

Basics

XML: extensible Markup Language

HTML: HyperText Markup Language

Tags Predefined (Static)

can create custom tags

<Salary> ... </Salary>

<menu> <food> toast </food>

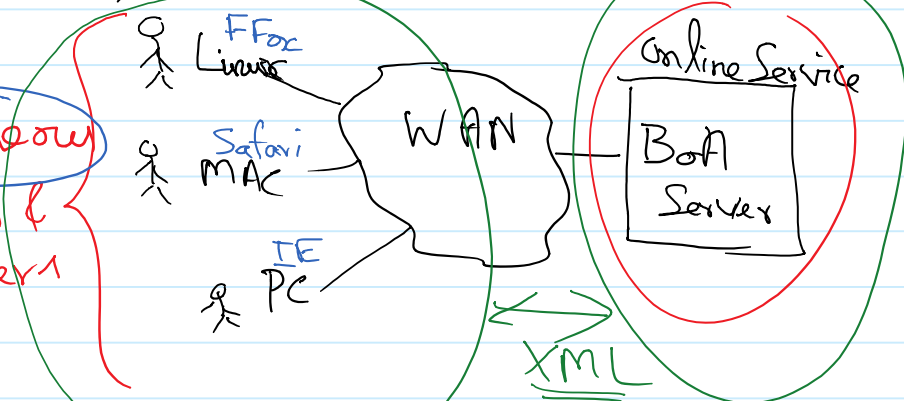
<food> tea </food>

</menu>

<html> ...
<title> ...
</title>
<body> ...
</body>
</html>

Why do we use XML?

heterogeneous
devices &
browsers



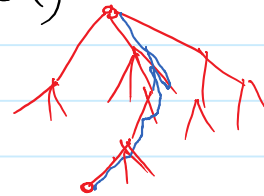
XML facilitates information exchanges between computers.

XML is human readable & machine readable

XML is used for interoperability among heterogeneous systems.

Characteristics of XML:

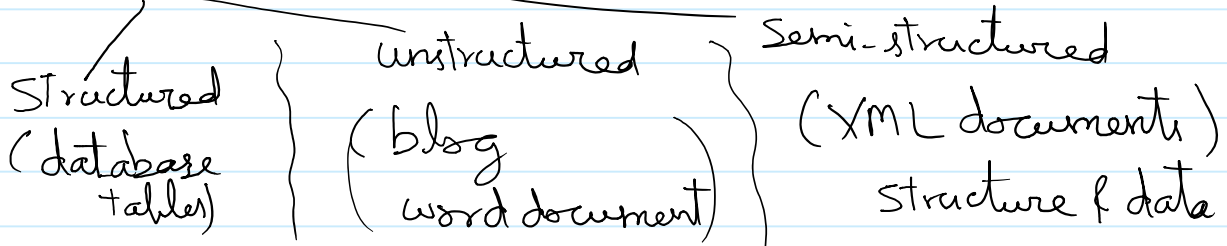
- Text based (all computers & human can interpret XML)
- Self descriptive (data along with structures)
(data and tags) through tags
- XML document is a tree (hierarchical)
Each node (element) of the document has a unique path.



Tags

- are case sensitive.
- must begin with alphabet or underscore
- can include digits, alphabet, -, _, .
- For each tag in the document,
there must be a start tag and end tag
`<book>...</book>`
- Self closing tag
`<pen />`
start end

Document



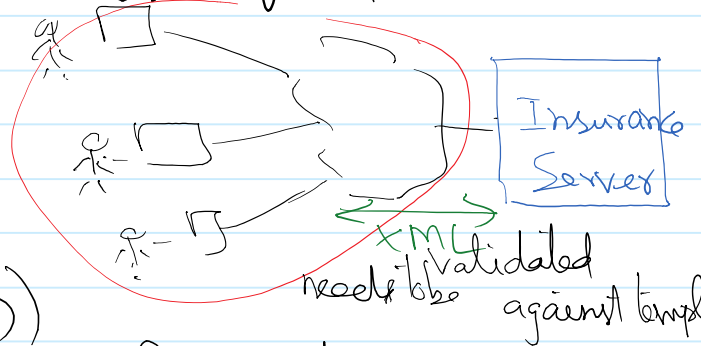
Components of an XML document

Element: $\langle \text{starttag} \rangle \text{value} \text{text} \langle \text{endtag} \rangle$
 nested elements
 $\langle \text{table} \rangle$
 $\langle \text{tr} \rangle \dots \langle / \text{tr} \rangle$
 $\langle \text{tr} \rangle \dots \langle / \text{tr} \rangle$

Attribute: Properties of an element

`<book cover="soft" class="fiction">Harry Potter</book>`

attribute values must be quoted.



