

Jazyk XML se v dnešní době stává stále důležitějším formátem pro uchování a výměnu dat. Porovnávání podobnosti XML dat hraje zásadní roli v efektivním ukládání, zpracovávání a manipulaci s daty. Tato práce se zabývá možnostmi jak zjišťovat podobnost mezi DTD. Napřed je navržena vhodná reprezentace DTD stromů. Následně je navržen také algoritmus, který je založený na editační vzdálenosti stromů. Nakonec se zaměřujeme na různé aspekty podobnosti, jako jsou například strukturální a lingvistické informace, a snažíme se je zahrnout do naší metody.