

Ziyang Zhang
107208048
Jan 25, 2020

Part One: Histories

1. Peter Naur is famously quoted as saying data science “deals with the data, while the actual relation of data to what they represent should occur in other fields.” Why do you think he’d choose to frame data science this way? What might be problematic in this statement?

According to this statement, I would personally think Peter is trying to say that, the relation between data and the data representation is somewhat unclear and indirect related. The problematic part of this statement is, the data that people collected during 1970s were hard to relate with the reality, because the technology strength wasn’t very advanced, so people might not clearly see the relationship between data and their representations.

2. In 2002, data science began to gain momentum as its own dedicated subfield. Compare and contrast the definitions of data science at that time, exemplified by the National Science Foundation, Data Science Journal, and Journal of Data Science, to those from Tukey & Naur in the 1970s.

In the early 1970s, people realized the importance of “data science”, however, nobody was able to tell the reason to explain how people should be utilizing data. In 2002, the definition of “data” became very specific, especially during 1994-2001 when people realized that certain data could be representing certain areas. Until 2002, the figure of “data” became “everything has something to do with data.”

3. Data continues to grow at an exponential rate today. List at least three technological factors that contribute to this growth and what role they play. List three major sources of data that contribute to this growth and at least one way they're being used.

1. Mobile phone, get people to access internet more easily
2. GPS, geographical data
3. Computer, better technology platform

1. Social Media, online streaming data
2. Online Mapping (e.g. google VR map)
3. Google

Part Two: Terminal Crash Course:

4. How would you move into each of the following directories from the shell? Why do you think it is important to have shortcuts for each of these directories for navigating the file structure?

- A. Root `cd /`
- B. Home `cd ~`
- C. Parent `cd ..`

The abbreviation is important which categorized the commands into the same parental list, easily for users to navigate the files.

5. Briefly describe what the following set of commands would achieve. What process would happen and what would be printed to the command line?

```
cd ~  
mkdir ./problem_set_1  
cd ..  
pwd
```

1. Change the directory to home directory
2. Make a new directory called “problem_set_1” within the home directory
3. Change the directory to the parent directory
4. Print working directory

This process is basically creating a new file inside the home directory and then back to the parental directory. The parent directory name would be printed to the command line.