# Topic Selection Proposal

## Introduction

Scientific applications often store datasets in self-describing data file formats, such as HDF5 and netCDF, however, existing solutions extract the metadata and store it in external database which offers a less efficient way to handle. In this project, I am going to achieve two tasks which are range query against numeric attribute values and compound metadata search with affix search against attribute name.

## Proposal

In this project, the goal is extending MIQS by implementing several specific functionalities, and based on my experience, I feel comfortable with Linux development environments, also delivering high quality C code. My plan is forking the project into my own repository and run the existing code first, then gradually working on the extending functionalities. The desired outcomes are successfully implementing the extending functions and try to make it perform better, I will look into the source code and find a good way to implement my functions.

## Tasks

* Range query against numeric attribute values
* Compound metadata search with affix search against attribute name and affix search/range query against attribute values

## Plan

|  |  |
| --- | --- |
| Date | Tasks |
| From 3-25 to 3-29 | Write project proposal and setup dev environment |
| From 3-30 to 4-10 | Finish range query against numeric attribute values |
| From 4-11 to 4-18 | Discuss with Jacob and finish Compound metadata search |

## Resources and Conclusion

The sources I need to study before I dive into the project is how radix tree works and the datasets stored in it. The general indexing data structures are also a needed researching material.