

Chapter 10

State-of-the-Art Information Retrieval Tools for Biological Resources

Shashi Bhushan Lal

*ICAR-Indian Agricultural Statistics
Research Institute, India*

Mohammad Samir Farooqi

*ICAR-Indian Agricultural Statistics
Research Institute, India*

Anu Sharma

*ICAR-Indian Agricultural Statistics
Research Institute, India*

Sanjeev Kumar

*ICAR-Indian Agricultural Statistics
Research Institute, India*

Krishna Kumar Chaturvedi

*ICAR-Indian Agricultural Statistics
Research Institute, India*

Dwijesh Chandra Mishra

*ICAR-Indian Agricultural Statistics
Research Institute, India*

Mohit Jha

ICAR-Indian Agricultural Statistics Research Institute, India

ABSTRACT

With the advancements in sequencing technologies, there is an exponential growth in the availability of the biological databases. Biological databases consist of information and knowledge collected from scientific experiments, published literature and statistical analysis of text, numerical, image and video data. These databases are widely spread across the globe and are being maintained by many organizations. A number of tools have been developed to retrieve the information from these

DOI: 10.4018/978-1-5225-2483-0.ch010

Copyright ©2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

databases. Most of these tools are available on web but are scattered. So, finding a relevant information is a very difficult, and tedious task for the researchers. Moreover, many of these databases use disparate storage formats but are linked to each other. So, an important issue concerning present biological resources is their availability and integration at single platform. This chapter provides an insight into existing biological resources with an aim to provide consolidated information at one place for ease of use and access by researchers, academicians and students.

INTRODUCTION

Sanger, way back in 1988, discovered successful methods for sequencing of proteins, Ribonucleic Acid (RNA) and Deoxyribonucleic Acid (DNA) which opened a new era in biological science (Sanger, 1998). In the world of biological sciences, when the advancement in sequencing technologies has made human genome sequenced to a great extent, one cannot imagine a world without nucleotide and protein sequences (Stretton, 2002). A very large number of researchers in life sciences have been working since then, which has generated an enormous amount of biological sequences and its derived data. SWISS-PROT database was developed in the year 1986 which had around seventy thousand protein sequences from five thousand different organisms.

After Sanger, emergence of further refined technologies for genome sequencing were discovered, namely Pacific Biosciences, Ion Torrent, 454/Roche, Illumina, SOLiD and so on. This technological advancements have led to the whole genome sequencing of a wide range of species across animals, insects, plants and human. Being the fastest growing area in biological science, there has been a remarkable increase in the volume of biological data. These data are sequences obtained from experiments, published literature and their computational and statistical analyses. The format of data obtained can be in the form of either text, numbers, videos, images or diagrams (Schaller, Mueller, & Sung, 2008). These data are available on the public domain for carrying out research in biological science.

Although a large number of these databases are available on open domain, yet their retrieval system has always been a challenge for the developers (Kamal et al., 2016; Singh, Sharma, & Dey, 2015). Many attempts have been made by researchers to integrate these databases and develop an efficient retrieval system. Information retrieval from these databases are maintained by various organizations to provide ease of access to the end users. But, due to the rapid growth in the biological data, requirement of computational resources for its storage and retrieval system always remain a challenge for computational experts to meet the growing needs.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/state-of-the-art-information-retrieval-tools-for-biological-resources/178375

Related Content

Tracking How a Change in a Telecom Service Affects Its Customers Using Sentiment Analysis and Personality Insight

Ammar Adland Abdelsadeq Khamis Elfergany (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1733-1749).

www.irma-international.org/chapter/tracking-how-a-change-in-a-telecom-service-affects-its-customers-using-sentiment-analysis-and-personality-insight/308572

Patterns Relevant to the Temporal Data-Context of an Alarm of Interest

Savo Kordic, Chiou Peng Lam, Jitian Xiao and Huaizhong Li (2010). *Dynamic and Advanced Data Mining for Progressing Technological Development: Innovations and Systemic Approaches* (pp. 18-39).

www.irma-international.org/chapter/patterns-relevant-temporal-data-context/39635

A BPMN-Based Design and Maintenance Framework for ETL Processes

Zineb El Akkaoui, Esteban Zimányi, Jose-Norberto Mazón and Juan Trujillo (2013). *International Journal of Data Warehousing and Mining* (pp. 46-72).

www.irma-international.org/article/bpmn-based-design-maintenance-framework/78375

Machine Learning Approaches for Sentiment Analysis

Basant Agarwal and Namita Mittal (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1917-1933).

www.irma-international.org/chapter/machine-learning-approaches-for-sentiment-analysis/150249

Improved Approximation Algorithm for Maximal Information Coefficient

Shuliang Wang, Yiping Zhao, Yue Shu and Wenzhong Shi (2017). *International Journal of Data Warehousing and Mining* (pp. 76-93).

www.irma-international.org/article/improved-approximation-algorithm-for-maximal-information-coefficient/173707