

This thesis surveys possibilities of clustering of advertisements, especially those for real estates. It defines clustering itself, its usage and typical requirements for clustering algorithms. We provide list of existing clustering methods and approaches, their properties and suitable application. We consider possibility of using them for clustering of millions of advertisements and based on that, we choose most suitable algorithm for this problem. We describe how to interpret advertisement as the point in multi dimensional vector space and this algorithm for clustering such points using locality of families of hash functions. We describe algorithm in detail, listing all of its parameters, estimating its complexity and expected results. In the following chapters we describe implementation of the algorithm in Java. We also describe database structure of underlying relational database. In the next chapter we present results of the algorithm based on real data and we compare the results with the expected results of the algorithm. In the end, we discuss possibilities for future extension of the clustering method.