**小组研究内容：基于区块链的代码版权保护方法研究**

**小组成员：徐嘉滢、代舒璐、卢慧颖、陈宏宇**

**写在前面**

我们小组要做的是基于区块链的代码版权保护方法，具体的实现方法为：运用antlr4工具将code转化为AST，将AST通过遍历、ngrams算法、MD5加密算法变成一串hash set，由此进行后面的相似度比较和上链过程。

**文件介绍：**

文件夹**Javalib** 即搭建ANTLR4环境时候使用，后文详述

文件夹**Test\_result** 即Code1和Code2从代码到Hash set的过程中间产物

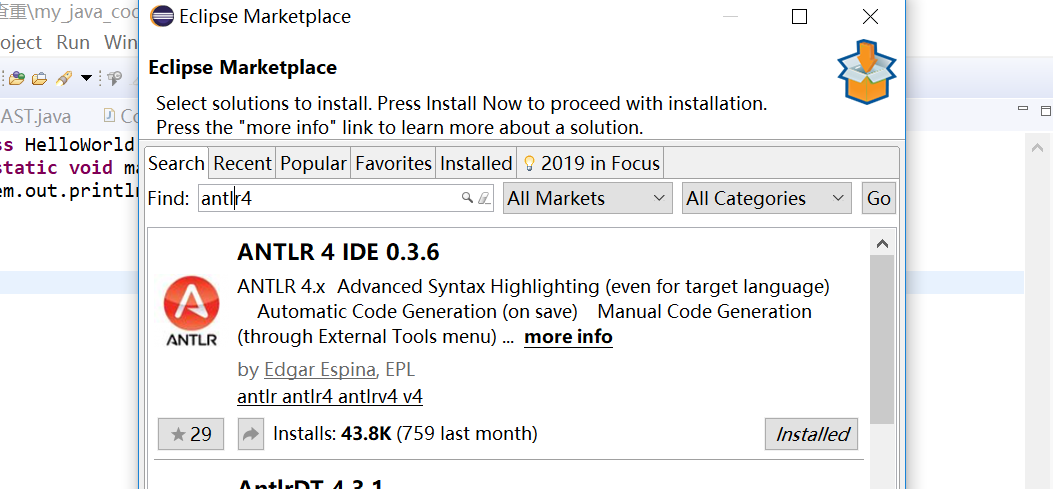
文件夹**Test\_code** 即Code1和Code2文件

两个**.jar**文件 即搭建区块链部分代码环境时候使用，后文详述

压缩包**My\_java\_code.zip** 即“My\_java\_code” project，import后可以查看所有执行代码

1. **安装antlr4**

打开eclipse, 从help -> Eclipse Marketplace -> 搜索 Antlr4 ->install安装即可



1. **将Javalib文件夹放入C盘**
2. **设置环境变量**

* 添加antlr-4.7.2-complete.jar （Javalib文件夹内的一个文件）到系统环境变量CLASSPATH中。

例如我的环境变量CLASSPATH值为：

C:\Program Files\Java\jdk1.8.0\_40\lib\; C:\Javalib\antlr-4.7-complete.jar;

