

In this work we expand upon the previous efforts to infer schema information from existing XML documents. We find the inference of structure to be sufficiently researched and focus further on integrity constraints. After briefly introducing some of them we turn our attention to ID/IDREF/IDREFS attributes in DTD. Building on the research by Barbosa and Menelzon (2003) we introduce a heuristic approach to the problem of finding an optimal ID set. The approach is evaluated and tuned in a wide range of experiments.