

AMD (MLDM – Machine Learning and Data Mining) – Module of Practice

Paulo Trigo Silva

1. Extend the Data Model

- a) Copy the scripts' folder from the previous practical lesson into a new folder (to use in this lesson).
- b) Change the previous scripts with the ones (with the same name) provided in this lesson.
- c) Drop all data in your database and execute the new "02_script_POPULATE_SCHEMA.txt" file.
- d) Confirm that your the database is properly populated (e.g., use "pgAdmin 4").
- e) Open the "03_script_CREATE_VIEW.txt" file and complete the "create view" code in order to obtain the output as presented in the bottom of the file (cf., below the [OUT] tag).

2. Export Data from the Model

- a) Open the "30_script_EXPORT_DATA.txt" file.
- b) Explore the SQL "COPY" statement in order to export the "v1" view (cf., previous item) and generate a CSV (icomma-separated values) formatted data file just like "view_export.txt". For detailed information on the "COPY" statement see "a01_postgresql-10-A4.pdf" (page 1394). You may also want to explore the "\o" psql command. For detailed information about psql see "d02_psql_withPostgreSQL.pdf".
- c) Explore the psql "\COPY" command in order to export the "v1" view (cf., previous item) and also obtain the "view_export.txt" file (see "b01_psql_withPostgreSQL.pdf").
- d) Write, in the "30_script_EXPORT_DATA.txt", comments on both previous approaches (SQL COPY statement and psql \COPY command) that describe their main differences.

3. Format Data Before Exporting

- a) Open the "03_script_CREATE_VIEW.txt" file.
- b) Build a view, named "v1_domain" that projects a single line (tuple) with some constant information; execute and test this new view.
- c) Now build a "SELECT" statement that combines both the views "v1_domain" and "v1" so that the "v1_domain" line (tuple) always appears on the top of "v1" data. You may need to operate (minor) changes on "v1_domain" and "v1".
- d) Use the previous "SELECT" statement to build a new view, named "v1_dataset". Execute and test this new view.
- e) Populate the data model with additional tuples and explore the views' behavior.

4. Dataset (automatic) build of “3RowHeader” Orange Format

- a) Consider the "orange data mining" input format described in:
`http://docs.orange.biolab.si/3/visual-programming/loading-your-data/#three-row-header-format`
- b) Open the “03_script_CREATE_VIEW.txt” file.
- c) Implement a view, “v1_class” that projects a single line (tuple) that indicates the “class” attribute.
- d) Now build a “SELECT” statement that combines the views “v1_domain”, “v1_class” and “v1” so that the resulting data matches the "3-Row-Header" input format.
- e) Use the previous “SELECT” to (re)build the view “v1_dataset”. Execute and test this view.
- f) Make sure you reproduce data and format presented in “dataset_3RowHeader.txt” file.

5. Dataset (automatic) build of “HeaderWithType” Orange Format

- a) Consider the "orange data mining" input format described in:
`https://docs.orange.biolab.si/3/visual-programming/loading-your-data/#header-with-attribute-type-information`
- b) Open the “03_script_CREATE_VIEW.txt” file. Follow a similar process as in the previous exercise and build the view “v2_dataset” (and any additional views that might be needed) so that so that the resulting data matches the "Header-with-Attribute-Type" input format.
- c) Make sure you reproduce data and format presented in “dataset_HeaderWihType.txt” file.

6. The “Kick-Off” of Final Project A

- a) Consider the “final project A” specification (cf., moodle, “Final Project” folder).
- b) Read project “Goal”, “Scenario A” and “Project Items”.
- c) Read “Project Item: 1” and write down (in English or Portuguese) your own “domain description” (do not exceed ½ A4 page).
- d) Analyze your “domain description” with your “MedKnow” client.
- e) Develop “Project Item: 2”.
- f) Analyze your “conceptual data model” with your “MedKnow” client.
- g) Develop “Project Item: 3”.
- h) Develop “Project Item: 4”.
- i) Show your “MedKnow” client a demo of your operational database model and implementation.
- j) Develop “Project Item: 5”.