

WEEK 2

Markup

What is markup?

- Markup is something that is used in order to change the way something is shown to the user and ultimately deals with the aesthetics or the sense of beauty or the sense of how nice something looks.

Information Representation

* computers are digital logic systems

- Computer ^{only} works only with 'bits'
 - Binary digits: 0 and 1
- Numbers
 - Place value: binary numbers (eg: 6 = 0110)
 - Two's complement: -ve numbers (eg: -6 = 1010)
- A bit is a binary digit means that it can take only two values - 0 & 1. Everything inside the computer has to be represented using these bits.
- The bits can represent any of the following:
 - i) Sequence of bits
 - ii) Decimal numbers
 - iii) Any alphabet

1 Byte = 8 bits
1 kiloByte = 1024 Byte
1 MB = 1024 KB
1 GB = 1024 MB

Bit is the smallest unit of storage.

apsara

Date: _____

Representing Text

→ ASCII
→ Unicode
→ UTF-8

Information Interchange

- Communicate through machines - either b/w machines or between humans
- Machines only work with 'bits' → i.e. 0 or 1.
- Standard 'encoding' → to interact with computers
 - some sequence of bits interpreted as a character

Interpretation

- What is '01000001'?
 - String of bits
 - Number with value 65 decimal
 - Character 'A'
- It is the matter of INTERPRETATION & CONTEXT.
 - the context in which we see this makes a big difference to how we interpret bits.

ASCII → it is basically for how we can transfer information between devices or between humans & machines.

- American Standard Code for Information Interchange
 - it is a standardized code, which describes how specific strings of bits should be interpreted.
- 7-bits : 128 different entities
 - 'a' to 'z' → 0 to 9
 - 'A' to 'Z'
 - Special characters : ! @ # \$ % & ()

Original ASCII was 7-bit code

Page No. _____

Unicode was sort of created as a consortium.

Unicode

- Allow codes for more scripts, characters
- "Universal Character Set" encoding - UCS
 - UCS-2 : 2 bytes per character (- max 65,536 characters)
 - UCS-4 : 4 bytes per character : 4 billion + characters

EFFICIENCY

- Most common language on Web ??
- Should all characters be represented with same no. of bits?
- EXAMPLE :

- Text document with 1000 words, approximately 5000 characters (including spaces)
- UCS-4 encoding : $32b \times 5000 = 160\,000$ bits
- ASCII encoding : $8b \times 5000 = 40,000$ bits
- Original 7-bit ASCII sufficient for English : $7b \times 5000 = 35k$ bits
- Minimum needed to encode just 'a' to 'z', numbers and some special characters : could fit in 6 bits : 30000 bits
- Optimal encoding based on frequency of occurrence

PREFIX CODING

1 st Byte	2 nd Byte	3 rd Byte	4 th Byte	Free Bits	Maximum Expressible Unicode Value
0xxxxxxx				7	007F hex (127)
110xxxxx	10xxxxxx			(5+6) = 11	07FF hex (2047)
1110xxxx	10xxxxxx	10xxxxxx		(4+6+6) = 16	FFFF hex (65535)
11110xxx	10xxxxxx	10xxxxxx	10xxxxxx	21	10FFFF hex (1114111)

Example

	A	X	女	不
Code Point	U+0041	U+05DD	U+597D	U+233B4
UTF - 8	41	D790	E5A5 BD	F0 A3 8E B4
UTF - 16	00 41	05 DD	59 7D	D8 4C DF B4
UTF - 32	00 00 00 41	00 00 05 DD	00 00 59 7D	00 02 33 B4

UTF - 8

- Use 8 bits for most common characters : ASCII subset
 - All ASCII documents are documents automatically UTF 8 compatible
- All other characters can be encoded based on prefix string.

→ More difficult for text processor :

- First check prefix
- Linked list through chain of prefixes possible
- Still more efficient for majority of documents

algorithm 3

N	E	S	A
SEARCH	CASE	OFFICE	PI
SEARCH	CASE	OFFICE	PI
SEARCH	CASE	OFFICE	PI
SEARCH	CASE	OFFICE	PI

Markup

Content vs Meaning
Types of Markup
(x) HTML

CONTENT

What is Markup?

→ Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed. Markup is very useful to make the display of text clear and easy to understand.

TYPES OF MARKUP

• PRESENTATIONAL

- WYSIWYG : directly format output & displays
- Embed Codes not part of regular text, specific to the editor

• PROCEDURAL

- Details on how to display:
 - change font to large, bold
 - skip 2 lines, indent 4 columns

• DESCRIPTIVE

- This is a <title>, this is a <heading>, this is a <para>.

Examples :

(1) MS Word, Google Docs etc.

- User Interface focused on 'appearance', not meaning
- WYSIWYG : direct control over styling
- Often leads to complex formatting & loss of inherent meaning.

(2) LaTeX, HTML (general *ML)

- Focus on meaning
- More complex to write & edit, not WYSIWYG in general

SEMANTIC MARKUP

LAU01AT01323099

→ Content vs Presentation

→ Semantics

• Meaning of the text

• structure or logic of the document

Introduction to HTML

HYPERTEXT MARKUP LANGUAGE

- HTML first used by Tim Berners Lee in original Web at CERN (~1989)
- Considered an application of SGML (Standard Generalized Markup Language)
 - Strict definitions on structure, syntax, validity
- HTML meant for browser interpretation

HTML Example :

```
<!DOCTYPE html>
<html>
<body>
<h1> My First Heading </h1>
<p> My First paragraph </p>
</body>
</html>
```

TAGS

- <h1></h1> paired tags
- Angle brackets <>
- Closing with /
- Case insensitive
- Location specific:
<DOCTYPE> - only at the head of doc.

NESTING

→ Hello

PRESERVATION vs SEMANTICS

- Tag: `strong` ` Hello `
- Tag: `b` ` Hello `

Which one is right? Which is better?

