Abstract

This study is to examine paraphrase detection (PD) for diagnostic purposes. Which is defined as the capability to find and discover the similarity between sentences that are written in a natural language? Where detecting similar sentences written in natural language is extreme importance and it is very essential for computer software used in plagiarism detection, Q and A automated systems, text mining, authorship authentication and text recapitulation. The goal of paraphrase detection is to detect whether two statements have the identical semantic or not. There is hundreds of empirical research in this direction. This study will focus on the discussion of recent studies of the PD methods and will categorize them in two categories, supervised learning and unsupervised learning. Also will give an idea about text similarity, machine learning and deep learning approaches. The performance of the selected researches is assessed by how accurate the F-measures are in detecting paraphrase in Microsoft Research Paraphrase Corpus (MSPR).


Index Terms

Computer Science Information Sciences

Keywords

Paraphrase Detection, Text Similarity, Machine Learning, Deep Learning, Microsoft Research Paraphrase Corpus.