Instructions for using the source code

**1. Configure the database.** The data file can be downloaded from the URI “<https://pan.baidu.com/s/1N56EzVhW0M-qPWZQzORGtw>”. The log file can be downloaded from the URI “<https://pan.baidu.com/s/1qwqD7CNCqjwRs1tgvpuYUQ>”.Then, you can load these two files to the local database “SQL Server 2017”.

**2. Import the project.** You can create a new java project by using Eclipse tool. Then, you can import the project from the folder to get all the source codes.

**3. Import all the jar files.** All the jar files are placed in the folder “lib”.

**4. Modify the configure file “com/hctc/parse/webfilter/BeanConfig.xml”.** You can modify the following configuration in the file “BeanConfig.xml”.

<bean id=*"featureFilterTask"* class=*"com.hctc.parse.webfilter.FeatureFilterRun"*>

<property name=*"dbFeatureFilterUtil"* ref=*"dbFeatureFilterUtil"* />

<property name=*"nodeTool"* ref=*"nodeTool"* />

<property name=*"linksTool"* ref=*"linksTool"* />

<property name=*"containStringFilterFlag"* value=*"false"* />

<property name=*"containsTagFilterFlag"* value=*"false"* />

<property name=*"linksRateFilterFlag"* value=*"false"* />

<property name=*"containChildTreeFilterFlag"* value=*"false"* />

<property name=*"simRateFilterFlag"* value=*"false"* />

<property name=*"tfIdfFilterFlag"* value=*"false"* />

<property name=*"dndFilterFlag"* value=*"false"* />

<property name=*"tfIdf"* ref=*"tfIdf"* />

<property name=*"dndM"* ref=*"dndM"* />

</bean>

1) If you use the filtering model MFPCS, you need to modify the property *containStringFilterFlag* to **true**, and modify these properties to **false** including *containsTagFilterFlag*, *linksRateFilterFlag*, *containChildTreeFilterFlag*, *simRateFilterFlag*, *tfIdfFilterFlag*, *dndFilterFlag*.

2) If you use the filtering model MFPCTA, you need to modify the property *containsTagFilterFlag* to **true**, and modify these properties to **false** including *containStringFilterFlag*, *linksRateFilterFlag, containChildTreeFilterFlag, simRateFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

3) If you use the filtering model MFPCT, you need to modify the property *containChildTreeFilterFlag* to **true**, and modify these properties to **false** including *containStringFilterFlag, linksRateFilterFlag, containStringFilterFlag, simRateFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

4) If you use the filtering model MFPLR, you need to modify the property *linksRateFilterFlag to* **true**, and modify these properties to **false** including *containsTagFilterFlag, containStringFilterFlag, containChildTreeFilterFlag, simRateFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

5) If you use the filtering model MFPSD, you need to modify the property *simRateFilterFlag* to **true**, and modify these properties to **false** including *containsTagFilterFlag, linksRateFilterFlag, containChildTreeFilterFlag, containStringFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

6) If you use the filtering model **TagAndString**, you need to modify the two properties *containStringFilterFlag* and *containsTagFilterFlag* to **true**, and modify these properties to **false** including *linksRateFilterFlag, containChildTreeFilterFlag, simRateFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

7) If you use the filtering model **TagAndLinks**, you need to modify the two properties *linksRateFilterFlag* and *linksRateFilterFlag* to **true**, and modify these properties to **false** including *containsTagFilterFlag, containChildTreeFilterFlag, simRateFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

8) If you use the filtering model **TagAndSim**, you need to modify the two properties *simRateFilterFlag* and *containsTagFilterFlag* to **true**, and modify these properties to **false** including linksRateFilterFlag, *containChildTreeFilterFlag, containStringFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

9) If you use the filtering model **LinksAndString**, you need to modify the two properties *linksRateFilterFlag* and *containStringFilterFlag* to **true**, and modify these properties to **false** including *containsTagFilterFlag, containChildTreeFilterFlag, simRateFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

10) If you use the filtering model **LinksAndSim**, you need to modify the two properties *linksRateFilterFlag* and *simRateFilterFlag* to **true**, and modify these properties to **false** including *containStringFilterFlag, containChildTreeFilterFlag, containsTagFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

11) If you use the filtering model **StringAndSim**, you need to modify the two properties *containStringFilterFlag* and *simRateFilterFlag* to **true**, and modify these properties to **false** including *linksRateFilterFlag, containChildTreeFilterFlag, containsTagFilterFlag, tfIdfFilterFlag, dndFilterFlag*.

In addition, you need to modify your connection parameters of database in the file “BeanConfig.xml”.

**5. Start the program.** You can execute the class “StartFilter” to start the program.