# Team: “The flying miners”
# Assignment 1


| Name | Student number |
|---|---|
| Dimitris Theodorakopoulos | 4620534 |
| Andreas Maruli Christian Pangaribuan | 4608259 |
| Lucio Claudio Guerchi | 4632397 |
| Ioannis Papadopoulos | 4565002 |


To complete this project, we worked in Python using the following frameworks:

- GitPython
- Jira

To install these frameworks, we executed the following commands in the command line using “C:/Python27/Scripts/pip.exe”:

- pip install GitPython
- pip install Jira
- pip install requests[security]

We installed git for Windows and executed the following command in order to get the “/lucene-solr” directory:

- git clone git://git.apache.org/lucene-solr.git

We also implemented an alpha version of this project in native Java, in order to cross-check our results.

The algorithm that we used in order to collect our data and fill in the table works as follows:

For every java **file**:
  For every **commit** that involved this file:
    If commit date between (2014-01-02 00:23:24 and 2015-02-01 00:01:00)
      then gather information about {**minor**, **major**, **total**, **ownership**}
    If commit date between (2015-02-01 00:02:00 and 2015-08-01 00:00:00)
      then 1) get issue-id from commit message
           2) gather information about **bug_count** from Jira using the issue-id

If the proportion of ownership for a particular file is above or equal to 5%, then the contributor is “major”, otherwise the contributor is “minor”.
The “ownership” property is calculated as the highest proportion of ownership for a particular file.
We determined if an issue-id is related to a bug or not by looking at the property “issuetype.name” in the json response from Jira. For example, if we look at the following the json reponse

https://issues.apache.org/jira/rest/api/2/issue/LUCENE-4797 for issue-id “LUCENE-4797”, we will see that variable **issuetype.name** is set to **”Bug”**.

Based on all of the above, we printed each row to the console to create our data table.