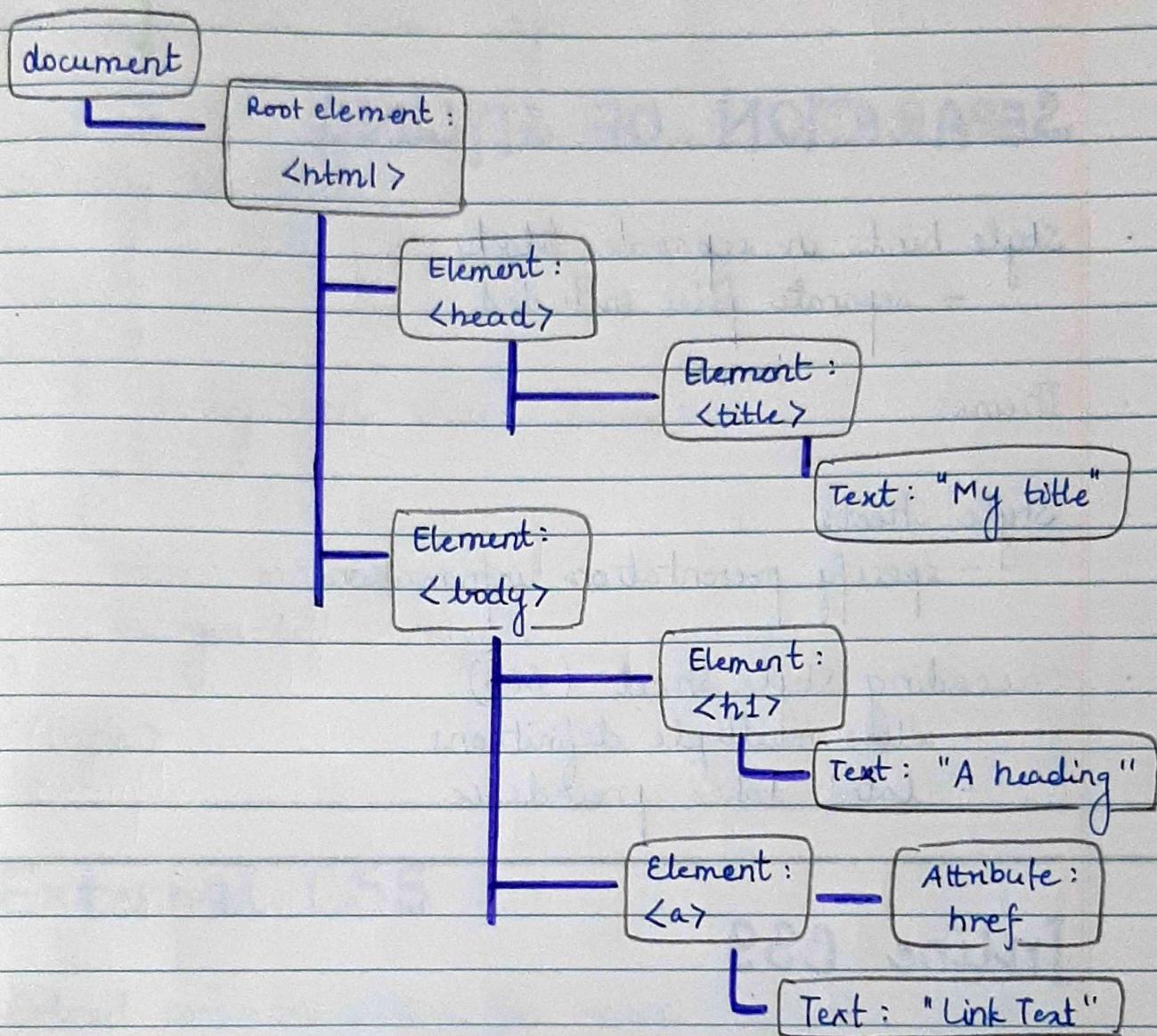
TIMELINES

- SGML based
 - 1989 - HTML Original
 - 1995 - HTML 2
 - 1997 - HTML 3,4
- XML based
 - XHTML - 1997 to mid 2010s
- HTML 5
 - first release 2008
 - W3C recommendation - 2014

HTML 5

- Block Elements : `<div>`
- Inline Elements : ``
- Logical Elements : `<nav>`, `<footer>`
- Media : `<audio>`, `<video>`

DOCUMENT OBJECT MODEL



- Tree structure representing logical layout of document
- Direct manipulation of tree possible!
- Application Programming Interfaces (APIs)
 - canvas ; offline ; web storage ; drag & drop
- JS primary means of manipulating
- CSS used for styling

Introduction to Styling

SEPARATION OF STYLING

- Style hints in separate blocks
 - separate files included
- Themes
- Style Sheets
 - specify presentation information
- Cascading Style Sheet (CSS)
 - allow multiple definitions
 - latest takes precedence

Inline CSS

- Directly add style to the tag
- Example : `<h1 style = "color : blue ; text-align : centre ;">`
A heading
`</h1>`

Internal CSS

- Embed inside `<head>` tag
- Now all `<h1>` tags in document will look the same - centrally modified.

```
<style>
body {
    background-color: linen;
}
h1 {
    color: maroon;
    margin-left: 40px;
}
</style>
```

External CSS

- Extract common content for reuse
- Multiple CSS files can be included
- Latest definition of style takes precedence

Responsive Design

- Mobiles & Tablets have smaller screens (different form factors)
- Adapt to screen - RESPOND
- CSS controls style - HTML controls content!

Bootstrap

- Commonly used framework
 - originated from Twitter
 - widely used now
- Standard styles for various components
 - Buttons
 - Forms
 - Icons
- Mobile first: highly responsive layout

Javascript

- Interpretted language brought into the browser
- Not really related to Java in any way - formally ECMA Script
- Why?
 - HTML is not a programming lang.
 - CSS is not a prog lang.
- Would still like to have "programmability" inside browser
- Not part of the core presentation requirements
 - very useful, but will be considered later.