Web Information System Design
No.3 Web Documents

Tatsuya Hagino (hagino@sfc.keio.ac.jp)

HTML

- Please write a simple HTML document.
  - include HTML specific features

```
<!DOCTYPE html>
<html>
<head>
  <link href="style.css" rel="stylesheet" type="text/css" />
  <title>My Page</title>
</head>
<body>
<h1>My Page</h1>
<p>This is just a <em>test</em> page.</p>
<ul>
  <li><a href="http://www.sfc.keio.ac.jp/">SFC</li>
</ul>
</body>
</html>
```

Documents

- Printed documents
  - Books
  - Newspapers
  - Articles
  - Posters

- Online (computer) documents
  - Word documents
  - PDF
  - Web pages
  - e-book

Online documents

- Free from the restrictions of paper:
  - paper size
    - paper height
    - paper width
  - page
    - number of pages
    - thickness
  - multimedia
    - audio and video
  - hyperlink
    - table of contents
    - index

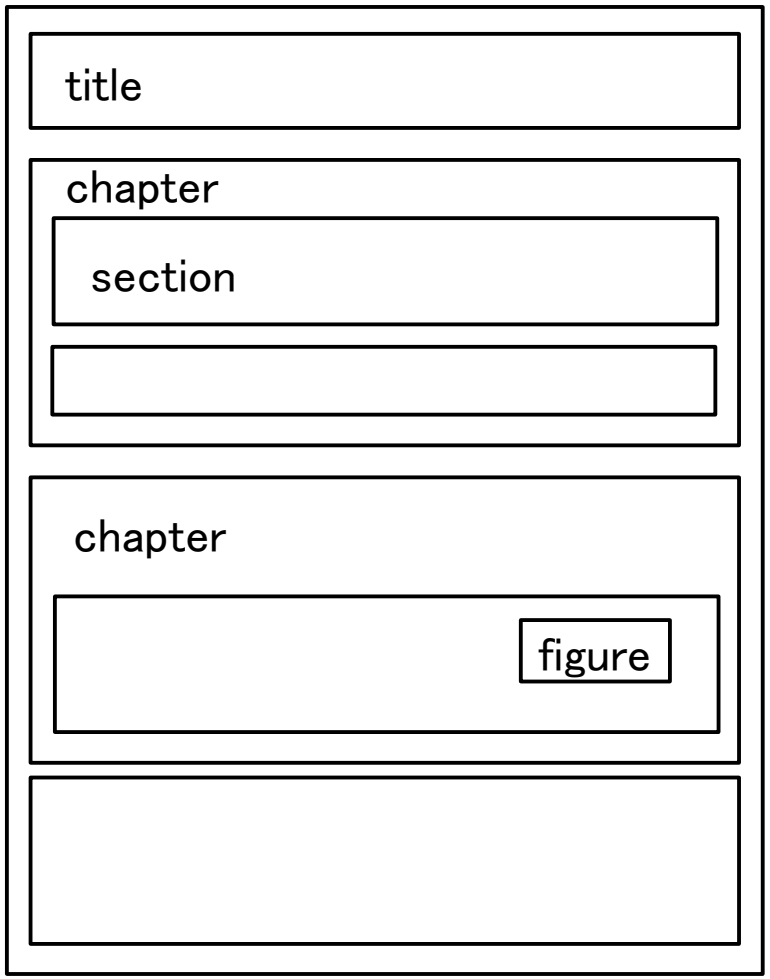
What is `document'?

- document = doc− + −ment
  - doc−: latin `docere' (= to teach)
  - −ment: make it noun
  - `document' is the thing which teaches something.
  - `document' is for telling some information.
  - Longman
    1. a piece of paper that has official information on it
    2. a piece of written work that is stored on a computer

- Similar words
  - documentary
  - documentation

Structured Document

- Documents have structure:
  - title
  - author
  - abstract
  - chapter, section, subsection
  - itemize, enumeration
  - figure, table
  - index, table of contents



Structured programming

C Programming Language

```
x = 0;
y = 0;
i = 1;
while (i <= 100) {
  if (i % 2 == 0) {
    x = x + i;
  }
  else {
    y = y + i;
  }
  i = i + 1;
}
```

