. Tyto opravy jsou kompaktně reprezentovány rekurzivními multigrafy, které se dají přeložit do konkrétních sekvencí editačních operací modifikujících datové stromy. Navrženy byly čtyři konkrétní algoritmy doplněné o prototypovou implementaci a experimentální výsledky. Nejvíce efektivní algoritmus heuristicky sleduje pouze perspektivní směry oprav a brání jakýmkoli opakovaným výpočtům.