In the presented work we study possiblities and limitations of benchmaring XML management systems (XMLMSs). We start by an analysis of existing solutions, continue by proposal and implementation of our XMLMS benchmark generator and finish with an application of it on a selected group of XMLMSs. The proposed benchmark is unique because it generates both testing data sets and operations (instead of being just a single fixed set). The work includes experimental results that depict advantages and disadvateges of the solution.