AHMat

Spectra Identification

Using Quasar 1.11.1

Procedure Manual

Contents

1. Introduction

1.1. Purpose of the Manual

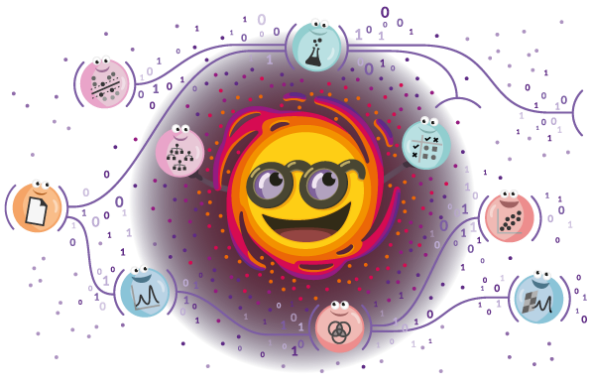
This manual aims to provide a guide to the **installation** and **use** of a custom spectra identification widget within the **Quasar** spectroscopy software.

With this widget, one is able to identify the **closest neighbour** to one spectrum, among the spectra of a specific database.

2. Quasar

2.1. Overview

Quasar is a **widget-based** modular software platform designed for the processing and analysis of **any type of data**, but specially for **spectroscopic** data. It is commonly used in a variety of branches such as materials science, chemistry, biology, and environmental analysis.



It was developed to support a **range of analytical methods**, including Raman, infrared (IR), ultraviolet-visible (UV-Vis), and X-ray spectroscopy. Quasar enables to perform data pre-processing, multivariate analysis, visualization, and classification within a single framework and a friendly interface.

Quasar supports data import from a **wide range of file formats** from the multiple experimental devices. It also offers scripting capabilities for advanced processing and batch operations. It is compatible with multiple operating systems, including Windows, macOS, and Linux. You can read the documentation here.

2.2. Installation

You can install Quasar downloading the installer in the “download” webpage of Quasar, that you can access clicking here.

Executing the installer will install all necessary Python packages as well as the necessary code to run the main program. The default route for the Quasar files will be:

Windows

- **C:\Users\YourUser\AppData\Local\Programs\Quasar**

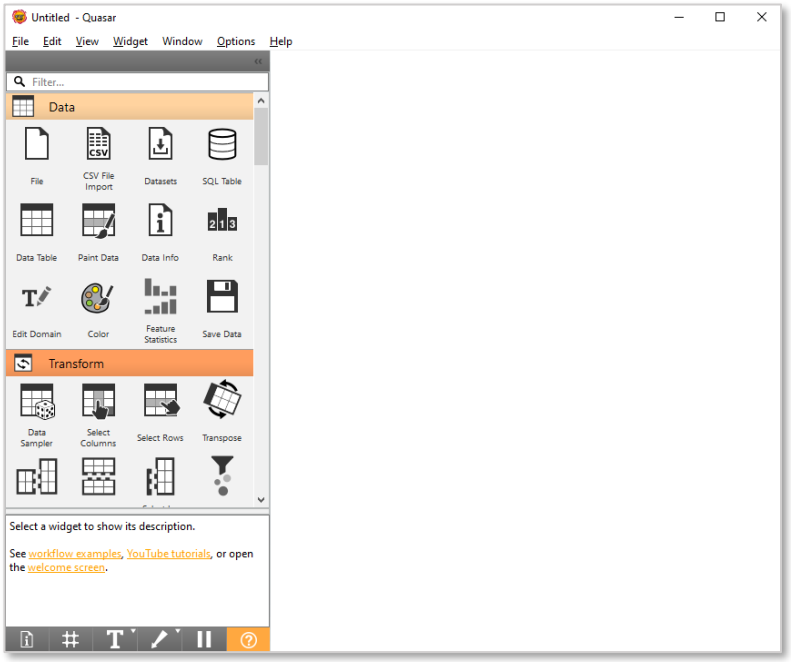
Linux

- **/home/youruser/.local/lib/python3.X/site-packages/Quasar**

macOS

- **/Users/youruser/Library/Python/3.X/lib/python/site-packages/Quasar**

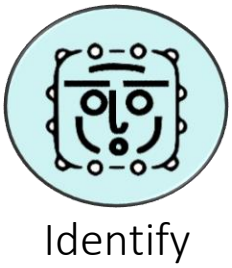
Upon successful installation, launching Quasar should present the main interface with access to the widget catalogue, as shown in the following image:



Once the user has proved that the installation was successful, they can proceed with the installation of the “Identify” widget.

3. The “Identify” Widget

3.1. Overview



The “Identify” widget is a custom widget developed into the Quasar platform, that provides automated spectra identification through comparison with a known reference database.

This widget is particularly suitable for applications such as spectral classification, chemical identification, or matching experimental spectra to reference libraries.

It uses a well-known neighbour finding algorithm that computes the distance between the unknown spectrum and each of the spectra of the database, and outputs the closest ones. The user is allowed to choose from multiple distance metrics.

