Summary of the Codes

* **code\_included\_ref\_download.py:** The bibliographic information of each reference included in a systematic review are extracted from the online Cochrane Library. This code is written in Python.
* **code\_getPM.java:** Titles of all included references in a systematic review are the input to these code. This java code returns a list of PubMED ids by using eutils search API. All the results which have less than 5 PubMED ids are considered to retrieve their citation information from PubMed. This is a Java code.
* **code\_json\_xml.java:** This code extracts inclusion statements (a data element of interest) from the JSON file corresponding to the PDF of each systematic review article. The PDF is converted to JSON using PDFx tool. The inclusion statements along with the reference are stored in XML format by this code. This code is written in Java.
* **code\_extract\_incl\_stmt\_from\_xml.py:** Extracts inclusion statements from the XML files and write them into a text file along with the references where they appear. This is a python code.
* **code\_create\_training\_set.py:** Creates the training corpus using the inclusion statements. The code takes the text file as input, where all the inclusion statement along with the references are stored. Moreover, all the references in plain text format are taken as the input. Therefore it extracts the actual sentence for each inclusion statement from the corresponding reference and stores the output in separate files. Similarly it generates some negative sentences for each inclusion statement and store them in different files. All these output files build the gold standard. This code is written in Python.
* **code\_extract\_inclusion\_stmt\_from\_test\_samples.py:** The inputs to this code are the training corpus created by the earlier code and the test samples in text format. The code extracts the inclusion statements from these test samples. SVM is used as the classifier in this code. This code is written in Python. Scikit-learn, a machine learning tool is used in this code.