UniCom
User Guide
by Daniel Otal Rodríguez. 2016

UniCom (stands for Universal Comparator) is a software which gives the user the ability to compare objects of any domain. The program can read massive datasets provided by the user and use that information to compare multiple objects with a reference object chosen by the user. The results are presented through a radar chart and the percentage of similarity between both items, calculated by the shared area on the chart. In order for *UniCom* to operate needs a CSV file, where the dataset must be stored. In the following chapters *UniCom* functionalities and all necessary instructions for use are further explained and detailed.

This program software was developed in 2016 by Daniel Otal Rodríguez with the major support of professor Adam Wojciechowski from Poznan University of Technology.

I. What is a CSV file?

A bit of background knowledge is required in order to be ready to use *UniCom*. Comma-separated values (CSV) files are a widely used practice of storing data in plain text. It is particularly common for moving and converting data between spreadsheet programs which run on incompatible formats. There are plenty of different specifications, but the most popular one (RFC 4180) indicates that each line in a CSV file is a data record and each record consists of one or more fields separated by commas. Normally the format is best used to represent sequences of records in which each record has an identical list of fields, like a single relation in a relational database. *UniCom's* CSV is not much different from the standard one, but it also has its own rules.

II. Formatting the UniCom CSV File

The requirements for the *Unicom* CSV File are:

● There are three types of data records in a CSV file formatted for *UniCom*: labels, reference object and compare object. There must be at least one data record of every type and they must be stored in the correct order. The first line of the file encompasses the names of the labels for every data field. Second line corresponds to the reference object. The following lines of the file are reserved for the compare objects, considering that there is no limit of entries of comparable items. As a result, *UniCom's* CSV file should have at least three lines.

● Label type of data records are alphanumerical, while reference and compare objects data fields are only numerical—where the dot is used as the decimal mark.

● Reference and compare objects only accept positive values or the zero number.

● There must be the same number of data fields for each data record. In case there is no value for a reference or compare object data record, the value of the data field will be considered zero.

- Field delimiters are semicolons (;), instead of commas. They must be located only between data fields without exception.

- A line break marks the end of a data record.

- The file must have the extension '.csv' so that the program recognizes the file.

- Upper and lower boundaries for the compare objects set can be entered by the user for chart scaling (see fourth 'Steps to Scale the Chart' section of the user guide).

III. CSV Example

Year;Horse Power;Price;Weight
2004;200;12000;1200
2006;160;15000;900
2005;150;18000;950

File in plain text 'cars.csv'

The basic example above runs perfectly on UniCom if it is saved in a CSV file and loaded into the program. In the figure below we can see 'cars.csv' running on UniCom. This dataset has four data records with four data fields. First red colored line is the label data record, which gives a name for every column if we picture the file as a spreadsheet. The second green line is the reference object and the last two white lines are the compare objects. More entries can be added to the file if the user edits the file by hand with a text editor.
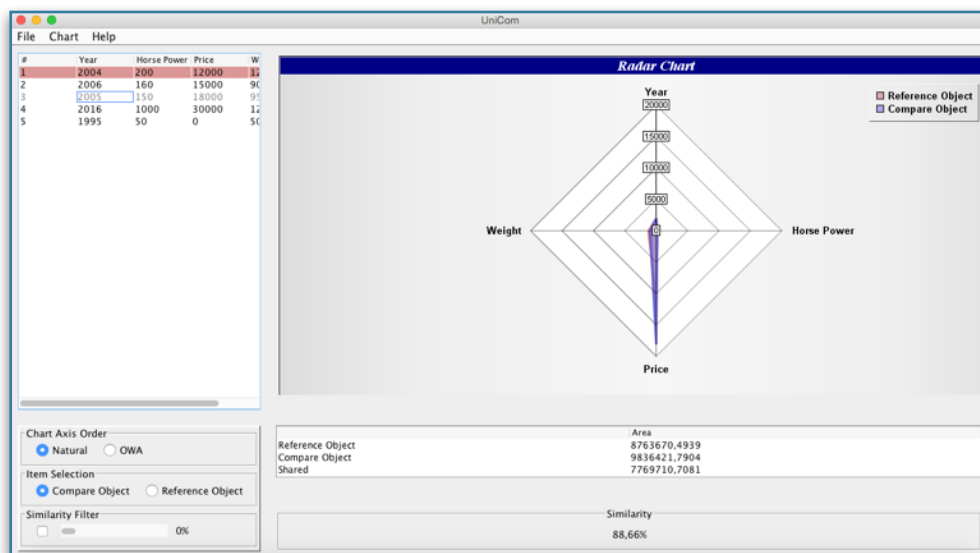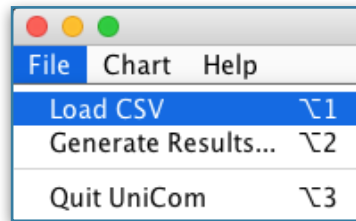

