



## *Text Classification using Ontology Graph Representation through Bag of Words*

<sup>1</sup>Kavya S N , <sup>2</sup>Dr Sudhamani M , <sup>3</sup>Jagadeesh M and <sup>4</sup>Abhishek M Anegundi

<sup>1</sup>Department of Computer Science, MMK & SDM MMV, Mysuru.

<sup>2</sup>Department of Computer Science, MMK & SDM MMV Mysuru,

<sup>3</sup>Department of Computer Science, University of Mysore.

<sup>4</sup>Department of Computer Science, University of Mysore.

### Abstract

Text classification is the process of automatically categorizing text documents into a set of predefined classes. Ontology deals with similar terms and relationship that can be used to describe and represent area of knowledge. In this work, we categorize text to a given ontology by using the Knowledge of the document, knowledge represented in the form of ontology for categorizing documents. We experimented the on 20 newsgroup mini dataset, 20 newsgroup large dataset and Reuters-21578 datasets. Support Vector Machine (SVM) classifier is used to classify the text documents and performance of the classifier is measured in terms of accuracy and f-measures.

**Keywords:** Text categorization, ontology, Support Vector Machine (SVM), knowledge representation, feature extraction.

### 1. Introduction

Text classification methods are mainly used to determine the belonging of text document. The purpose of classification is to identifying the belongingness of text. Rapid development of information technology directly affects an increase in the amount of data. Automatic text classification methods enable the organization or categorization of a set of documents into different categories or classes. Many classification problems have been solved manually by the use of some rules commonly written by hand. Instead of using hand-written rules, the text categorization approaches use machine learning methods to learn automatic classification rules based on human labelled documents. It is obvious labelling is easier than defining rules. The rapid growth in data volume increases the complexity of classification; it also requires a lot of time and human effort for manual classification. Therefore, automated classification of the electronic documents needed. Hence, text categorization can be considered as an effective method for automatic assignment of documents to the predefined categories according to their context. In this paper, we consider the tasks of the automated classification of 20NewsgroupLarge dataset, 20 Newsgroup Mini dataset, REUTERS-21578(Transcriptions).

Text classification is the process of assigning tags or categories to text according to its content. It's one of the fundamental tasks in **Natural Language Processing (NLP)** with broad applications such as sentiment analysis, topic labelling, spam detection, and intent detection. Unstructured data in the form of text is everywhere; emails, chats, web pages, social media, support tickets, survey responses, and more. Text can be an extremely rich source of information, but extracting insights from it can be hard and time-consuming due to its unstructured nature. Businesses are turning to text classification for structuring text in a fast and cost-efficient way to enhance decision-making and automate processes.

It is observed that information extracted is enormous, adequate and are unstructured in nature. Automatically retrieving such huge information from a database could be one of the most difficult tasks in the field of information retrieval. In machine learning and text processing, Text classification plays an important role in deciding the label for the document. Much research work has been done in single label text classification problems, where single-labelled document means a document belongs to single class. Few works have done in

*Bharath N*

PRINCIPAL

MMK & SDM Mahila Mahavidyalaya  
Krishnamurthunuram, Mysore-570 004