

This thesis deals with the problematics of detecting documents, which are so similar one to another, that we can consider them to be (nearly) identical and that in collections having up to millions of documents. The greatest aim of this thesis is a comparison of new, fast algorithms designed to solve this task with current algorithms, which due to their complexity cannot be used for large collections. The thesis contains an implementation of both new and current methods of solving the given task together with applications that are designed to experimentally compare these methods.