# 实验二 掌握WEKA软件工具的搭建环境并学会使用

# 一、实验目的

1. 掌握Weka软件工具的搭建环境；
2. 学会使用Weka的基本操作；

# 二、实验步骤

## 1、了解Weka

Weka的全名是怀卡托智能分析环境（Waikato Environment for Knowledge Analysis），是一款免费的，非商业化的，基于JAVA环境下开源的机器学习（machine learning）以及数据挖掘（data mining）软件。它和它的源代码可在其官方网站下载。有趣的是，该软件的缩写WEKA也是新西兰独有的一种鸟名（新西兰秧鸡），而Weka的主要开发者同时恰好来自新西兰的怀卡托大学（The University of Waikato）。

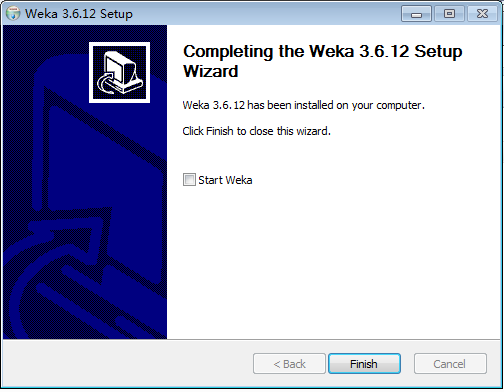
## 2、安装Weka

### （1）下载

官网下载安装包：<https://www.cs.waikato.ac.nz/ml/weka>

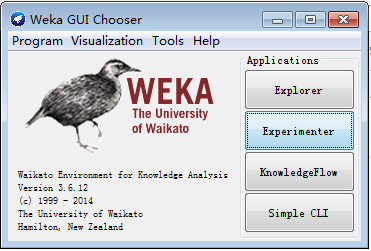
### （2）安装

打开安装包一直next，没有安装jre需要安装，安装好jre后，Weka安装完成。



**图 1 安装完成**

## 3、界面介绍



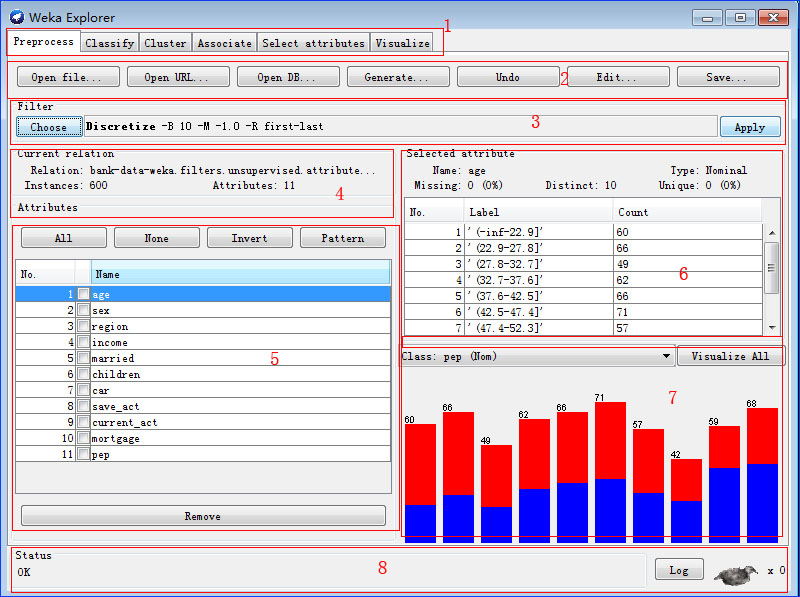
**图 2 Weka GUI界面**

（1）Explorer界面：实验主要用到的。

（2）Experimenter界面适用于基于不同数据集的不同机器学习方法的大规模性能比较。

（3）KnowldgeFlow界面是Weka的图形界面。

（4）命令行界面。



1. 区域1的几个选项卡是用来切换不同的挖掘任务面板。

具体功能如下：

Preprocess（预处理）：选择数据集，并以多种方式对其进行修改

Classify（分类）：训练用做分类或回归的学习方案，并对他们做出评估

Cluster（聚类）：学习数据集的聚类

Associate（关联）：学习数据的关联规则并对其评估

Select attributes(选择属性)：在数据集中选择最相关的部分

Visualize（可视化）：查看不同的二维数据点图并与其互动

（2）区域2是一些常用按钮。包括打开数据，保存及编辑功能。

（3）区域3中“Choose”某个“Filter”，可以实现筛选数据或者对数据进行某种变换。数据预处理主要就利用它来实现。

（4）区域4展示了数据集的一些基本情况。

（5）区域5中列出了数据集的所有属性。勾选一些属性并“Remove”就可以删除它们，删除后还可以利用区域2的“Undo”按钮找回。区域5上方的一排按钮是用来实现快速勾选的。

（6）在区域5中选中某个属性，则区域6中有关于这个属性的摘要。注意对于数值属性和分类属性，摘要的方式是不一样的。图中显示的是对数值属性“income”的摘要。

（7）区域7是区域5中选中属性的直方图。若数据集的最后一个属性（我们说过这是分类或回归任务的默认目标变量）是分类变量（这里的“pep”正好是），直方图中的每个长方形就会按照该变量的比例分成不同颜色的段。要想换个分段的依据，在区域7上方的下拉框中选个不同的分类属性就可以了。下拉框里选上“No Class”或者一个数值属性会变成黑白的直方图。

（8）区域8是状态栏，可以查看Log以判断是否有错。右边的weka鸟在动的话说明WEKA正在执行挖掘任务。右键点击状态栏还可以执行JAVA内存的垃圾回收。

# 三、实验总结

* 1. 通过本次实验，掌握了Weka的下载及安装方法；
  2. 了解了Weka的相关信息；