Abstract

This paper presents a methodology for tackling the authorship verification problem. The approach is based on comparing the similarity between a given unknown document against the known documents using a graph representation that captures the syntactic sequence of texts and a graph similarity measure. An unknown document can be classified as having been written by the same author if the majority of the comparisons surpass a predefined threshold. The best results were obtained on the Clef PAN 2014 dataset: 79% for the Spanish and 68% for English, showing that the proposed methodology could be a way for determining a document authorship.

References


Index Terms

Computer Science Security

Keywords

Authorship Verification, Syntactic Sequence Graph, Graph Similarity