DMDW LAB – 1

AIM: Build Data Warehouse and Explore WEKA

THOERY:

Weka (pronounced to rhyme with Mecca) is a workbench that contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions. The original non-Java version of Weka was a Tcl/Tk front-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities in C, and Make file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains, but the more recent fully Java-based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research.

Advantages of Weka include:

* Free availability under the GNU General Public License.
* Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform
* A comprehensive collection of data preprocessing and modeling techniques
* Ease of use due to its graphical user interfaces

**Data Warehouse Introduction:**

Start with an introduction to data warehousing. Explain what a data warehouse is and its significance in business intelligence and decision support systems.

**Data Warehousing Process:**

Describe the key stages in the data warehousing process, including data extraction, transformation, loading (ETL), and data modeling.

**Data Sources Selection:**

Choose appropriate data sources for the data warehouse. Discuss the sources' relevance to the business problem or application under study.

**Data Preprocessing:**

Explain the importance of data preprocessing in data mining. This includes data cleaning, data transformation, and data reduction techniques.

**Data Integration:**

Discuss the challenges of integrating data from multiple sources into a unified data warehouse. Explain how data integration is performed and why it is crucial.

**Data Modeling:**

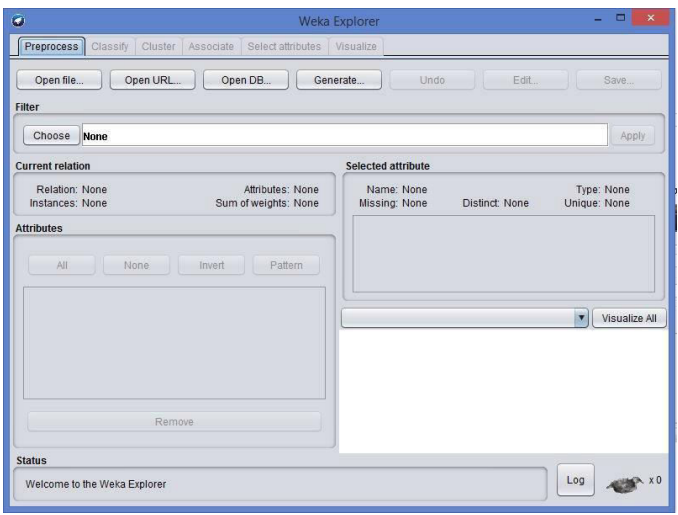
Describe the process of designing and creating a data model for the data warehouse. Discuss the difference between a star schema and a snowflake schema and when to use each.

OUTPUT:

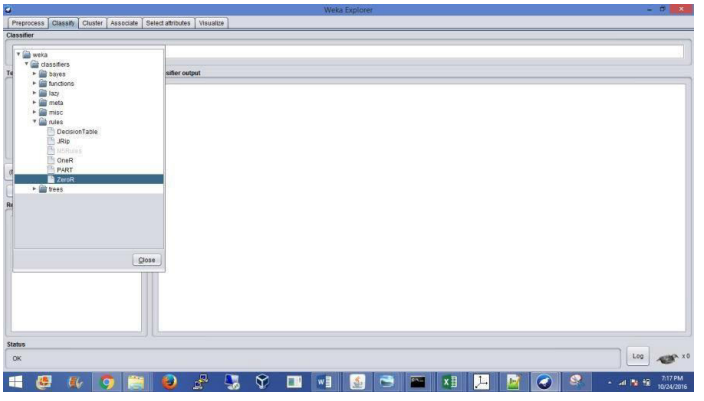
Exploring WEKA tool:



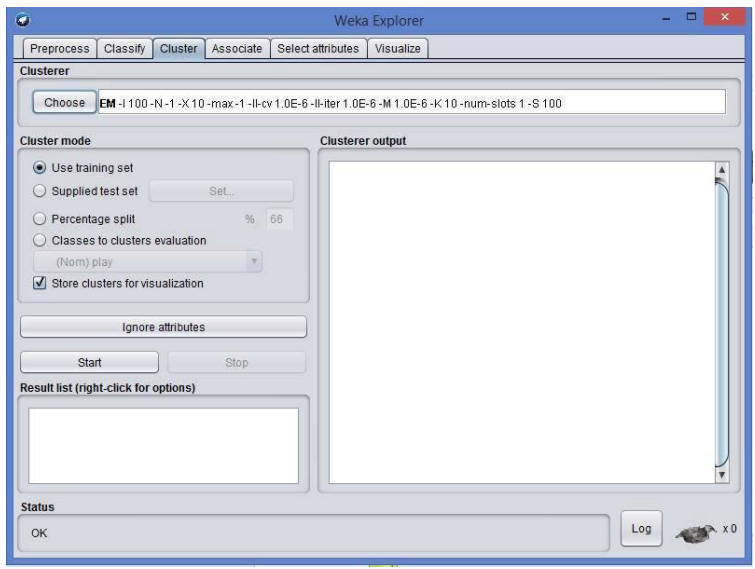
Weka GUI



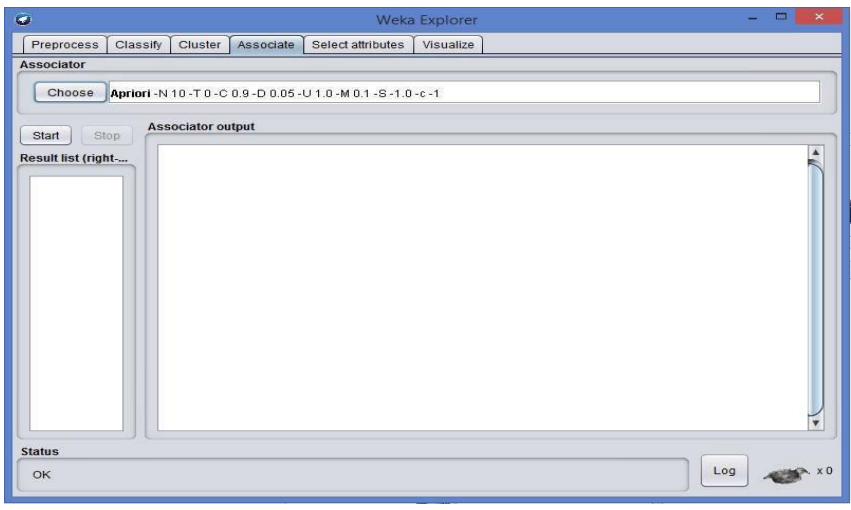
Pre-Processor



Classifier



Clustering



Association

CONCLUSION:

Successfully explored the WEKA tool.