

Abstract

XML (Extensible Markup Language) is a general specification to create a custom markup language. XML is classified as an extensible language, because it allows users to define their own markup elements, purpose of XML is to help information systems in a variety of data structures, particularly through the internet, besides that it is also designed to encode the documents, and for serialization data. This makes the XML is widely adapted and made the researchers to make for XML Information Retrieval. In this final task will be develop structured Information Retrieval using XML. SIMNOMERGE is the algorithm used for rank document. For the weighting in the SimNoMerge algorithm used three kinds of weightings. Weighting using the TF, IDF, and TF-IDF. weightings output are then analyzed. This is to measure the performance of the algorithm based on specific weights.

Result showed that the algorithms can be used to *SimNoMerge* do ranking of XML documents, although a test score is not high precision, whereas for test recall having a better value, this indicates that the algorithm can perform SimNoMerge of ranking result in many irrelevant documents. By comparison the value of averages of precision and recall from the testing, was found that applying preprocessing with combination of all three types of weighting was better to apply on structured information retrieval instead of not applying preprocessing.

Keywords : *Structured Information Retrieval*, SimNoMerge, TF, IDF, TF-IDF