

Abstract

In this paper we present a clustering method for XML documents. Our step is twophase based: we first automatically extract the structure from the document; we then use it as model of representation to classify the document that it represents. The matching of the documents' structures is based on the calculation of their similarities. For the experimentation we used the INEX