**Web crawling to Match miRNA ID to expression**

**Narendra Banerjee, Rayshawn Walker**

**Advisor: Lin Chen**

**Abstract**

A web crawler was developed to match miRNA IDs to corresponding expressions. A miRNA ID usually corresponds to multiple expressions and an expression can be from multiple miRNAs. The complexity of building a connection between miRNA IDs and expression is a O(n^3). In our study of Cancer in African American’s and Caucasians, the search was implemented by a time-consuming manually process. Here we designed a web crawler to conduct matching by automatically searching and aligning records deposited in different databases hosted on internet. Two input files from Dr. Hirendra Banerjee were used as the inputs which contains transcript ID and gene expressions respectively. The web crawler in Python processed those information to create a output file which contains the matched records between miRNA ID and its corresponding transcript ID. The matching time has been reduced to a few hours from several months that it used to be.