Abstract

The solution for a feature selection problem is presented in the field of document image processing. The choice of shape features for describing glyphs of historical documents is a non-trivial task since the variations of glyphs in different documents is innumerable. Hence, the manual selection of shape features would be a cumbersome task. To select a subset of features from a given set a genetic algorithm is used which optimises the result of a clustering process by x-means. The result of x-means is evaluated by using different quality measures. The optimisation methodology is illustrated within a case study, in which the selection of an appropriate set of features is a crucial part of the system. The intended application supports a user who is transcribing historical documents by showing him similar occurrences of a given glyph.

References

- SBPK Berlin, Phillipps 1870, fol. 11v, 1870.
Choosing Shape Features by means of Genetic Algorithms for Glyph-clustering of Historical Documents

Choosing Shape Features by means of Genetic Algorithms for Glyph-clustering of Historical Documents

DocEng '12, New York, NY, USA, 2012. ACM.


**Index Terms**

Computer Science

Artificial Intelligence

**Keywords**

Document Image Processing  Genetic Algorithms  Feature Selection  Shape Descriptions  Glyph Clustering

X-Means