* 添加Javalib到环境变量Path中

例如我的环境变量PATH值为：

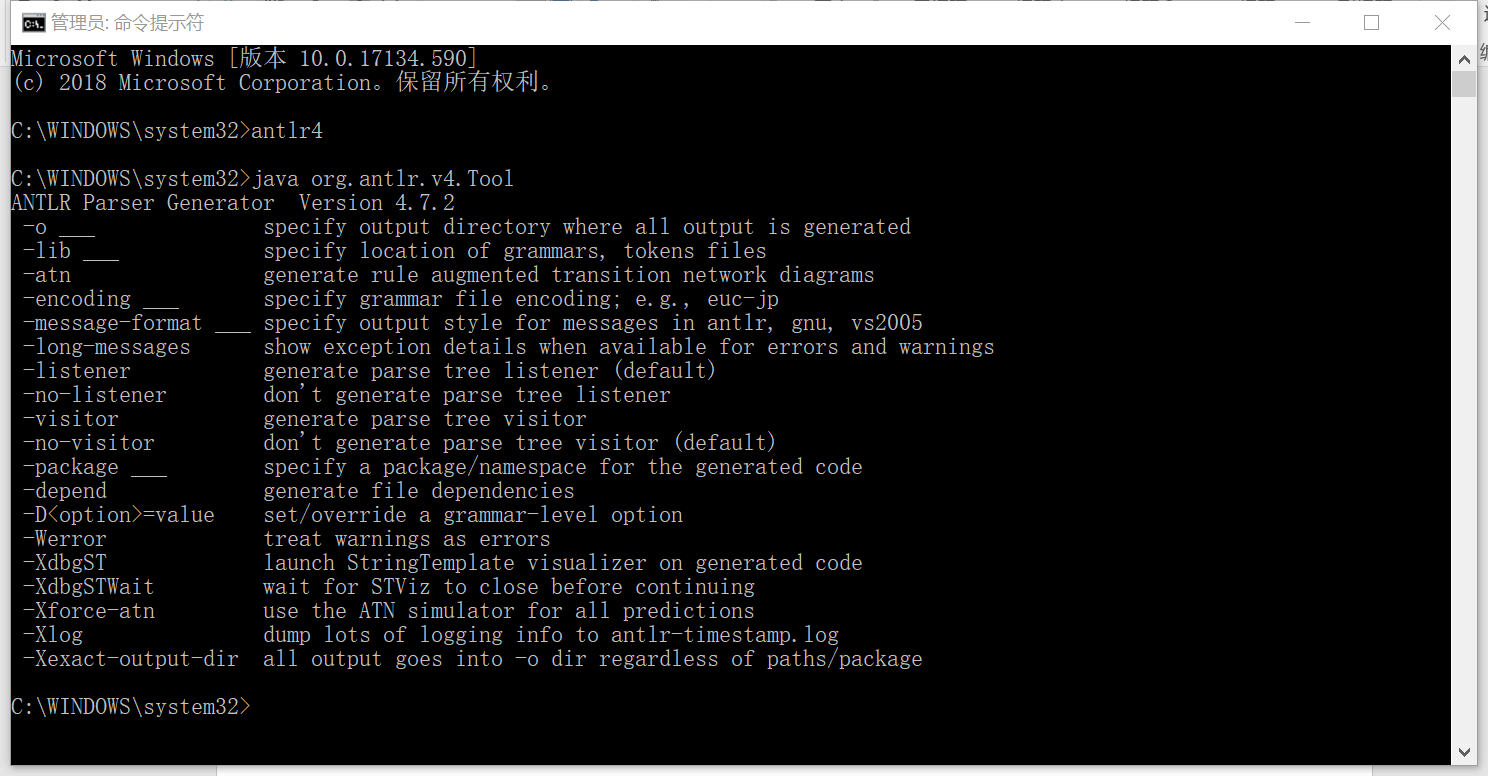
…; C:\Javalib

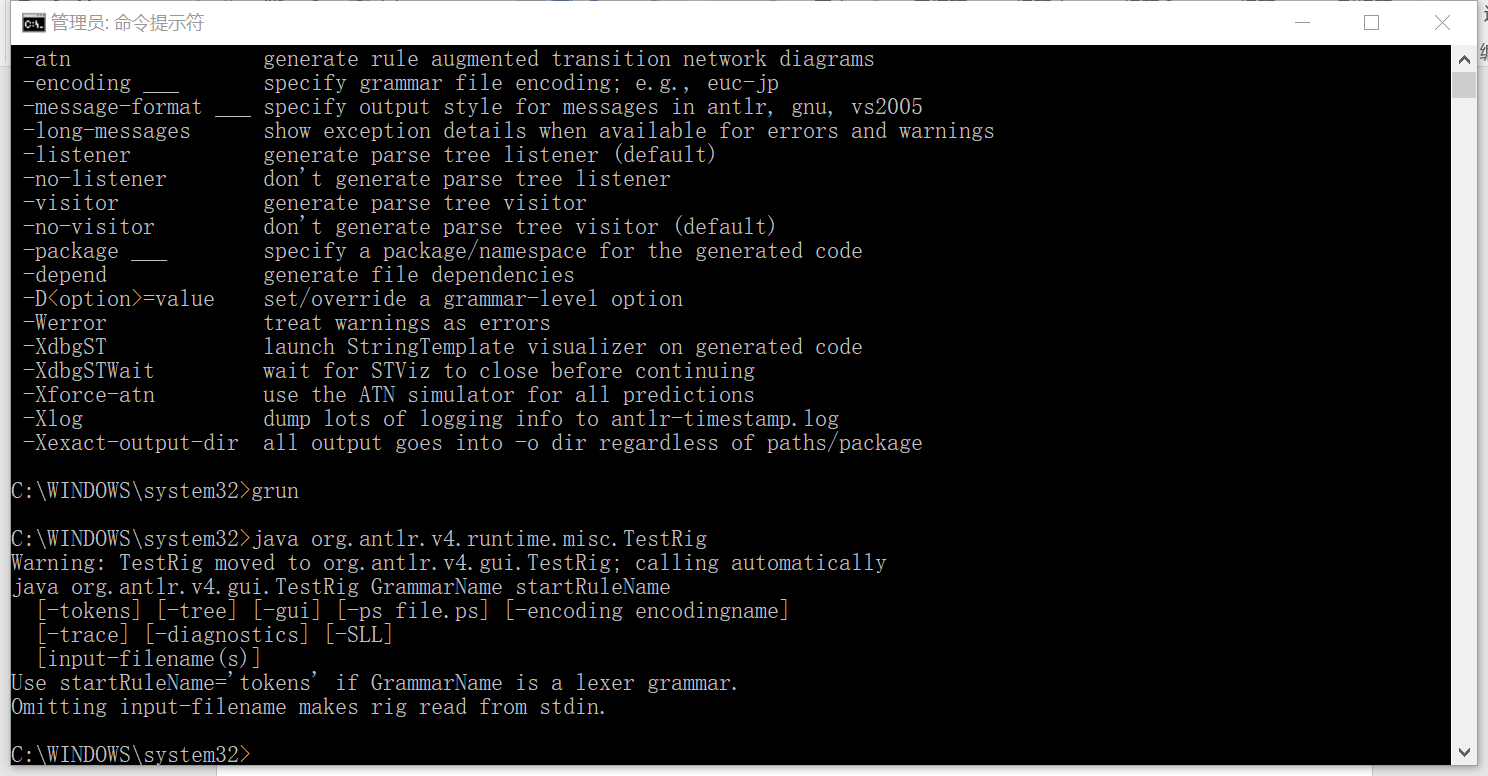
