**Title:**  Mining Software Repositories

**Members:**
Russle Brock Wilcox (wilcox6)             Anu Pardeshi (apardesh)

**Abstract:**
Software Repositories keep the complete history of any file in the project, including who modified what, when, and the delta of the modification. They can be used to mine important information like how the project evolved, how the developers collaborated, their contributions, important milestones in the project’s development, predict changes in a software system, bring some insight on the design of the software, warn about the change couplings that need to be propagated to the other entities help during program comprehension in general and enlighten dependencies between parts of a system. These applications of data mining to software repository inspire us to summarize current approaches and recent developments.

We will build off of a previous survey, [1], and for each of the major topics discussed we will include some recent developments. We will also consider whether the list of topics needs to be revised, with groups added, removed, or merged. Topics, along with some of the papers we will read, include:

- Metadata analysis: [2], [3]
- Static source code analysis: [4]
- Source code differencing and analysis: [5], [6]
- Software metrics: [7], [8]
- Visualization: [9], [10], [11]
- Clone detection: [12], [13], [14], [15]
- Frequent-pattern mining: [16], [17]
- Information-retrieval methods: [18], [19]
- Classification with supervised learning: [20], [13]
- Social network analysis: [4], [21]