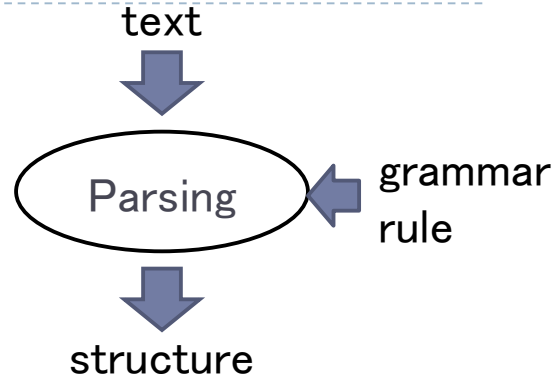
BASIC

```
100 let x = 0
110 let y = 0
120 let i = 1
130 goto 170
140 let i = i + 1
150 if int(i/2)=i/2 then 190
160 goto 210
170 if i <= 100 then 150
180 end
190 let x = x + i
200 goto 140
210 let y = y + i
220 goto 140
```

- No goto statement!
- Structured programs are easy to understand and easy to express algorithm

How to give structure

- Use specific syntax
  - structured programming language
  - mathematical formula



- Embed codes which specify structure
  - LaTeX environment
  - SGML mark-up
    - `mark-up' = marking up
    - Traditionally, instructions by editors written with a blue pencil
    - In SGML, use tags

SGML: Standard General Markup Language

- ISO standard in 1986.

ISO 8879:1986

0 Introduction

This International Standard specifies a language for document representation referred to as the "Standard Generalized Markup Language" (SGML). SGML can be used for publishing in its broadest definition, ranging from single medium conventional publishing to multi-media data base publishing. SGML can also be used in office document processing when the benefits of human readability and interchange with publishing systems are required.

DTD: Document Type Definition

Define the structure of documents

```
<!ELEMENT chapter - - (title, section+)>
<!ELEMENT title o o (#PCDATA)>
<!ELEMENT section - - (title, paragraph+)>
<!ELEMENT paragraph - - (#PCDATA)>
```

Document

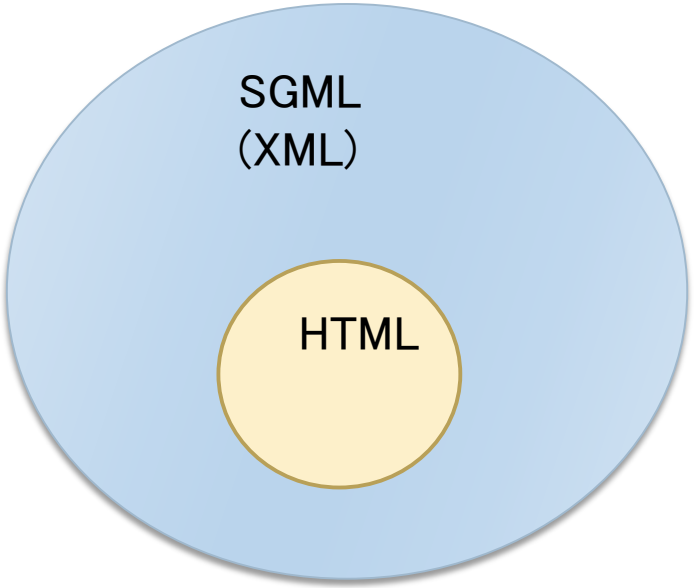
Use tags to specify the structure

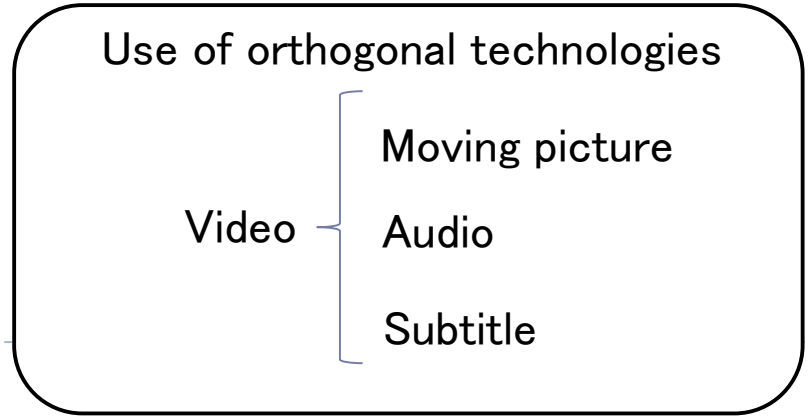
```
<chapter>
<title>Web documents</title>
<section>
What is HTML
<paragraph>HTML is the format for
Web documents.</paragraph>
<paragraph>HTML is used
widely.</paragraph>
</section></chapter>
```

What is HTML?