*File 'cars.csv' running on Unicom*

IV. Steps to Load a CSV File

There are two ways of loading a CSV file:

1. Go to 'File' and select 'Load CSV'. Use the file browser to select the CSV file. If you want to load another CSV file, the process can be performed again.

UniCom 'File' menu

2. Click on the 'Load CSV File' button which appears in the middle of the screen just after running the program. Use the file browser to select the CSV file. This option is only available for the first CSV loaded by the user. If you want to load another CSV file you should follow the previous file menu instructions (section 1).
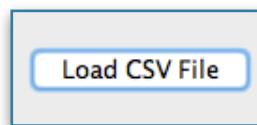
'*Load CSV File*' button

V. Steps to Scale the Chart

If the user wants to scale the dataset to normalize the data and to improve the visualization of the chart, *UniCom* has the feature of scaling the graph in the range of [0, 100]. To accomplish this with maximum precision, the user should add two fake compare objects to the CSV: the maximum possible value for each variable of the compare object and the second one for its minimum possible value. With both limits stablished, chart scaling can be done with greater accuracy. Otherwise, if both limits are not set by the user, *UniCom* takes the maximum and minimum values found in the dataset.

Following the 'cars.csv' example shown in the past section, the user should add two more lines to the CSV to scale the chart properly:

File 'cars.csv' with maximum and minimum boundaries

Last two yellow lines are the upper and lower boundaries respectively. They can be located anywhere between the compare objects set as they will be treated by UniCom as an ordinary compare object. If the user does not want to set the limit for one of the values specifically, he can always enter an average number for that compare object data field as *UniCom* will ignore it.

To visualize the scaled chart, the user can find this option under the 'Chart' menu after loading the CSV into the program. Clicking on the checkbox menu option 'Scaled Chart' switches the graph

and similarity results to the scaled ones. The user can go back to the unscaled chart simply by clicking again on the same menu option.

VI. Area and Similarity Results

One of the similarity results for an entry from the 'cars.csv' dataset

Due to the graph changing its dimensions when the values are scaled, every compared object has two sets of data results: the unscaled and the scaled one. Both sets consist of the same variables shown in the table above: the reference and compare objects chart areas, the shared chart area between these two items and the resulting similarity percentage.

It is in the hands of the user to decide which of the data results suits him better, as every dataset is different and needs different considerations. Given the diversity of data which can be found in a dataset, the option to scale its values also serves as a manner to improve graph's readability and to normalize all data under the same scale.

VII. Generate Results…

Underneath the 'File' menu there is the option 'Generate Results…'. It generates a CSV file (following *UniCom* format) with the dataset which is loaded in the program and the currently selected reference object, but adding new data fields for the calculated areas in the chart and the similarity results of every entry. This file may help to process the data with external programs.


```
Year;Horse Power;Price;Weight;Area;Common Area;Common Percentage;Scaled Area;Scaled
Common Area;Scaled Common Percentage
2004.0;200.0;12000.0;1200.0;8763670,4939;8763670,4939;100;4288,49534288,4953;100

2004.0;200.0;12000.0;1200.0;8763670,4939;8763670,4939;100,00;4288,4953;4288,4953;100,00
2006.0;160.0;15000.0;900.0;8057752,8484;7335101,6359;83,70;3144,9926;2920,3718;68,10
2005.0;150.0;18000.0;950.0;9836421,7904;7769710,7081;88,66;3598,8727;3169,0899;73,90
2016.0;1000.0;30000.0;1200.0;31484416,9000;8763670,4939;100,00;17879,9333;4288,4953;100,00
1995.0;50.0;0.0;500.0;490468,9196;490468,9196;5,60;0,0000;0,0000;0,00
```

