ASSIGNMENT 2

- Explian three-dimensional data indexing ?
- Synonyms
- Construction of 3D scenes
-
- Definition
- Semantic queries on indexed objects allow the reuse of the 3D scenes.
-
- Introduction
- Nowadays, the 3D is a highly expanding media. More particularly with the emergence of dedicated standards such as VRML and X3D, 3D animations are widely used on the Web. The continuous evolution of computing capabilities of desktop computers is also a factor that facilitates the large deployment of 3D information contents. At the same time the demand in term of 3D information is becoming more and more sustained in various domains such as spatial planning, risks management, telecommunications, transports, defense, and tourism. 3D information should represent a real world scene as accurately as possible and should exhibit properties (like, topological relations) to allow complex spatial analysis [1].
-
- The construction of a 3D scene is a complex and time consuming task. Thus, being able to reuse the 3D scenes is a very important issue for the multimedia…

.•how do you use pandas to make a dataframe Out of n_dimensional arrays ?

- Here’s a very simple example to convert an array to a dataframe:
- import pandas as pd # Create the dataframe df = pd.DataFrame(numpy_array)
- df = pd.DataFrame(numpy_array, columns=[‘digits’, ‘words’]) …
- df = pd.DataFrame(numpy_array, index=[‘day1’, ‘day2’, ‘day3’, ‘day4’], columns=[‘digits’, ‘words’])

- What role does pandas play in data cleaning ?
- Data cleaning is the most important task that should be done as a data science professional. Having wrong or bad quality data can be detrimental to processes and analysis. Having clean data will ultimately increase overall productivity and permit the very best quality information in your decision-making. Following are some reasons why data cleaning is essential:
-
- data cleaning
- Image source: by me
-
- 1. Error-Free Data: When multiple sources of data are combined there may be chances of so much error. Through Data Cleaning, errors can be removed from data. Having clean data which is free from wrong and garbage values can help in performing analysis faster as well as efficiently. By doing this task our considerable amount of time is saved. If we use data containing garbage values, the results won't be accurate. When we don't use accurate data, surely we will make mistakes. Monitoring errors and good reporting helps to find where errors are coming from, and also makes it easier to fix incorrect or corrupt data for future applications.
-
- 2. Data Quality: The quality of the data is the degree to which it follows the rules of particular requirements. For example, if we have imported phone numbers data of different customers, and in some places, we have added email addresses of customers in the data. But because our needs were straightforward for phone numbers, then the email addresses would be invalid data. Here some pieces of data follow a specific format. Some types of numbers have to be in a specific range. Some data cells might require a selected quite data like numeric, Boolean, etc. In every scenario, there are some mandatory constraints our data should follow. Certain conditions affect multiple fields of data in a particular form. Particular types of data have unique restrictions. If the data isn't in the required format, it would always be invalid. Data cleaning will help us simplify this process and avoid useless data values.
-
- 3. Accurate and Efficient: Ensuring the data is close to the correct values. We know that most of the data in a dataset are valid, and we should focus on establishing its accuracy. Even if the data is authentic and correct, it doesn't mean the data is accurate. Determining accuracy helps to figure out the data entered is accurate or not. For example, the address of a customer is stored in the specified format, maybe it doesn't need to be in the right one. The email has an additional character or value that makes it incorrect or invalid. Another example is the phone number of a customer. This means that we have to rely on data sources, to cross-check the data to figure out if it's accurate or not. Depending on the kind of data we are using, we might be able to find various resources that could help us in this regard for cleaning.
-
- 4. Complete Data: Completeness is the degree to which we should know all the required values. Completeness is a little more challenging to achieve than accuracy or quality. Because it's nearly impossible to have all the info we need. Only known facts can be entered. We can try to complete data by redoing the data gathering activities like approaching the clients again, re-interviewing people, etc. For example, we might need to enter every customer's contact information. But a number of them might not have email addresses. In this case, we have to leave those columns empty. If we have a system that requires us to fill all columns, we can try to enter missing or unknown there. But entering such values does not mean that the data is complete. It would be still being referred to as incomplete.
-
- 5. Maintains Data Consistency: To ensure the data is consistent within the same dataset or across multiple datasets, we can measure consistency by comparing two similar systems. We can also check the data values within the same dataset to see if they are consistent or not. Consistency can be relational. For example, a customer's age might be 25, which is a valid value and also accurate, but it is also stated as a senior citizen in the same system. In such cases, we have to cross-check the data, similar to measuring accuracy, and see which value is true. Is the client a 25-year old? Or the client is a senior citizen? Only one of these values can be true. There are multiple ways to for your data consistent.
-
- By checking in different systems.
- By checking the source.
- By checking the latest data

- Explain the notion of pandas plotin?
- Pandas uses the plot() method to create diagrams. We can use Pyplot, a submodule of the Matplotlib library to visualize the diagram on the screen. Read more about Matplotlib in our Matplotlib Tutorial
-
- Here are the steps to plot a scatter diagram using Pandas.
- Step 1: Prepare the data. To start, prepare the data for your scatter diagram. …
- Step 2: Create the DataFrame. Once you have your data ready, you can proceed to create the DataFrame in Python. …
- Step 3: Plot the DataFrame using Pandas

- What is the difference between the series and data frame?
- Answer: Series is a type of list which can take integer values, string values, double value and more. Series can only contain single list with index, whereas dataframes can be made of more than one series or we can say that a dataframes is a collection of series that can be used to analyse the data