Chapter 14
A Comparative Analysis of Software Engineering Approaches for Sequence Analysis

Muneer Ahmad
King Faisal University, Saudi Arabia

Low Tang Jung
University Technology PETRONAS, Malaysia

Noor Zaman
King Faisal University, Saudi Arabia

ABSTRACT

DNA is considered the building block of living species. DNA sequence alignment and analysis have been big challenges for the scientists for many years. This research presents a comparative analysis of state of the art software engineering approaches for sequence analysis, i.e. genome sequences in particular. Sequence analysis problems are NP hard and need optimal solutions. The underlying problems stated are duplicate sequence detection, sequence matching by relevance, and sequence analysis by approximate comparison in general and by using tools, i.e. Matlab and multi-lingual sequence analysis. The usefulness of these operations is also highlighted, and future expectations are described. The proposal describes the concepts, tools, methodologies, and algorithms being used for sequence analysis. The sequences contain the precious information that needs to be mined for useful purposes. There is high concentration required to model the optimal solution. The similarity and alignments concepts cannot be addressed directly with one technique or algorithm; a better performance is achieved by the comprehension of different concepts.

INTRODUCTION

Sequences are logical units that contain vital information, for instance consider biological sequences that compose of nucleotide base pairs in the form of A (Adenine), T (Thymine), G(Guanine) and C (cytosine). The structure and position of these pairs in sequence determine the personality, habits and inheritance characteristics of species.

The mining of useful information from the vast repositories of sequence data brings interesting results related to genes and their functional properties, the main attention and focus of biologists is to differentiate species on behalf of these functional characteristics, many different solutions have been proposed that claim to bring optimal results. It is worth knowing that direct matching in sequence repository data is not efficient and may bring inaccurate and slow results, so going beyond the exact match is necessary for optimality.

Modern computational technology and good devices has made the job of scientists relatively easy in bringing accurate results, this reflection is quite positive in micro-array DNA technology and image data-sets comparison techniques where huge bulky genetic data is approximately compared promptly.

The data is spread over chips and relevancy is determined. The other tools like MATLAB, TRADOES and EBMT are now broadly used for sequence manipulation. FASTA and BLAST are also very popular in biological researchers for sequence comparisons, different people have developed many tools for analysis of not only the genetic sequences but corpora sequences, the lexical analysis explores the hidden resources in these structures, global alignment tools have replaced local one and multiple alignment techniques have given way to know more about diversity in functional properties of species in sequences.

People are interested in mining some kind of association rules in genetic and lexical data, these rules will better help to understand the patterns in data and further exploration may lead to more knowledgeable and interesting results that could not be available by query application phenomenon. The query application only generates views that are provided through datasets within a confined domain and redefined rules in the form of queries, later solutions present the query enhancement techniques but that are not as optimal as direct rule generation from datasets.

Scientists now use latest systems in biotechnology for storage of genetic data, employing data ware housing techniques and analyzing the DNA sequences, it is not limited to computations but can solve many different complex biological problems.

A BRIEF DISCUSSION OF RESEARCH OVER SEQUENCE ANALYSIS

We have presented a brief summary of research done for sequence analysis, Bansal (1995) has presented a sophisticated way for the presentation of information in the form of a frame work which helps understanding multiple sequences as ADT (abstract tool to integrate information achieved from proposed idea). The authors tried to develop a library that could contain generic high level language for generation of useful phylo-genetic tree. Complex analysis at multiple sequences had been performed by derived groups of amino acids in homologous protein (sharing some common properties along with identification of constrained columns. A help from PROLOG TOOL was taken to apply proposed frame work.

Kappen and Salbaum (2003) had presented a sequence comparison approach that could annotate a comparison between a mouse chromosome 9 and a human chromosome 15, the data draft sequences had been obtained from genetic databases and a complex map containing 14 genes has been presented as a genome map, the framework described in the paper for data interpretation and demonstration can be quite helpful for generation of more complex maps provided time constrained is kept

Related Content

A Contribution to the Specification of Model Transformations with Metamodel Matching Approach
Karima Berramla, El Abbassia Deba, Abou El Hassen Benyaminaand Djilali Benhamamouch (2017). *International Journal of Information System Modeling and Design (pp. 1-23).*
www.irma-international.org/article/a-contribution-to-the-specification-of-model-transformations-with-metamodel-matching-approach/204369

Is Modeling a Treatment for the Weakness of Software Engineering?
Janis Osisand Erika Asnina (2015). *Handbook of Research on Innovations in Systems and Software Engineering (pp. 411-427).*
www.irma-international.org/chapter/is-modeling-a-treatment-for-the-weakness-of-software-engineering/117935

Requirements to Products and Processes for Software of Safety Important NPP I&C Systems
Vladimir Sklyar, Andriy Volkoviy, Oleksandr Gordieievand Vyacheslav Duzhyi (2022). *Research Anthology on Agile Software, Software Development, and Testing (pp. 212-246).*
www.irma-international.org/chapter/requirements-to-products-and-processes-for-software-of-safety-important-npp-ic-systems/294466

Adapting Agile Practices During the Evolution of a Healthcare Software Product
Danilo F. S. Santos, André Felipe A. Rodrigues, Walter O. Guerra Filhoand Marcos Fábio Pereira (2022). *Research Anthology on Agile Software, Software Development, and Testing (pp. 777-791).*
www.irma-international.org/chapter/adapting-agile-practices-during-the-evolution-of-a-healthcare-software-product/294494

Best Practices Guidelines for Agile Requirements Engineering Practices
Chetankumar Pateland Muthu Ramachandran (2010). *Handbook of Research on Software Engineering and Productivity Technologies: Implications of Globalization (pp. 1-14).*
www.irma-international.org/chapter/best-practices-guidelines-agile-requirements/37020