Inputs

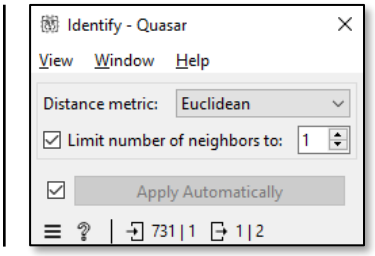
- **Database** (*Table*): A dataset containing multiple reference spectra. Each row represents one known spectrum, and each column corresponds to a spectral feature (e.g., energy or wavelength).
- **Unknown spectrum** (*Table*): A dataset containing a single spectrum to be identified. The feature domain may differ from the database; if so, the unknown spectrum will be interpolated onto the database's feature axis.

Outputs

- **Identification** (*Table*): A subset of spectra from the database identified as most similar to the unknown spectrum. The result includes a new column indicating the computed distance between the unknown and each selected spectrum.
- **Comparison** (*Table*): Same as the Identification output but with an additional spectrum corresponding to the instance of the unknown spectrum. The results include an additional column indicating the origin of the spectrum, whether "unknown "or "neighbour".

Parameters

- **Distance metric**: Select from a range of distance metrics (e.g., Euclidean, Cosine, Pearson). The choice affects how spectral similarity is quantified.
- **Number of neighbours**: Defines how many of the closest matching spectra to include in the output.



When the database and the unknown spectrum have different feature sets, the unknown spectrum is interpolated onto the database's feature axis using linear interpolation. The widget then computes distances between the preprocessed unknown and each spectrum in the database using the selected metric. The closest matches are output, optionally sorted by similarity.

3.2. Installation

To install the widget, first you have to download the necessary files. You can download them in this link. In the downloaded zip file, you will find:

- A python file with the extension '.py' containing the code of the widget. You must place this file in:

Linux
- **/home/youruser/.local/lib/python3.X/site-packages/Quasar/widgets/evaluate**

Windows
- **C:\Users\youruser\AppData\Local\Programs\Quasar\Lib\site-packages\Orange\widgets\evaluate**

macOS
- **/Users/youruser/Library/Python/3.X/lib/python/site-packages/Quasar/widgets/evaluate**

- Two image files with extensions '.svg', and '.png'. You must place those two in:

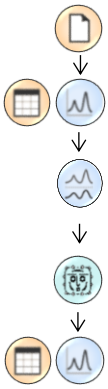
Windows
- **C:\Users\youruser\AppData\Local\Programs\Quasar\Lib\site-packages\Orange\widgets\evaluate\icons**

Linux
- **/home/youruser/.local/lib/python3.X/site-packages/Quasar/widgets/evaluate/icons**

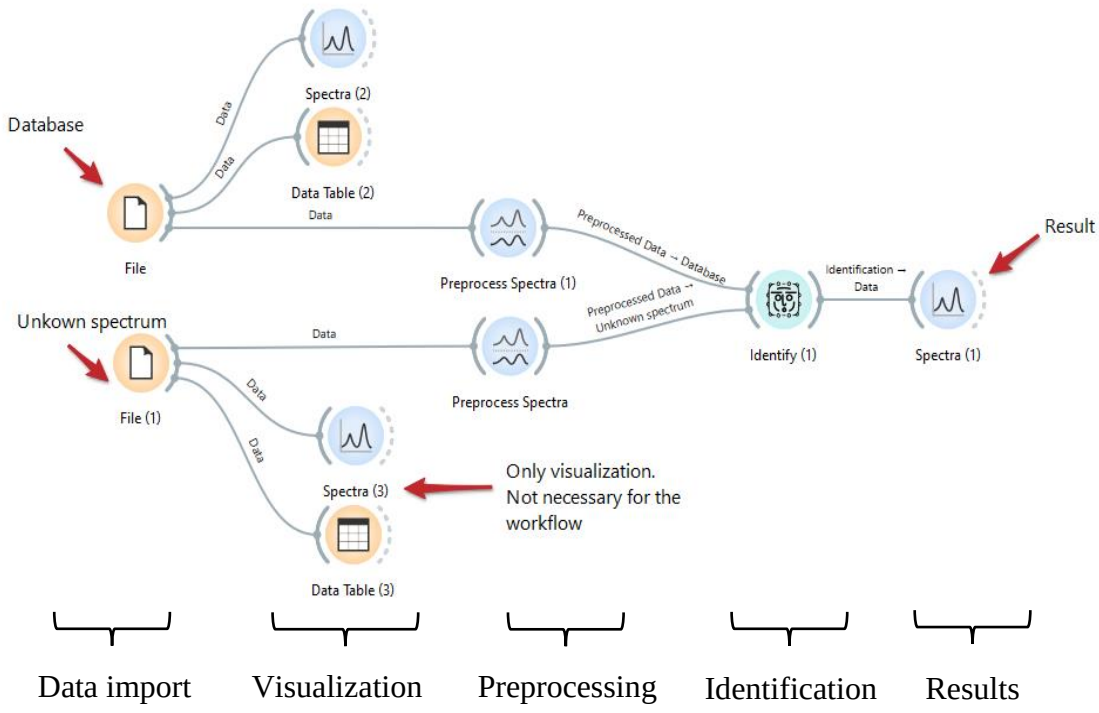
macOS
- **/Users/youruser/Library/Python/3.X/lib/python/site-packages/Quasar/widgets/evaluate/icons**

