

Text Formats

- word processor files are binary
- text files have advantages
- human-readable
- easier to maintain (version control)
- easier to exchange
- easier to convert (to web page, to PDF, ...)

Markup

- how to use text-based formats to describe documents?
- add directives to text to mark regions
- content elements
- structure
- style

Markup Languages

- markup languages define rules for marking text
- Markdown, ReStructured Text, ...
- e.g. stars for italics:

In the mid 1960s, there were so many British groups in the US music charts that the Americans called it the **British invasion**.

Markup Types

- semantic vs stylistic
- “emphasis” is semantic
- “italics” is stylistic
- *semantic is preferred*

WYSIWYM

- not WYSIWYG
- for example, where will lines be wrapped?
- WYSIWYM: What You See Is What You Mean

HTML

- HyperText Markup Language
- a document consists of *elements*
- e.g. paragraph: `p`, emphasized text: `em`

Tags

- elements are represented by *tag* pairs
- opening tag: tag name in angular brackets
- closing tag: add slash in front of tag name

In the mid-1960s, there were so many British groups in the US music charts that the Americans called it the ``British invasion``.

Paragraph Example

<p>The Beatles are regarded as the most important and influential band in the history of rock music. They are the best-selling band with more number one albums and singles than any other band or artist.</p>

Whitespace

- whitespace is insignificant

In the mid-1960s, there
 were so many British

 groups in the US music charts that the Americans
 called it the *British*
invasion

Comments

- anything between `<!--` and `-->` is a comment
- ignored by the browser

```
<p>The Beatles are regarded as the most important  
and influential band <!-- after Herman's Hermits -->  
in the history of rock music. They are  
the best-selling band with more number one albums  
and singles than any other band or artist.</p>
```

Attributes

- elements can have attributes
- name-value pairs
- opening tag: `<element name="value">`

Attribute Example

```
<p title="Introductory information about The Beatles">  
  The Beatles are regarded as the most important  
  and influential band in the history of rock music.  
  They are the best-selling band with more number one  
  albums and singles than any other band or artist.  
</p>
```

Nesting Elements

- elements can be nested

```
<p>In the mid 1960s, there were so many British groups  
in the US music charts that the Americans called it  
the <em>British invasion</em>.</p>
```

Nesting Rule

- subelement opens later and closes earlier
- invalid:

`<p>In the mid 1960s, there were so many British groups
in the US music charts that the Americans called it
the British invasion</p>.`

Self Closing Tags

- some elements can not have subelements
- single tag, ends with `/>`
- thematic break: `hr`
- line break: `br`

Self Closing Tag Example

```
<p>The Beatles are regarded as the most important and  
influential band in the history of rock music.</p>
```

```
<br/>
```

```
<p>The band was formed in Liverpool<br/>in 1960.</p>
```


Element Hierarchy

- nesting elements creates a hierarchy
- top level element: `html`
- subelements:
 - metadata of document: `head`
 - body of document: `body`

Document Template

- document type
- document language attribute

```
<!DOCTYPE html>
<html lang="en">
  <head>
    METADATA OF DOCUMENT
  </head>
  <body>
    CONTENT OF DOCUMENT
  </body>
</html>
```

Indentation

- hierarchy is reflected in indentation to improve readability

```
html
├── head
└── body
```

Metadata Elements

- title: `title`
- others: attributes of `meta` (self closing)
 - character set
 - author, copyright
 - keywords

Metadata Example

```
<head>  
  <title>The Beatles</title>  
  <meta charset="utf-8" />  
</head>
```

Content Elements

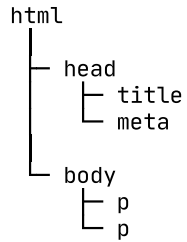
- paragraphs, lists, tables, ...
- intra-paragraph: emphasis, links, ...
- embedded: image, audio, video, ...
- headers, sectioning

Full Example

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title>The Beatles</title>
    <meta charset="utf-8"/>
  </head>
  <body>
    <p>The Beatles are regarded as the most important and
      influential band in the history of rock music.</p>

    <p>The band was formed in Liverpool in 1960.</p>
  </body>
</html>
```

Hierarchy of Example



Emphasis

- emphasized: **em**

`<p>In the mid 1960s, there were so many British groups
in the US music charts that the Americans called it
the British invasion.</p>`

└ body
 └ p
 └ p
 └ em

Importance

- important: **strong**

```
<p>Their famous lineup, called  
<strong>The Fab Four</strong>, consisted  
of the following members:</p>
```

Character References

- how to represent `<`?
- not to be confused with an opening tag
- character reference: `&name;`

Character Reference Examples

Character Reference

< <

> >

& &

“ "

Unicode References

- numeric character references: `&#number;`
- usually written in hex: start with `x`

Unicode Reference Examples

Character Reference

< <

< <

Ě Ğ

€ €

Ω Ջ

Non-breaking Space

- make sure line wrapping doesn't separate words: ` `;

```
<p>Of the albums, "Sgt.&nbsp;Pepper" is considered  
to be one of the...</p>
```

Preformatted Text

- to preserve format: `pre`

```
<p>Here's how you create an unordered list  
in Markdown:</p>
```

```
<pre>  
- item 1  
- item 2  
- item 3  
</pre>
```


Lists

- unordered list: `ul`
- ordered list: `ol`
- subelements are list items: `li`

