Name: Wakeland Branz

Date: 04/03/2024

ACTIVITY 1

Data, Data, Data!

The ease with which data can be stored, cataloged, and searched in an ever more-connected society is an important point to understand. Data storage has both positive and negative implications, and being able to recognize these, as well as how your actions can contribute to the data being stored, are addressed in the following activity.

1. List two or three sites that you access on a regular basis. Youtube.com, github.com

2. In your group, find the privacy policy for your given site and identify two or three pieces of information that are collected for users of the site.

Aggregated data (e.g. diagnostics), demographics information

File Types

3. What does the word *delimited* mean? Why is this necessary when talking about data files?

Delimited means to have a boundary. This is necessary because it allows data files to be able to store data in an organized manner with defined organizational methods.

4. Given a data file, how can you determine the type of data that might be contained in a specific column?

You can determine the type of data that might be contained in a specific column through basic reasoning (for example, this column has text, must be a string - this column has numbers in the decimal place, must be a float.)

Posing Questions

5. On your own, identify two broad areas of interest that you have. Examples include health, sports, etc.

Olympic weightlifting, computer science

6. For each of the two areas of interest you identified, determine one question to which you might want to know the answer. These should be questions that are not easily answered with an online search. Two examples of questions that are easy to answer with an online search are who won the 1998 Superbowl (The Denver Broncos), and what is the height of the world's tallest building (at the time of publication, Burj Khalifa, 2717 ft/828 m).

What is the optimal diet plan for somebody in the 67kg weight class who is attempting to bulk into the 73kg weight class while doing Olympic style weightlifting. What is the best development environment to learn x64 Intel assembly?

Discuss your two questions with your partner and refine your questions based on their feedback. Write your updated questions below.

What's the optimal diet to bulk from 67kg to 73kg for Olympic weightlifting?What's the best environment to learn x64 Intel assembly?

Finding Data Sets

Your teacher will provide you with several sites that provide free access to data sets. Spend a few minutes looking at the sites.

7. Are there data sets that might apply to one of the two questions that you refined in question 6? Find two different data sets that might be used to answer one of your questions. List the site and any search criteria used to find the data sets. If not, consider revising one of your questions so that you can identify an applicable data set.

The cereal data set can apply to my first question as it relates to a diet plan. My second question does not relate to any data set, so a better question would be: x64 Intel assembly documentation.

8. How many records are in each of the data sets you identified? Describe one benefit of using a larger data set with more records.

There are 78 records in each of the data sets identified. A larger data set allows for more information to be stored and used in search results.

Check Your Understanding

9. With your partner, discuss one way that user data captured by a site (whether knowingly or unknowingly) has contributed to an improvement in the service provided. Have there been any positive impacts of this data outside of the service or website?

On YouTube, the ability to store user search results anonymously allows for videos to be suggested more personally, improving the user's experience. This can give people a better ability to learn certain skills and improve their intelligence to help others.

10. Do you know how the data in the data set you identified was collected? If so, please describe. If not, describe one way that the data might have been collected.

I do not know how the data set I identified was collected. I would guess that it was parsed from numerous cereal brands based on nutritional data from their websites.

11. In your opinion, are there situations where the benefit provided from the data collected is worth the risk to personal privacy? Why or why not?

I believe that when data is extremely tightly controlled and limited, it can be beneficial to the user because it can provide them with a better experience throughout the app.
This must be done responsibly though, and all users must understand exactly what data is being used and where it is going to.