Thomas Hartmann

Application of text and image mining algorithms in a two-stage classification of patent documents from the field of mechanical engineering

Abstract

In the practical work of machine-engineering exists a very successful *method of developing new ideas* of products. The method stimulates the pictorial-intuitive associating in the brain which shall lead to new product ideas. The basic idea of this method for finding ideas is the fact, that new technical designs can very often be combined by well-known principles of former techniques. The patent documents and their images contain the necessary information. In the present scientific paper is explained how the method has been optimized. The selection of thematically related patents is handled automatically and therefore the method is accelerated.

Procedures of automatical text- and image-classifications have been examined and tested by means of a successful handling of the digitized patent-documents. Various algorithms for text- and image-mining have been used to extract suitable features of the digitized documents so as to assign the documents to a category.

Out of these findings a way of computer based selection of patents has been developed. The selected patents fulfill two requirements: they contain a certain technical function (e.g. locking something) in as much as possible spread applications (e.g. locking the heel-clamp of ski-boots , kitchen-cupboards or loading door of washing-machines) and they additionally contain patent-images which are suitable for the *method of developing new ideas.*

The results of the reference-tests, which is in step with actual practice, show the successful application of the new developed procedure at patent-documents.

As a result of this research work there is a recommendation of detailed process steps and all necessary parameters.

Consequently, design engineers are able to apply the promising *method of developing new ideas* which is now based on an accelerated procedure.