

Data Model

* Introduction To Modeling Web Data:

=> Modeling Web data involves the structuring and organizing the data which is available on the web.

Modeling Web Data used to make the data searchable and analyzable.

Effective web data model ensures that data is stored efficiently and can be retrieved in effective way.

We can Model the Web Data in the mainly Two manner.

- 1) Structured Data
- 2) Semi-Structured Data
- 3) Unstructured Data

1 Structured Data:

Structured Data is highly organized and easily searchable in database.

Structured Data has Fixed Format and schema for store the data.

Mainly Structured data is store in the tabular form like rows and column format.

The structure of the data is defined by a schema, which specifies data types and constraints.

Structured Data format support various data type like integer, string, date etc.

In Structured Data format, we can use SQL for perform any complex query.

Ex. Employee Records

2 Semi-Structured Data:

Semi-Structured Data mainly does not used rigid schema like Structured Data.

But, Semi-Structured Data still contain some organized structure.

Semi-Structured Data structure is flexible compare to the structured Data.

Semi-Structure Data can be represent Hierarchical Format like tree structure.

In Semi-Structure Data Format, We can use XPath, XQuery for perform any complex query.

Semi-Structure Data can be represent using formates like XML, JSON or YAML.

Semi-Structured Data can be used Self-describing tag or attributes to store the data.

* Web Data Management with XML :

=> XML stands for Extensible Markup Language which is a versatile format for representing structured and semi-structured data.

Mainly XML is used to store the semi-structured data.

It provides a standardized way to encode documents and data structures.

=> XML Syntax and structure:

-> XML Document Structure:

<root>

<child>

<subchild> ... </subchild>

</child>

</root>

- 1 Root: This is top-level element that contains all other element.
- 2 Child: All the element are nested within this element.
- 3 <subchild>: subchild contain other element.
- 4 Attributes: Provide metadata about elements and is part of start tag.

⇒ XML Standards:

1 XML Schema (DTD):

XML Schema is also called XML Schema Definition which is alternative to DTD use.

XML Schema is used to define structure of an XML Document which is also used to validate the XML Document.

XML Schema is defined whole XML document in the form of tree structure.

The main purpose of an XML schema is to define the legal building blocks of an XML document.

XML Schema contain element and attributes that can appear in a document.

2 XSLT:

XSLT stands for Extensible Stylesheet Language Transformations.

→ XLink Attribute:

- XLink:

- xlink: Specifies the type of link

- xlink:type: Specifies the simple link

- xlink:href: Specifies the target resources URL

- locator: Elements defines the resources to be linked

- arc: Defines relationships between the resources.

Ex. Simple Link:

```

<book
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xlink:href="http://brainspot.org/book">
</book>
  
```

5 XPointer :

XPointer stands XML Pointer Language which is used to extend XPath to provide powerful addressing mechanisms.

XPointer allows you to ^{link} specific element or attributes within an XML document.

XPointer also link specific points within the text of an XML element or range of text within XML document.

XPointer can be used in the XLink to provides proper addressing.

Syntax : #Xpointer (expression)

where, # → URI Fragment identifier
expression → An XPath expression

Ex. #Xpointer (/book/book[1])

XPointer allows for fine-grained linking and referencing within XML documents.