THE UNIVERSITY OF MELBOURNE

Workshop 2

COMP20008

Elements of Data Processing

Zijie Xu

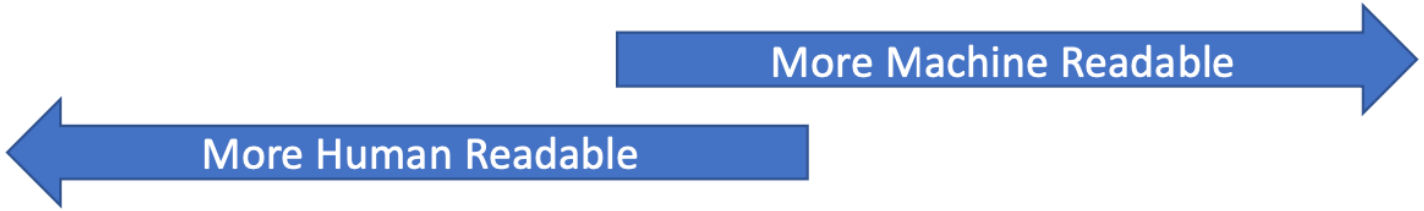
Agenda

- Data formats
- XML
- JSON

Data formats

Categories of data formats

Unstructured	Semi-Structured	Structured
Text files/documents	CSV	Databases
Audio	Webpages (HTML)	Spreadsheets
Video	XML	
Social media data	JSON	



XML

e**X**tensible **M**arkup **L**anguage

- Markup language
  - A well-formed set of rules that “markup” the data with structural information (see lecture slides -> XML syntax)
  - Not a programming language -> XML does not do anything
- Extensible
  - User-defined tags to represent elements
- Represents a tree-like data structure with a single root

JSON

Java**S**cript **O**bject **N**otation

- A lightweight data-interchange format
- A JSON object is a key-value pair enclosed in curly braces
  - { "key": "value" }
  - Keys: string
  - Values: string, number, object, array, boolean, or Null
- Arrays are ordered lists of values enclosed in square brackets
  - [value1, value2, ...]
- Can represent a nested key-value pairs

Example

name	region	weapon	rarity
Raiden Shogun	Inazuma	Polearm	5
Noelle	Mondstadt	Claymore	4

```
<?xml version="1.0"?>
<characters>
  <character>
    <name>Raiden Shogun</name>
    <region>Pyro</region>
    <weapon>Polearm</weapon>
    <rarity>5</rarity>
  </character>
  <character>
    <name>Noelle</name>
    <region>Mondstadt</region>
    <weapon>Claymore</weapon>
    <rarity>4</rarity>
  </character>
</characters>
```

```
{
  "characters": [
    {
      "name": "Raiden Shogun",
      "region": "Inazuma",
      "weapon": "Polearm",
      "rarity": "5"
    },
    {
      "name": "Noelle",
      "region": "Mondstadt",
      "weapon": "Claymore",
      "rarity": "4"
    }
  ]
}
```

HTML vs XML

HTML

- For creating web pages
- Pre-defined tags and attributes
- Focus is on display and rendering, usually not validated

XML

- For data representation
- Custom tags and attributes
- Can be validated against a schema to ensure data consistency

XML vs JSON

XML

- More verbose with start and end tags
- Supports attributes for elements
- Ideal for document-based structures and data with complex hierarchical relationships

JSON

- Concise and easier to read
- Simple key-value pair structure without attributes
- Great for data interchange between web applications and APIs.

THE UNIVERSITY OF MELBOURNE

Thank you

More Resources: Canvas