1. **cmd查看是否设置成功**

首先，打开一个CMD命令行。

然后分别输入antrl4 和 grun 命令；

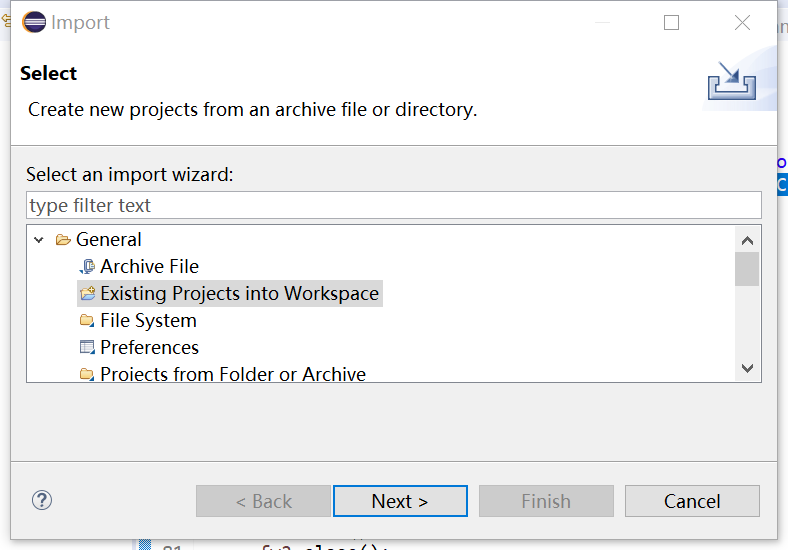
如果看到如下图所示内容，则表示成功安装。



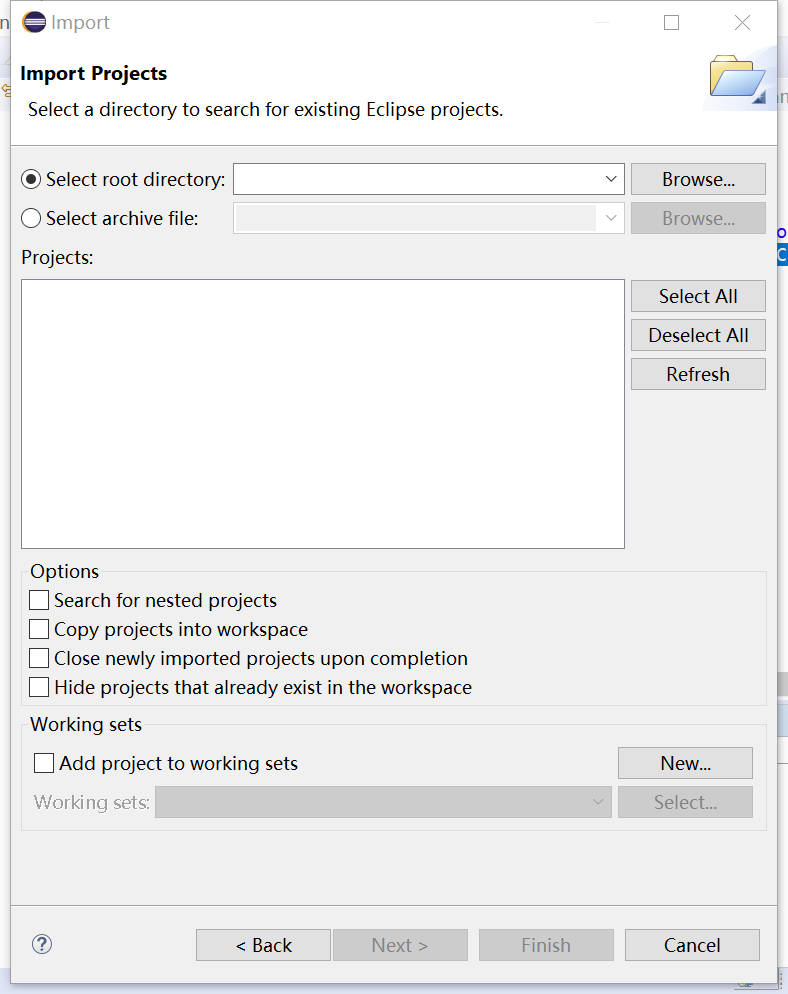


