XML documents and related technologies represent one of the most widespread ways how data on the Web are maintained and interchanged. Unfortunately, many of the real-world documents contain various types of consistency issues that prevent their successful automated processing.
In this thesis we focus on the problem of the structural invalidity and its correction. In particular, having one potentially invalid XML document modeled as a tree, and a schema in DTD or XML Schema languages modeled as a regular tree grammar, our goal is to find all the minimal corrections of this tree.
The model we proposed builds on top of the recursively nested structures of correction multigraphs, where the shortest paths are being found. For this purpose we formally introduce three correction strategies with different pruning optimizations applied. According to the experiments we performed, the refinement correction strategy not only significantly outperforms all the other existing approaches, but also guarantees important characteristics the others cannot.