Document Type Definition (DTD)

defines structure (pattern) for XML documents.

extended Regular expression. (set theory)

(pattern) ? (occurs one time or none) - optional
+ (occurs at least one time)
* (no occurrence, one occurrence, - ; any no. of occurrences)

Entity is an XML document

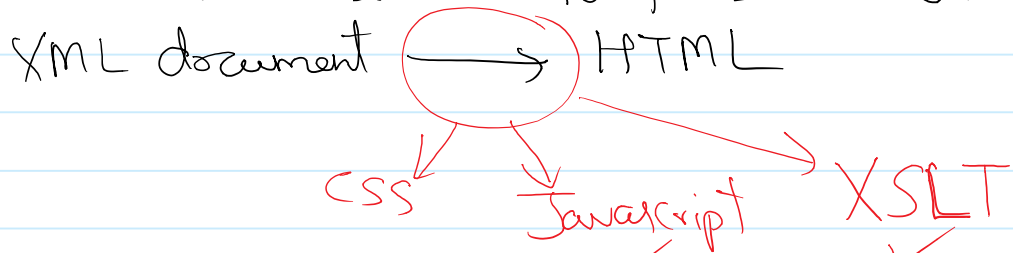
create a shortcut for some descriptive data

internal external

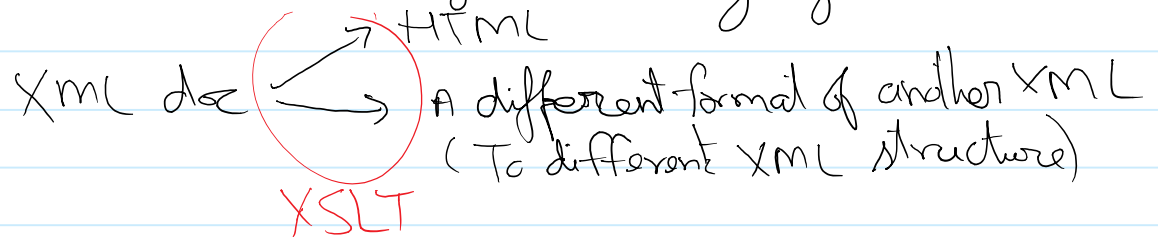
<Entity shortcut "description"> <Entity shortcut SYSTEM " " >
label URI/URL

Transformation

XML document is fully text based
- it is not suitable for presentation.

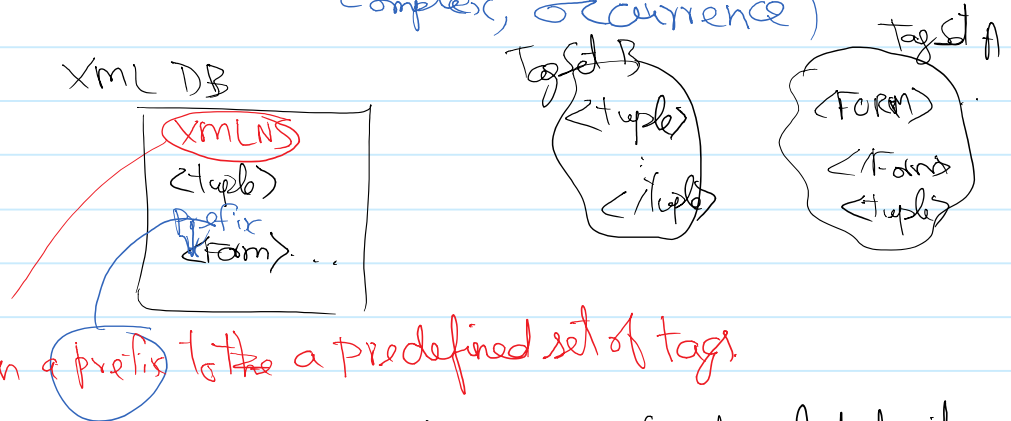


XSLT (extensible Stylesheet Language Transformation)



XML Databases

XML database is an extension of an XML document with constraints (key, KeyRef, unique, Sequence of sub elements, Complex, Occurrence)



For validation of an XML database, structural definition of the database is used.