1. **将my\_java\_code.zip 解压缩，在eclipse中import一下my\_java\_code这个project**

点击File -> import



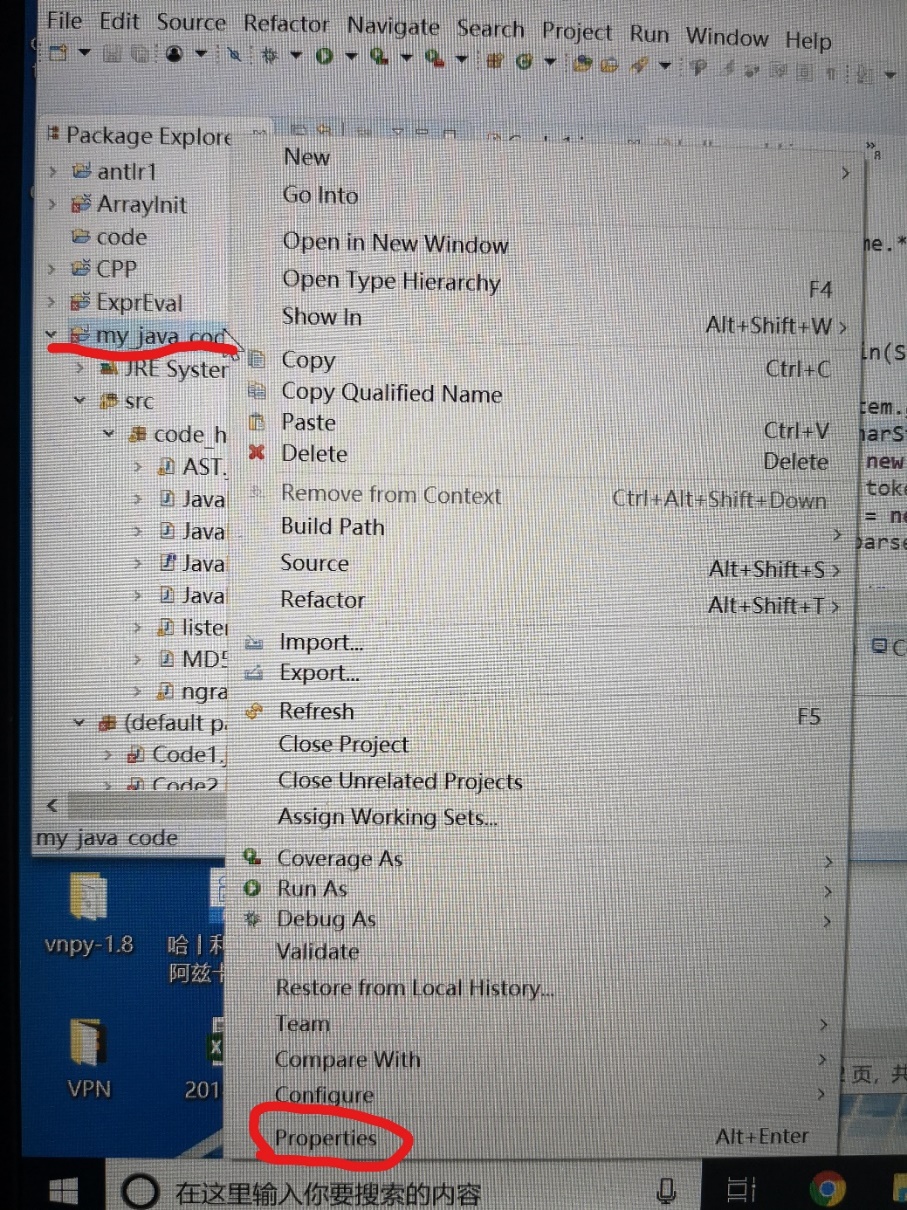
