

The aim of this thesis is to research the current possibilities and limitations of automatic generation of synthetic XML and JSON documents used in the area of Big Data. The first part of the work discusses the properties of the most used XML data generators, Big Data and JSON generators and compares them. The next part of the thesis proposes an algorithm for data generation of semistructured data. The main focus of the algorithm is on the parallel execution of the generation process while preserving the ability to control the contents of the generated documents. The data generator can also use samples of real data in the generation of the synthetic data and is also capable of automatic creation of simple references between JSON documents. The last part of the thesis provides the results of experiments with the data generator exploited for the purpose of testing database MongoDB, describes its added value and compares it to other solutions.