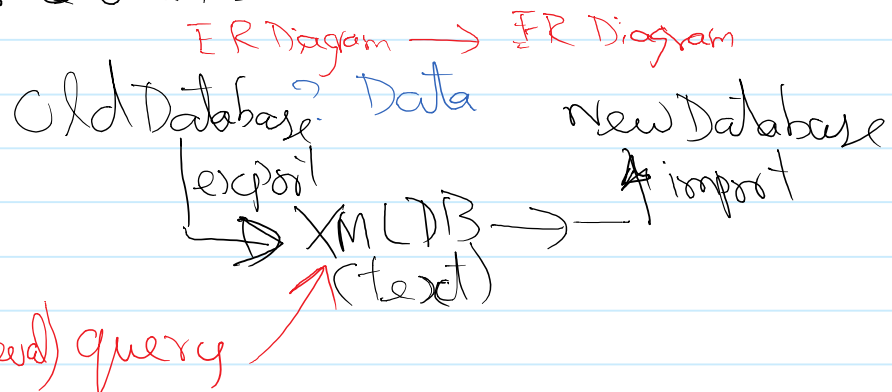
XML Schema
(ER diagram for a DB)
XML Schema for XML DB

Ch 12: Company DB (XML DB)
University DB

text based
tree structure. (no cycles)

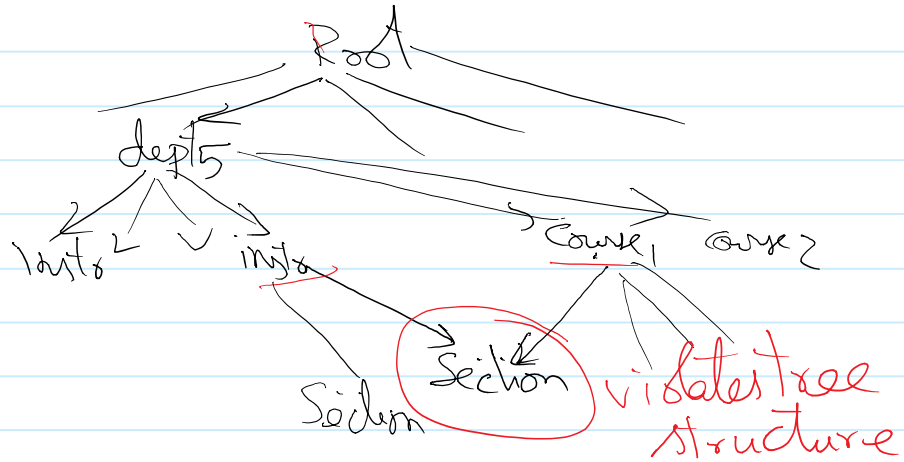
Need for XML DB

For the migration of a production DB
from one DBMS To another DBMS

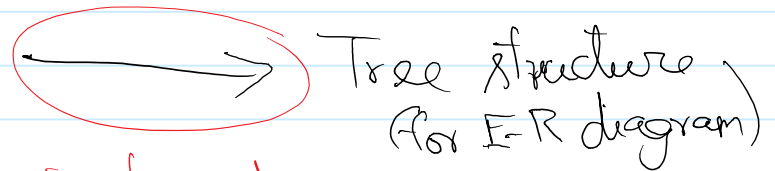


(retrieval) query

University DB
association

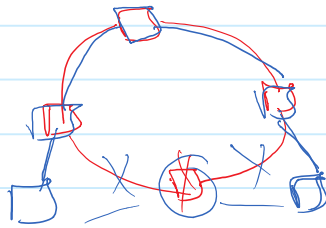


Structure of the DB
E-R diagram



Tree structure
(for E-R diagram)

? cycles $\Rightarrow$ break each cycle $\Rightarrow$ Tree with no cycle.



Querying XML document

Query against XML Document

XPath

XQuery

Path of a search element in an XML document.

(file)

(disk)

/breakfast menu / food / price > 5

Root

// price > 5

NOT a comment

Starting from root

Traverse any no. of nodes down the tree till price element is found.

(//) =>

/ * /

any no. of nodes

Path / absolute (begin from the root)

relative

XQuery: extension of XPath with additional control flow structures. for XML documents
(Similar to SQL for relational DB)

FLWOR
(Flow)

For ----- iteration

Let ----- collection to a variable (aggregate functions can be invoked on the collection)

Where ----- Search condition

Order by ----- Sort criteria

Return ----- (mandatory) ... output result as an XML

Create / test / run XQueries using BaseX

www.w3schools.com