- HTML
  - SGML application
  - Hypertext documents

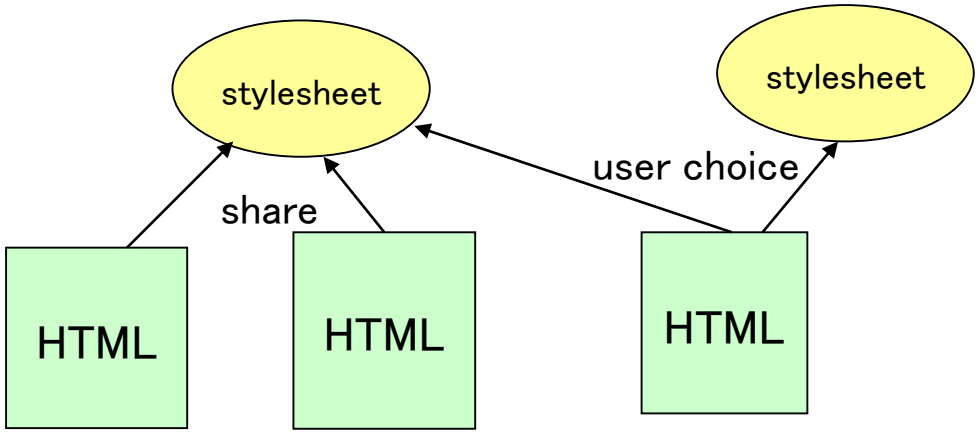
- HTML features
  - Separation of content and presentation
  - Style is specified by CSS
  - Mix orthogonal technologies
    - Content: HTML
    - Style: CSS
    - Programming: Javascript



