CSCU9T4/F – The Assignment 2025

# Task 2: XML Schema Creation and Validation

## c) Justify your design choices. (max 100 words)

# I chose a hierarchical structure with nested elements because things like stops are part of routes in the real world a hierarchical structure represents a real-world relationship between these two pieces of data, it also improves things such as parsing efficiency due to the data that is related is grouped with each other. I used identifiers such as bus and driver ids as attributes rather than elements because of the fact that attributes are better at storing simple metadata, this change simplifies parsing without the parser having to access child nodes.

# Task 3: Designing XML and JSON Structures

## d) Justify your design choices (max 200 words)

XML and JSON are both hierarchical and can be used to model the relationships between each of the historical artifacts, events and people and this creates obvious connections between entities in real life to be understood by the program.

Using the standardized real life date formats(BCE/CE) allows you to remain consistent when sorting using dates while maintaining human readability .  
  
organizing different entities into different sections help separate the data and improve performance when trying to answer a question because only sections where it hold relevant data have to be searched instead of searching the whole dataset.  
  
using IDs to reference information about entities allows for a more maintainable structure for the data while avoiding things such as duplication and help to reduce document size and redundancies for a larger datasets.

Using consistent naming conventions for all of the entities improves the readability of the program and reduces possible confusion while coding for others and myself.  
  
separation of things such as meta data (ID, Dates) from the core content allows less data to be searched, this is done by excluding certain info which means nothing to the search.

# Task 4: Working with XML and JSON in Java

## 2. Justify your choices (max 100 words)

I used Dom for my XML parsing because it produces a in memory tree representation which makes it easy to traverse and change data in a document it also allows for in depth queries to be made. Sax does in fact use less memory but because of the size of the current document the ease of use of Dom makes it a clear choice for this application.

I used Jackson in my Json parser due to the better performance and the scalability for future expansion onto the JSON file, Jackson is typically used in enterprise applications, although GSON is good for being simple and easy to read.

## 3. Formulate and Answer Two Questions

Write your two questions here below and the answers that you get

**Question 1:** **What is the distribution of artifacts by century?**

**Answer to Q1:**

**15th Century CE: 1 artifacts**

**18th Century CE: 1 artifacts**

**2nd Century BCE: 1 artifacts**

**Question 2: Which individuals have the most connections by type?**

**Answer to Q1:**

**John Adams:**

**- Artifact connections: 1**

**- Event connections: 1**

**- Individual connections: 1**

**- Total: 3**

**Thomas Jefferson:**

**- Artifact connections: 1**

**- Event connections: 1**

**- Individual connections: 1**

**- Total: 3**

**Pierre-François Bouchard:**

**- Artifact connections: 1**

**- Event connections: 2**

**- Individual connections: 1**

**- Total: 4**

**Johannes Gutenberg:**

**- Artifact connections: 1**

**- Event connections: 0**

**- Individual connections: 0**

**- Total: 1**

**Napoleon Bonaparte:**

**- Artifact connections: 0**

**- Event connections: 1**

**- Individual connections: 1**

**- Total: 2**

## 4. Comparison of XML and JSON usage (max 400 words)

Write your answer to task 4.4 here, discuss the comparison between your two implementations.  
  
XML and JSON have big differences when it come to their usage both are hierarchical and can be easily used for nested relationships as well as being extremely human readable but when it come to the file size of XML it generally has a larger file size compared to JSON although in this case the XML file size is a single kilobyte smaller then JSON but at scale this would not be the case.  
  
A major difference between the two is the structure they are laid out in while writing JSON and XML files, XML uses tags that’s encompasses data (<fruit>pear</fruit>".)  
while JSON more adds some sort of value to the entity which looks like this

(“Weather”< “Rainy”).  
  
  
In terms of parsing XML with the use of DOM or SAX it requires a more extensive setup to take place using Xpath, while JSON was more direct in terms of interaction during querying I personally found it to be more readable from a programming perspective and was required less lines of code while maintaining the same functionality by poxy improving maintainability of the code.   
  
XML has better data integrity than JSON with the use of schemas to validate datasets and make data follow strict rules and it generally is used when data must be made sure to be correct and structured while JSON more compares to something like a list.  
  
Json is more suited to a shopping cart for a website than Historical data like this program is made to store, XML is the better option for this specific program to ensure the data is historically correct and precise.