

# Jordan University of Science and Technology

## A Hierarchical K-NN Classifier for Textual Data

**Authors:** Rehab Duwaiir, Rania Al-Zubaidi

**Abstract:** This paper presents a classifier that is based on a modified version of the well-known K-Nearest Neighbors classifier (K-NN). The original K-NN classifier was adjusted to work with category representatives rather than training documents. Each category was represented by one document that was constructed by consulting all of its training documents and then applying feature selection so that only important terms remain. By this, when classifying a new document, it is required to be compared with category representatives and these are usually substantially fewer than training documents. This modified K-NN was experimented with in a hierarchical setting, i.e. when categories are represented as a hierarchy. Also, a new document similarity measure was proposed. It focuses on co-occurring or matching terms between a document and a category when calculating the similarity. This measure produces classification accuracy compared to the one obtained if the cosine, Jaccard or Dice similarity measures were used; yet it requires a much less time. The TrechTC-100 hierarchical dataset was used to evaluate the proposed classifier.