

# Edit Distance

## Introduction

Given two strings “word1” and “word2”, what is the minimum number of operations required to convert the first word “word1” into the second word “word2”. That was a very simple sentence that describes the edit distance measure.

Edit distance is a measure of similarity between two strings. Its mission is to determine the minimum number of changes required to convert string x into string y, which is done by using one of the three operations (Insert/Delete/Replace).

## Application

Edit distance is a very common measure used nowadays, many applications are supported by the edit distance measure especially the “Levenshtein distance”, following are common usages of edit distance

- Natural language processing: NLP is used by edit distance to detect spelling mistakes and then suggest some corrections for a misspelled word.
- Bioinformatics: bioinformatics is related to genetics which uses edit distance technology to quantify the similarity of DNA sequences, it can be viewed as strings of the letters A, C, G, and T.
- Speech Recognition: some speech recognition systems use edit distance measures to find very difficult audio signals, it finds a close match between a new voice and the stored old voice.
- Plagiarism Detection: identifying copied content when working on an assignment or graduation research projects is very important, here comes the edit distance role, its features allow it to detect stolen information such as stealing another person’s words or intellectual property without having credit.

However, edit distance has some disadvantages as well. It does not consider the semantic or the meaning of the strings. In addition, it does not consider the weight or importance of each character.