Chapter 5
State-of-the-Art Information Retrieval Tools for Biological Resources

Shashi Bhushan Lal
ICAR-Indian Agricultural Statistics Research Institute, India

Anu Sharma
ICAR-Indian Agricultural Statistics Research Institute, India

Krishna Kumar Chaturvedi
ICAR-Indian Agricultural Statistics Research Institute, India

Mohammad Samir Farooqi
ICAR-Indian Agricultural Statistics Research Institute, India

Sanjeev Kumar
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 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.

In this chapter, a comprehensive compilation of existing biological databases have been provided at one place for easy accessibility of researchers, academicians and students. The organization of this chapter is as follows: first section deals with the nature of biological data and its organization, followed by types of biological databases. In the second section classification of biological databases, its detailed description with few examples are presented. Third section discusses about the means of interconnection between these databases followed by the drawbacks in the biological databases explained in the fifth section. In the sixth section, retrieval system of these biological databases have been discussed. Finally, conclusions are presented in the last section.

Related Content

Metadata for Search Engines: What can be learned from e-Sciences?
Magali Roux (2012). *Next Generation Search Engines: Advanced Models for Information Retrieval (pp. 47-77).*
www.irma-international.org/chapter/metadata-search-engines/64420

Comparative Analysis of Machine Learning-Based Algorithms for Detection of Anomalies in IIoT
Bhupal Naik D. S., Venkatesulu Dondetiand Sivadi Balakrishna (2022). *International Journal of Information Retrieval Research (pp. 1-55).*
www.irma-international.org/article/comparative-analysis-of-machine-learning-based-algorithms-for-detection-of-anomalies-in-iiot/298647

Interactive IR in Digital Library Environments
Iris Xie (2008). *Interactive Information Retrieval in Digital Environments (pp. 116-152).*
www.irma-international.org/chapter/interactive-digital-library-environments/24526

Artificial Intelligence and NLP-Based Chatbot for Islamic Banking and Finance
Shahnawaz Khanand Mustafa Raza Rabbani (2021). *International Journal of Information Retrieval Research (pp. 65-77).*
www.irma-international.org/article/artificial-intelligence-and--nlp--based--chatbot-for-islamic-banking-and-finance/280527

Gravizor: A Graphical Tool for the Visualization of Web Search Engines Results With Multi-Agent Based Modeling
Abdelkrim Bouramoul (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 1338-1359).*
www.irma-international.org/chapter/gravizor/198602