File 'cars_results.csv' generated by the 'Generate Results…' feature

Taking the example of 'cars_results.csv', we can see how new content has been added to the original CSV which is colored in white. Now, there are new labels at the first and second line (with no break line) of the file: 'Area', 'Common Area', 'Common Percentage', 'Scaled Area', 'Scaled Common Area' and 'Scaled Common Percentage'. All of them represent the column titles for the similarity results which follow. The green line is the current reference object selected on *UniCom*,

which depends on the entry the user chose as the reference to compare the rest of the dataset (second line of original CSV file by default). In this example, last six data fields stand for the similarity results of the reference object and as is to be expected, similarities are always 100% since the object is being compared to itself. The rest of the CSV lines are all the dataset of comparable objects: yellow numbers are the values for the calculated chart areas, while orange numbers are the similarity results between the compare object and the reference object given in percentage.

VIII.Panel Control Options

1. **Natural VS. OWA Chart Axes Order.** In the left control panel of the main UniCom screen, under the table that lists the comparable items, there are two radio buttons to control the order of the chart axes. These buttons are called 'Natural' and 'OWA' (which stands for Ordered Weighted Averaging). The 'Natural' mode distributes the axes in order of appearance in the CSV—that is, if we take a look at the 'cars.csv' example previously explained in section 3, axes will be ordered clockwise as: 'Year', 'Horse Power', 'Price' and 'Weight'. On the other hand, the 'OWA' mode order algorithm takes the selected reference object, sorts its data from highest to lowest and applies this order to the chart axes—in other words, 'cars.csv' after applying OWA ordering method would look in clockwise direction like: 'Price', 'Year', 'Weight' and finally 'Horse Power'.
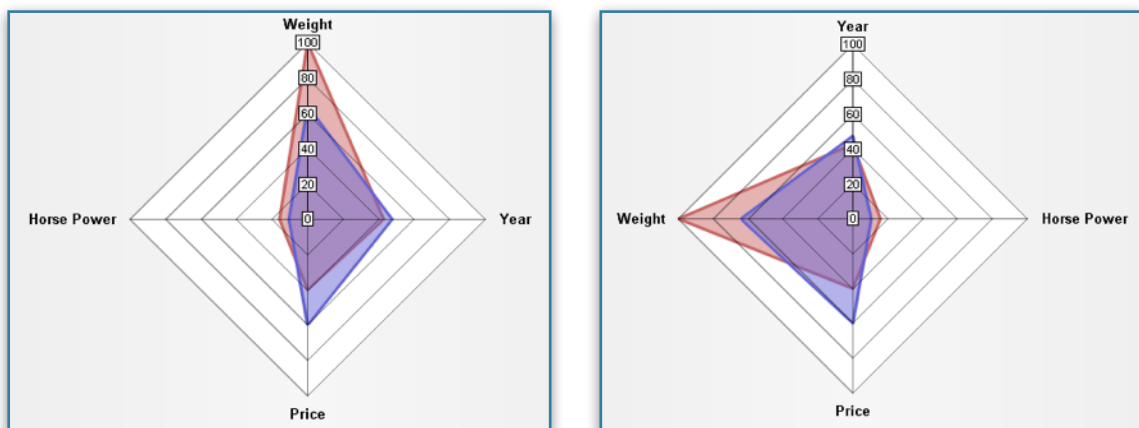
Charts with 'Natural' order (right) and with 'OWA' order (left)

Similarity results do not change whether or not the chart axes are ordered in respective ways.

2. **Item Selection.** This feature is essential to navigate the list of compare objects and be able to select different items on the table. To select a new reference object to compare all items, the user should select the 'Reference Object' item selection mode and select the desired entry on the table. This effectuates all data calculations so the user can retrieve this new information taking a look at the charts and the similarity results, or maybe generating the results in bulk with the 'Generate Results…' feature (see section 7). If the user wants to compare another entry instead, the 'Compare Object' mode must be selected. Selecting a new entry to compare it with the reference object is as easy as clicking on the item the user wants to compare.

3. **Similarity Filter.** Rather than a filter, this panel control option overshadows the object rows with a lower similarity percentage than the selected one. The checkbox located on the left side of the scrollbar enables or disables this functionality.