WEB Scraping with Jrnime.

Introduction:-

web scraping is the most important thing in data science. Our objective in this article is to make a simple workflow in Jrnime to see how Jrnime scrapes the data from the data form the website. Here we have to take website of corona cases and other information for India. You can find the website here.

(i) webpage Retriever node:-

webpage Retriever node provides the facility to connect with the website and generate the xml file, so as you can see in the below dialog box of web retriever node you have to just specify your respective URL in the connection setting in our case we have to specify (www.businessinsider.com) this web site in URL.

You can find a more detailed description of the webpage retriever node here.

(ii) x-path:-

So, our next step is to parse the xml file to extract the table data from it. x-path node helps you out to parse the xml file, as you can see below the configuration to x-path we have to specify the x-path.

using the add path button you can add a different path and x-path summary displays how many paths you have.

(iii) ungroup.

our next step is to transform the table into a data file we use pivoting node. and after this, we have to ungroup the column using ungroup node.

(iv) Excel (or) CSV writer:-

our final step is to write a Excel (or) CSV writer file to this fetched data and knime provides facilities to do this by simply using Excel or CSV writer node.

So, By executing this full workflow at the end in the CSV or Excel you get the fetch data of data-wise billionair list.

correlation

* If you to test wether the correlation coefficient between two variable is significantly different from 0 you have to conduct correlation.

* Go to the top bar -> click Regression -> correlation under classical section -> choose at least two variables.

* Let's run the correlation to investigate whether the correlation coefficient between the covert and overt antisocial behaviour (covert and overt) is significantly different from 0.

- from now on, we will use the alpha level of .05 because it is most commonly used.

- to do so. move the two variables, overt and convert. under the variables section.

- the correlation coefficient between convert and overt is 0.386 and sig. value is lower than 0.001 according to the output in the output panel.

- therefore. the correlation coefficient of 0.386 is significantly different from 0.