3.3. Example workflow

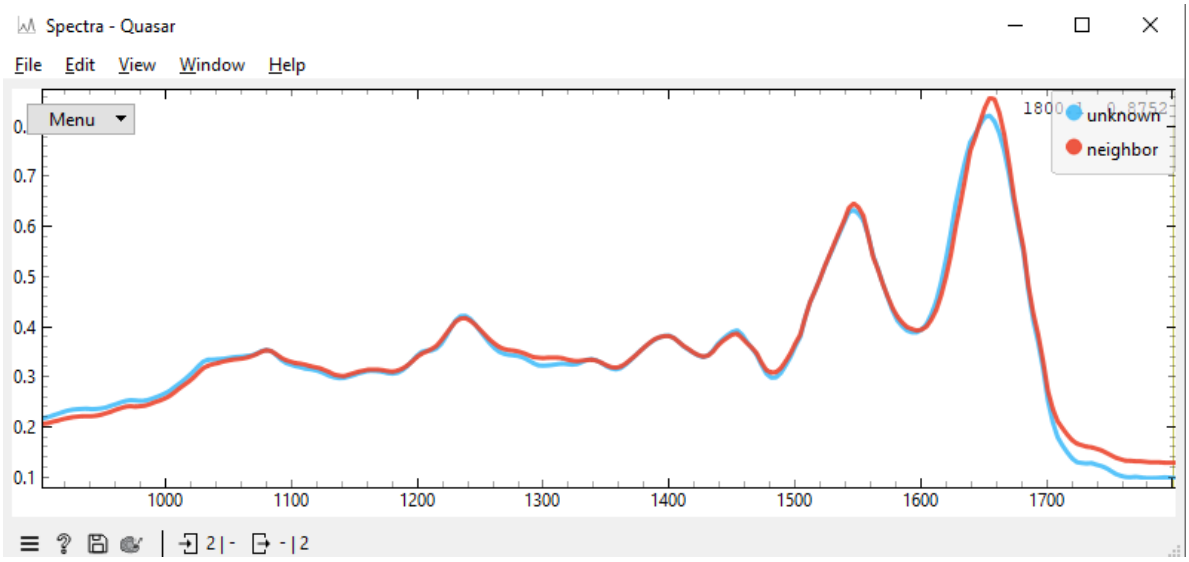
A typical workflow using the "Identify" widget in Quasar involves the following steps:



1. **Load Spectral Data**: Import a dataset (e.g., Raman spectra) using the **File → Open** menu or drag-and-drop functionality.
2. **Visualize data**: Use the widgets "Data Table" or "Spectra" to visualize your unknown spectrum and, if needed, the database.
3. **Preprocess the Spectrum & database**: Apply baseline correction, smoothing, or normalization as needed using Quasar's widget "Preprocess spectra". **Remember that you will also have to apply those corrections to the database.**
4. **Use the Identify Widget**: Take the 'Identify' widget from the Evaluate panel or searching 'Identify' in the widget search bar.
5. **Visualize Results**: You can use a "Data Table" or a "Spectra" widget again to visualize the results of your identification. If you want to see them in comparison with the original, choose the output "Comparison" double-clicking in the interconnection between the 'Identify' widget and the result spectra widget.



The result of the identification can be plotted using the "Spectra" widget that will display the closets neighbours. If needed, the output "Comparison" will also include the input unknown spectrum to ease the visualization. A normal plotted output can be seen in the following image.



4. Databases

4.1. Where to find them?

You will find a handful of spectra databases from many different experimental techniques in this link. These include Raman, Infrared, and even XRF.

You can also find your own databases like the RRUFF Project for Raman spectroscopy or the Aldrich FT-IR Spectra Library for Infrared searching in external websites. Databases found in external websites must be converted or formatted according to the Quasar's file opening requirements. To modify the database structure, Excel is a useful tool, as explained in the following section.

4.2. Create your own database

In order to create your own database your need the spectra of several identified samples. Each spectrum must contain one X-column and one Y-column.

Each of those spectra must have the exact same X-column, and will have to be arranged in any specific shape that Quasar can read using the "File" widget. One of the most comprehensive methods to arrange the data is utilizing Excel. The process to follow is as described here:

- Labels: They are the names of the identified samples. The labels are placed in the first column beyond the first row.

- X-axis: Either wavenumbers, Raman shift, energy, etc. They represent the x-axis of the spectra. Because all spectra must have the same X-axis, this one in unique and is placed in the first row beyond the first column.

- Y-axis: Either intensities, absorbance, transmittance, etc. They represent the y-axis of the spectra. This is unique for every label or spectrum. Therefore, every label will have its spectra associated. The Y-axis of each spectrum is placed to the right of its label matching each point of the X-axis.