点击Existing Projects into Workspace



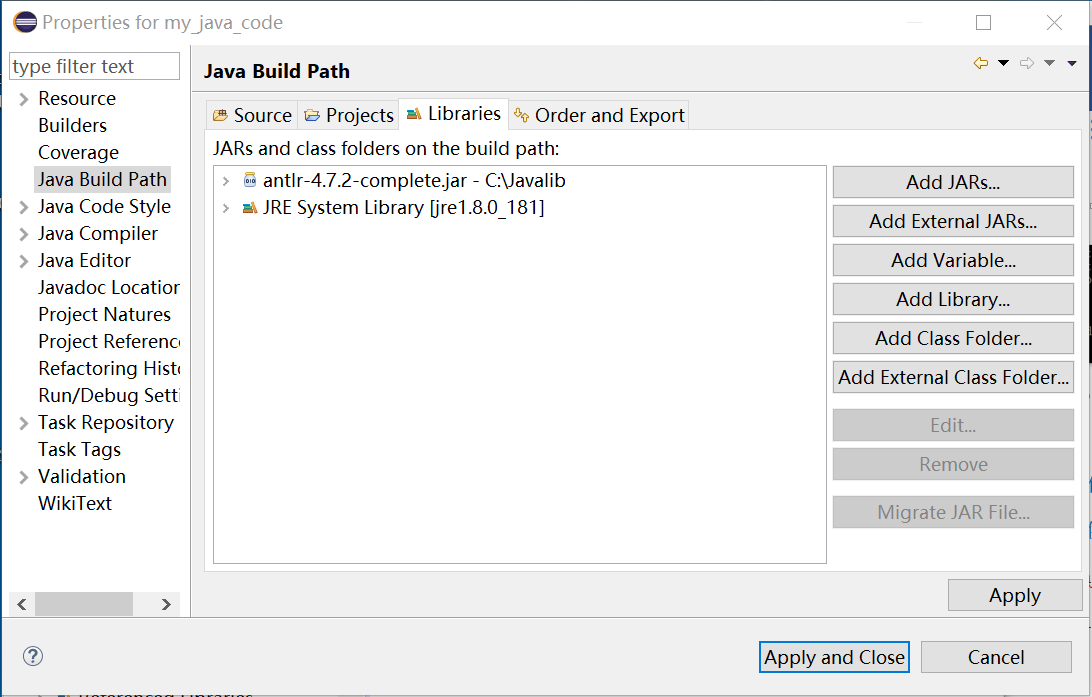
点击Browse…,找到my\_java\_code这个project，然后就可以import了

