

XML documents and related technologies represent widely accepted standard for managing and exchanging semi-structured data. However, surprisingly high number of XML documents is affected by well-formedness errors, structural invalidity or data inconsistencies. The aim of this thesis is the analysis of existing approaches resulting to the proposal of a new correction framework.

The introduced model involves repairs of elements and attributes with respect to single type tree grammars. Via the inspection of the state space of an automaton recognising regular expressions, we are always able to find all minimal repairs. These repairs are compactly represented by recursively nested multigraphs, which can be translated to particular sequences of edit operations altering data trees. We have proposed four particular algorithms and provided the prototype implementation supplemented with experimental results. The most efficient algorithm heuristically follows only perspective repair directions and avoids repeated computations using the caching mechanism.