



PURDUE POLYTECHNIC INSTITUTE
Department of Computer and Information Technology

CNIT 27200: Lab #9

25 pts

Due Date

- The entire lab is due via Blackboard on Wednesday, April 15th, by 11:59pm. 25 pts.

STEP 1: The fact that massive amounts of data are being collected about you can be both a blessing and a curse. Consider one of these two scenarios:

- A grocery store chain tracks every purchase you make over time as a loyalty member at each of their locations; the company even tracks your movements over the duration of each visit to the store.
- Amazon tracks every search you have ever performed on the website, every purchase you make, as well as all data gathered on customer transactions related to product returns.

For one of the above two scenarios, discuss the following issues:

1. How could this type of information collection benefit you as an individual? Be specific and give at least two examples.
2. How does collecting this type of information benefit the grocery store or Amazon? Why do they want to collect this information? Again, provide at least two examples by researching online and include the source citation in APA format.
3. How could this information collection be negative for an individual (even someone with nothing to hide)?

4. What are the 4 big data components? Explain each component and how it relates to the scenario you chose for this step.

- 1.
- 2.
- 3.
- 4.

STEP 2A: Watch the following TedTalk about Google Books:

https://www.ted.com/talks/jean_baptiste_michel_erez_lieberman_aiden_what_we_learned_from_5_million_books?utm_campaign=tedsread&utm_medium=referral&utm_source=tedcomshare

STEP 2B: Try Google N-Gram Viewer

When you enter phrases into the Google Books Ngram Viewer, it displays a graph showing how those phrases have occurred in a corpus of books (e.g., "British English", "English Fiction", "French") over the selected years. <https://books.google.com/ngrams/info>

1. Open the N-Gram viewer (<https://books.google.com/ngrams>) , and specify an n-gram phrase to test (at least 3 words). Describe how your n-gram words are related (i.e., the overall subject you are investigating).
2. Provide a screen shot and the describe the graph that was produced.
3. Provide an **objective description** of the popularity of each n-gram relative to one another and over time. Is the popularity of these search terms increasing or decreasing over time? Has the relative popularity of the terms changed at all over time? Are there distinct moments in time when the popularity of the search term has abruptly increased or decreased?

4. Provide a **subjective discussion** about the relative popularity of the terms and their popularity over time. Why do you think some of the n-grams are more popular than others?