

## ACKNOWLEDGEMENT

I would like to begin by thanking my supervisor **Dr. Debasis Dash**, Senior Principal Scientist of CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), for giving me the opportunity to work in their labs. Their immense understanding and curiosity in research allowed me to learn many things. Their guidance and encouragement were indispensable in completing this dissertation.

Also, I would like to thank Rintu Kutum for constant encouragement and help in understanding Python programming.

My sincere thanks goes to all the staff, scientist and PhD scholars at CSIR-IGIB, whose kindness and friendly nature provided a very enjoyable research environment.

I wish to extend my sincere gratitude to all faculty members of Jaypee University of Information Technology, for providing me such a platform of excellence to do my dissertation.

**Aagam Mishra**

## **List of Figures:**

**Figure-01:** Flowchart representing the extraction of table contents and hrefs\_id

**Figure-02:** Flowchart representing the extraction of xml files

**Figure-03:** Flowchart representing the reading and writing of data from xml ids

**Figure-04:** Top 10 species proteomics data in jPOST database

## List of Tables

**Table-01:** Representation of frequencies of species Obtained from the jPOST database