Lecture Notes-Bob

Today my lecture is not about smart material or robotics that I am majoring in.The following project is all of my personal interest in other fields. So if Mrs Zhao think it’s not OK, I will choose another project about smart material after this class.

So, Let’s get down to the business.This is the title of this lecture.

First, I want to throw out a question ‘What’s surrounding us?’ There are many answers :air,soil,people… But when you surfing the internet, reading news from your APP on your smart-phone or just press the buttons on the screen, you are making a incredible flow of data that flows to all over this world, all over this WorldWideWeb.So my answer is that data is surrounding us.

Data a set of values of ***qualitative*** or ***quantitative*** variables which classify everything in a different view. Data analysis is so important. Almost every filed needs data analysis to dissect phenomenons. Such as in science to do research, in economy to evaluate the risk of stock or else.

There are some common types of data in usual life.(……………)

Data analysis has various categories. Narrative, Descriptive, statistics/mathematical, audio-optical, etc.

Narrative means describe something with other publicatios or any else data source.Descriptive usuall happens in social science to reflect the community.Statistical analysis concentrate on the numeric and number to study the law of finance, physics or something else.

As for this project ,I tend to realize a simple descriptive analysis for myself.It should only involve some plain aspects of this category.So the goal is to gain Dowjones Stocks data and make the visualization.

Then we encounter a question :How to get data?For some company,such as Douban, Google provide its API for users to gain some useful data.But this is not applicable to this project because we should apply for the authentication. There are some other data sites providing dataset. In this project, I choose web crawler to crawl the data or message for myself. WebCrawler is a program or automated script which browses the WorldWide Web in a methodical, automated manner.Like the spyder,from a given point ,it can get everywhere of this web.Here is the basic kind structure of web crawler.First we should give the initial URL that we want crawl and add this url to the frontier.We pull out one url from the frontier ,then get the page,parse it,add urls we extract to the frontier.If frontier is empty,then stop, otherwise circles this procedure.But for me ,I just have to access two webpages :this is about the stock’s symbol and the next is about detail data.

Since we have determine the strcture,it’s time to take action.Here are some tools or libraries I choose.Request is a powerful library for handling HTTP requests.It can easily access web pages. Re is a builtin module of Python,it can be used in regular expression.BeautifulSoup provides powerful html parser to analyse web pages to generate data.Pandas/Numpy/Matplotlib are the core libs of Scipy. Scipy is (………………) .We use these to group data and get charts of the result. At last,using wxPython to build GUI window.

Next is some details of the code of dealing with unordered data.We can quickly skip these.There is the last outcome. And here is the gif .