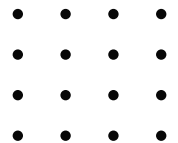Data Modeling Patterns

SEMI STRUCTED DATA USING JSON

What is Semi-Structured Data?

- Not purely structured, and also not completely unstructured

- Contains some level of orgranization but not a rigid schema

Key Characteristics of Semi-Structured Data

Flexible Schema

Does not adhere to a strict, predefined schema

Partial Consistency

Not all data instances need to have the same attributes

Human Readable

Labels and tags, making it more accessible

Hierarchy

Enabling the representation of nested and related data elements

Semi-Structured Data Example

Emails:

A mix of structured elements like sender, recipient, subject, and date, alongside unstructured text within the email body

XML, JSON, CSV:

XML --> encoding structured data using tags
JSON --> structure composed of key-value pairs.
CSV --> store tabular data with values separated by commas

Web Pages:

Organized using HTML tags, which provide a hierarchical structure to the content
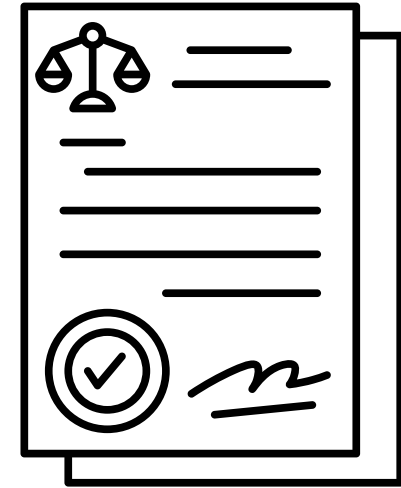
JSON(JavaScript Object Notation)

- Lightweight
- Human-readable
- Easy for machines to parse and generate,
- Organizes data in key-value pairs
- Supports nested structures.

```
{
    "article_id": 3214507,
    "article_link": "http://sample.link",
    "published_on": "17-Sep-2020",
    "source": "moneycontrol",
    "article": {
        "title": "IT stocks to see a jump this month",
        "category": "finance",
        "image": "http://sample.img",
        "sentiment": "neutral"
    }
}
```

Legal Document in JSON

- The Indian Penal Code (IPC) of 1860
- The Civil Procedure Code (CPC) of 1908
- The Hindu Marriage Act (HMA) of 1955

JSON(JavaScript Object Notation)

Current Text
- **Content**
- **POS Tagging of Content**

Law Name
- **POS Tagging of Law Name**

Title
- **Title Name**
- **POS Tagging of Content**
  - **Chapter**
    - **Chapter Name**
    - **POS Tagging**
      - **Article**
        - **Article Content**
        - **POS Tagging**

Transient
- **Content**
- **POS Tagging of Content**

Missing Data in JSON

Using null for missing value

```
[
  {
    _id: ObjectId('67ab2c935c74c7b1afcb0ce7'),
    name: 'Sahil',
    age: 21,
    standard: 10,
    school: 'KV'
  },
  {
    _id: ObjectId('67ab2d305c74c7b1afcb0ce8'),
    name: 'Mohit',
    age: null,
    standard: 8,
    school: 'KV'
  }
]
```

Missing Data(age not entered)

```
[
  {
    _id: ObjectId('67ab31db5c74c7b1afcb0cea'),
    name: 'Sahil',
    age: 21,
    standard: 10,
    school: 'KV'
  },
  {
    _id: ObjectId('67ab31df5c74c7b1afcb0ceb'),
    name: 'Manan',
    standard: 9,
    school: 'DPS'
  }
]
```

Actual Structure

```
[
  {
    _id: ObjectId('67ab2c935c74c7b1afcb0ce7'),
    name: 'Sahil',
    age: 21,
    standard: 10,
    school: 'KV'
  }
]
```

Thank You