1. **添加.jar文件**

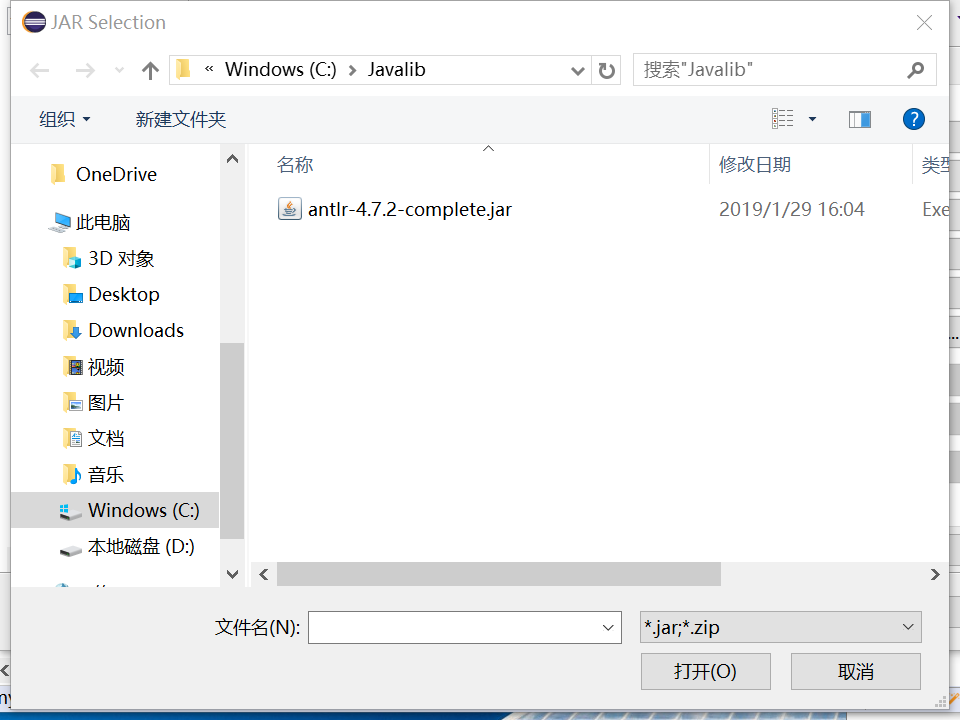
右键点击Package Explorer中my\_java\_code这个project，点击Properties



