Aim:

To use mongoDB to process semi structured and unstructured data collection using python mongoDB interface.

Theory:

PyMongo:

The PyMongo distribution contains tools for interacting with MongoDB database from Python. The bson package is an implementation of the BSON format for Python. The pymongo package is a native Python driver for MongoDB. The gridfs package is a gridfs implementation on top of pymongo

We need to make sure that the PyMongo distribution installed. If so, in the Python shell, the following should run without raising an exception:

>>>import pymongo

Semi structured data:

It is a form of data that refers to data that does not contain any format structure of data models associated with RDBMS or other forms of data tables , but none the less contain other markers to separate other semantic elements and enforce hierarchy of records and fields within the data.

Semi structured data is increasing since the advent of the internet where full text documents and DBMS are not the only forms of data anymore and different applications need a medium for exchanging information.

Unstructured data:

It refers to information that either does not have a predefined data model or is not organised in a predefined manner. It is typically text but may contain multimedia .Results generated by such data are irregular and it is difficult to understand using traditional computer programs as compared to data stored in fielded form

Update:

Pymongo can update documents in a number of different ways inserting one by this way.

Import pymongo
conn=pymongo.connection()
db=conn.tutorial
db.things.insert(<key>:<value>)

db.things.findone(<key>:<value>)

Now,we can use the update method to modify document.
db.things.update(key:value)

Removing data

Removing documents,collections and database is easy in mongoDB.This first example removes a specific document.
db.things.remove('name':'KIck')
We can use db.object.dropCollection() method to drop the collection
db.dropCollection('things')

Conclusion

Hence,we have learnt mongoDB-python connectivity using pymongo and performed insert,delete,update and remove operations on mongoDB database.