```
└─ ul
   └─ li
   └─ li
   └─ ...
```

```
└─ ol
   └─ li
   └─ li
   └─ ...
```

Unordered List Example

`<p>Their famous lineup, called "The Fab Four", consisted of the following members:</p>`

```
<ul>
  <li>John Lennon (rhythm guitar)</li>
  <li>Paul McCartney (bass guitar)</li>
  <li>George Harrison (lead guitar)</li>
  <li>Ringo Starr (drums)</li>
</ul>
```

Ordered List Example

`<p>The Beatles released 13 albums in the UK:</p>`

``

`Please Please Me (1963)`

`With the Beatles (1963)`

`...`

`Abbey Road (1969)`

`Let It Be (1970)`

``

Definition Lists

- definition list: `dl`
- every item is a term: `dt`, and a description: `dd`

```
└─ dl
    ├── dt
    ├── dd
    ├── dt
    ├── dd
    └── ...
```

Definition List Example

```
<dl>
  <dt>John Lennon</dt>
  <dd>Lennon released several successful albums...</dd>

  <dt>Paul McCartney</dt>
  <dd>McCartney had a long lasting solo career...</dd>

  <dt>George Harrison</dt>
  <dd>Harrison had moderate success...</dd>

  <dt>Ringo Starr</dt>
  <dd>Starr also released singles and albums...</dd>
</dl>
```

Images

- image: `img`
- source attribute: `src`
- source can be a relative path or a web address
- alternate text: `alt`
- always provide one for accessibility

Image Example

```

```

```
<p>Of the albums, "Sgt. Pepper" is considered to be  
  one of the best rock albums of all time.</p>
```

Headings

- six levels of headings
- h1, h2, ..., h6
- no structure between the levels

Heading Example

```
<h1>The Beatles</h1>
...
<h2>History</h2>
<p>The band was formed in Liverpool, England...</p>
...
<h2>Discography</h2>
<p>The Beatles released 13 albums in the UK:</p>
...
<h2>Later Careers</h2>
<p>After the band broke up...</p>
...
```

Links

- anchor: `a`
- destination address attribute: `href`
- absolute addresses: starting with `http://`
- relative addresses: starting from the current path

Link Example

<p>

Although their popularity increased in the UK,
at first they were rejected by major record companies.
A well-known story is about Decca Records
turning them down, saying that "guitar groups are
on the way out". Finally, in 1962,
they signed with EMI.

</p>

Images as Links

- images can be made into links
- nest under anchor element

```
<a href="http://en.wikipedia.org/wiki/Sergeant_Pepper">  
    
</a>
```

Section Anchors

- in-document links
- `id` attributes for specifying target

Section Anchor Example

```
<p>...Although <a href="#history">they started out  
playing rock and roll</a>, they later evolved...</p>
```

```
...
```

```
<h2 id="history">History</h2>  
<p>The band was formed in England...</p>
```

Tables

- top level: `table`
- each row: `tr`
- each cell: `td`

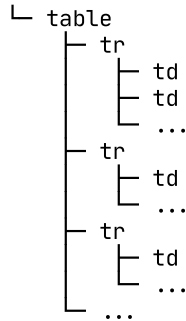


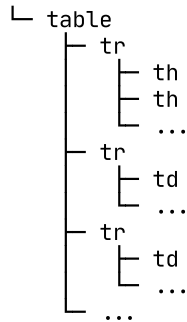
Table Example

```
<table>
  <tr>
    <td>No</td>
    <td>Title</td>
    <td>Lead Vocals</td>
    <td>Duration</td>
  </tr>
  <tr>
    <td>1</td>
    <td>Sgt. Pepper's Lonely Hearts Club Band</td>
    <td>McCartney</td>
    <td>2:02</td>
  </tr>
</table>
```


Header Row

- in header row:

th instead of td

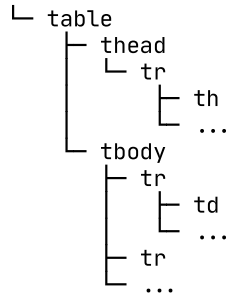


Header Row Example

```
<table>
  <tr>
    <th>No</th>
    <th>Title</th>
    <th>Lead Vocals</th>
    <th>Duration</th>
  </tr>
  <tr>
    <td>1</td>
    <td>Sgt. Pepper's Lonely Hearts Club Band</td>
    <td>McCartney</td>
    <td>2:02</td>
  </tr>
</table>
```

Row Groups

- header: `thead`
- data: `tbody`
- summation: `tfoot`



Row Groups Example

```
<table>
  <thead>
    <tr>
      <th>No</th>
      <th>...</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <td>1</td>
      <td>...</td>
    </tr>
  </tbody>
</table>
```

Multi-column Cells

- span multiple columns: `colspan` attribute

```
<tr>
  <th colspan="2">Title</th>
  <th>Lead Vocals</th>
  <th>Length</th>
</tr>
<tr>
  <td>1</td>
  <td>Sgt. Pepper's Lonely Hearts Club Band</td>
  <td>McCartney</td>
  <td>2:02</td>
</tr>
```

Multi-row Cells

- span multiple rows: `rowspan` attribute

```
<tr>
  <td>4</td>
  <td>Getting Better</td>
  <td rowspan="2">McCartney</td>
  <td>2:48</td>
</tr>
<tr>
  <td>5</td>
  <td>Fixing a Hole</td>
  <td>2:36</td>
</tr>
```