看见下面窗口，点击右方Add External JARs…



选择.jar文件，然后点击打开。



最后点击Apply and Close。

1. **Code处理**

这里我们分别对文件夹“测试代码Codes”中三个Code进行了一系列（最终成为hash set）处理，此处以Code1.java为例，其它两个同理。

打开project -> src -> code\_handle

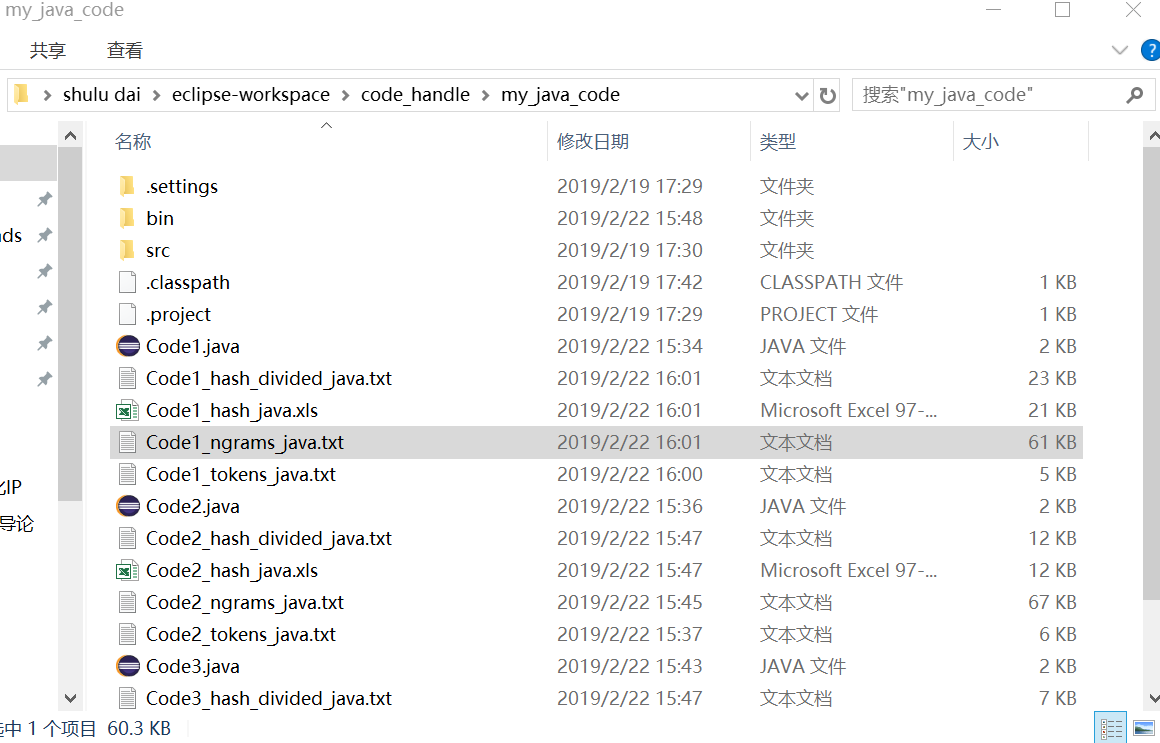
首先，运行AST.java，实现 Code1.java 到 AST 再遍历AST得到tokens的转变，结果存在Code1\_tokens\_java.txt中。

然后，运行ngrams.java，实现tokens到ngrams的转变，结果存在Code1\_ngrams\_java.txt中。

最后，运行MD5Hash.java，实现 ngrams 到 hash set的转变，结果存在Code1\_hash\_divided\_java.txt和 Code1\_hash\_java.xls中。

最终得到Code1的hash set.

最后运行得到的这些结果都在“…\eclipse-workspace\code\_handle\my\_java\_code”这个文件夹里。



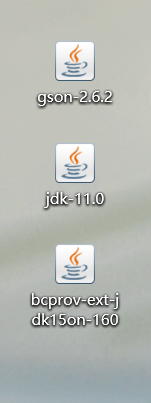
1. **其它**

以上仅以Code1为例，在尝试转化Code2和Code3的时候，请注意修改代码中读取和写入的文件的名称，即把每个“Code1”都换成“Code2”或“Code3”。

例：把“Code1\_tokens\_java.txt”改为“Code2\_tokens\_java.txt”或“Code3\_tokens\_java.txt”.

**后面这一部分我们需要完成代码的相似度计算，随后判定是否为原创文档，将通过判定的代码特征值上链…**

1. **User Library中放入gson-2.6.2与bcprov-ext-jdk15on-160**



下载gson与bcprov文件，在eclipse的Window中找到Preferences，在左边列表中找到Java🡪Build Path🡪User Libraries 新建bouncycastle\_lib、gson\_lib后选择Add External JARs并选中相对应文件，然后在Package Explore中选择test，右键Build Path🡪Add Library🡪User Library选中bouncycastle\_lib和gson\_lib后选择结束。

**10.运行Noobchain中的main函数得到区块链**

例子实现了上链与源代码文档原创性判断，代码2抄袭代码1所以判断后不进行上链操作，由于例子链上只存放了一个代码所以没有出现merkleroot，当同一个区块里存放多个代码特征值，代码还能实现merkleroot的功能。

（因codefingerprint数据量大，需要在线下txt文件中查看）

**11.得到的blockchain\_get.txt部分自动保存**在

“…\eclipse-workspace\code\_handle\my\_java\_code”目录下