

Abstract: This master thesis analyses the methods used for duplicity document detection and possibilities of their integration with a web search engine. It offers an overview of commonly used methods, from which it chooses the method of approximation of the Jaccard similarity measure in combination with shingling. The chosen method is adapted for implementation in the Egothor web search engine environment. The aim of the thesis is to present this implementation, describe its features, and find the most suitable parameters for the detection to run in real time. An important feature of the described method is also the possibility to make dynamic changes over the collection of indexed documents.