Assignment 1: Abstract and Introduction summary

- To process biological data we need a lot of computational resources and time, both of them are costly for us, we would like to process the data in the least time possible with the least resources available.
- the problem is, biological data are enormous in size and very complex, we need efficient ways to analyze the data
- the kind of analysis we are interested in is pattern matching, we would like to search for specific pattern which can range from searching for specific disease or mutation ..etc
- Finding such similarities is a challenging research area, comprehending Big Data, that can bring a better understanding of the evolutionary and genetic relationships among the genes
- the paper will study and examine different kinds of algorithms for pattern matching, and also assessing complexity, performance
- Keyword search and matching are techniques to discover patterns inside specific strings. Algorithms for matching, are used to discover matches between patterns and input strings
- we can consider the pattern as P and is shorter than the text T, and we need to find all occurrences of P in T to extract valuable insights
- There are two main approaches for string matching, first is exact matching, for instance: Smith-Waterman (SW); Needleman Wunsch (NW); Boyer Moore Horspool (BMH); Dynamic Programming; Knuth Morris Pratt (KMP). Second approach, is approximate matching, also known by Fuzzy string searching, for instance: Rabin Karp; Brute Force
- in this paper, we apply those concepts in Protein, DNA and RNA, using for that effect, different types of string matching algorithms
- examples : NW algorithm, Boyer Moore (BM), SW algorithm, Hamming Distance, Levenshtein Distance, AhoCorasick (AC), KMP, Rabin Karp, and

CommentZwalter (CZW)