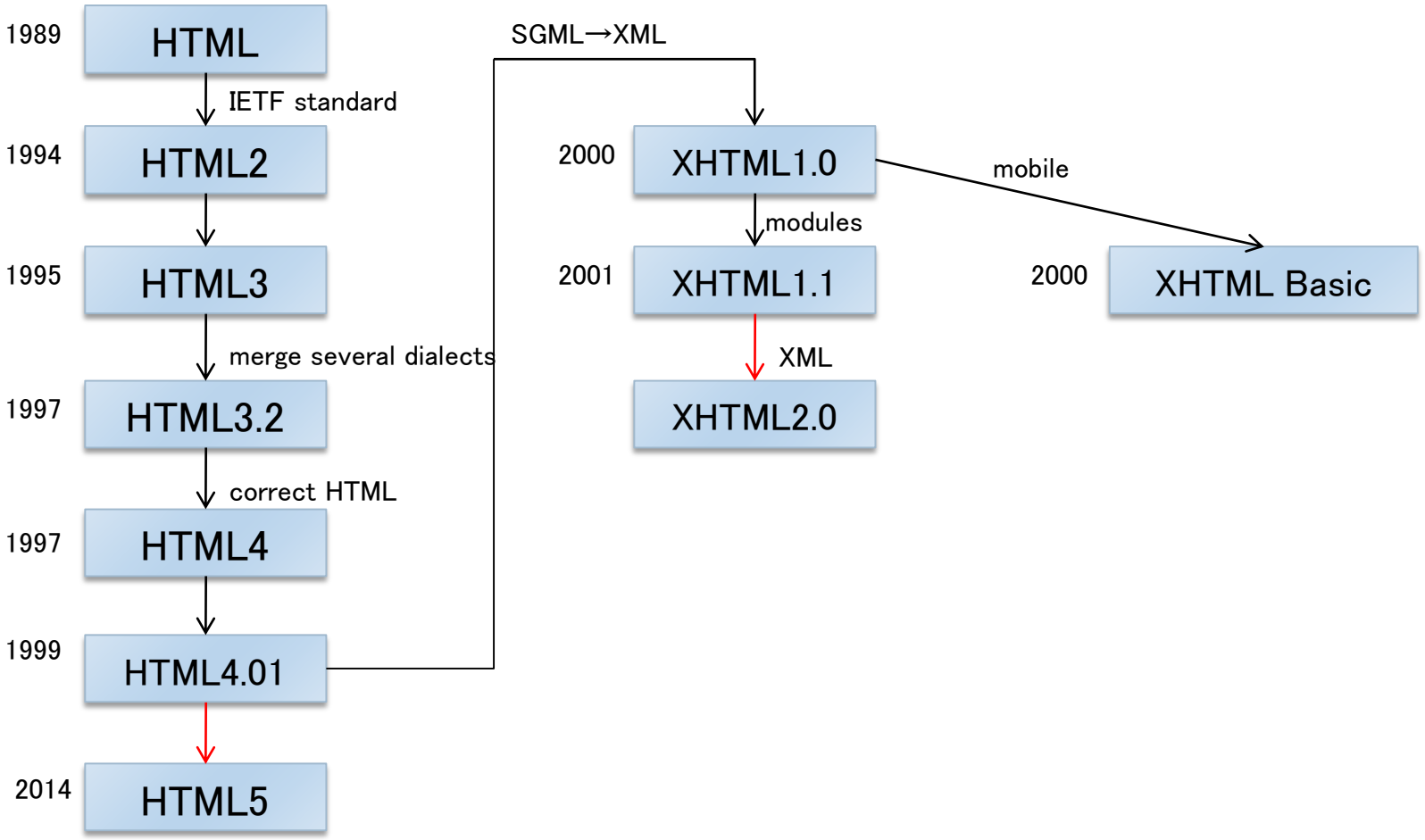


Separation of content and presentation

- Merit of separation
  - The role of HTML is clear.
  - Possible to change style without changing content.
  - Share the same style for multiple documents.
  - Keep the uniformity of the site.
  - User can choose or change the style.
    - For better accessibility



HTML versions and history

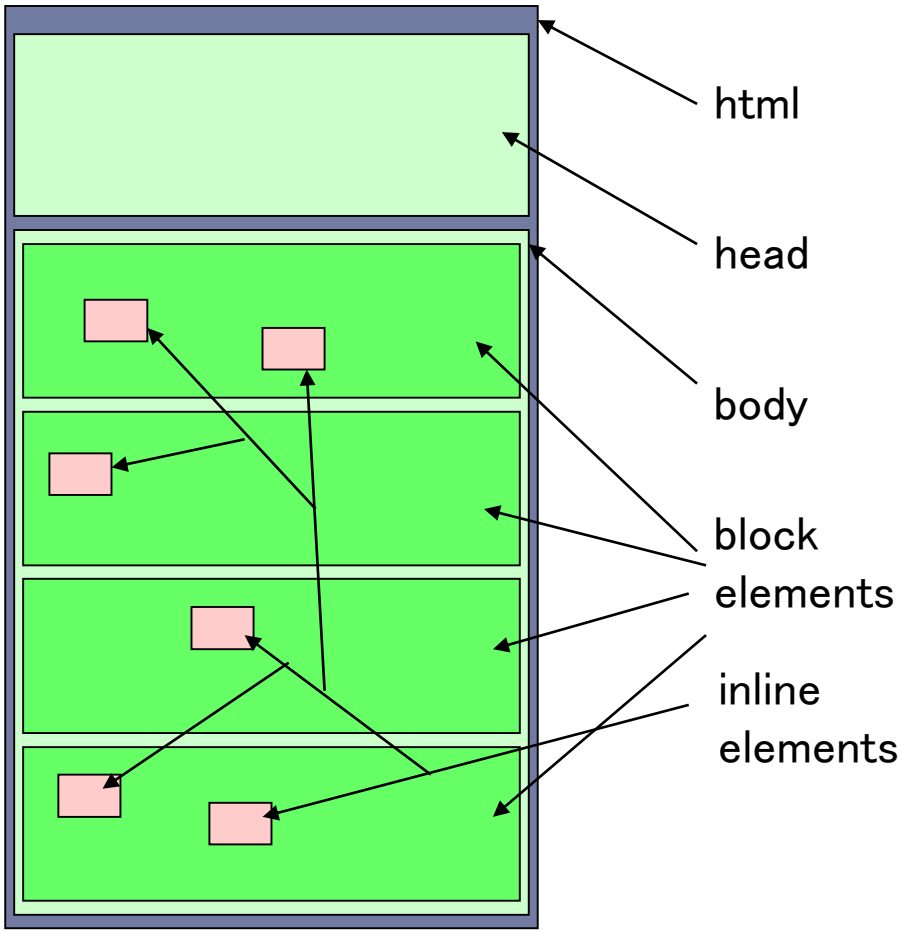


Classify HTML tags

- Structure tags
  - address, article, aside, audio, blockquote, body, canvas, caption, col, colgroup, dd, details, dialog, div, dl, dt, embed, fieldset, figcaption, figure, footer, form, h1–6, head, header, html, iframe, img, legend, li, map, menu, nav, noscript, object, ol, p, pre, q, script, section, summary, table, tbody, td, tfoot, th, thead, title, tr, ul, video
- Semantic tags
  - abbr, cite, code, del, dfn, em, i, ins, kbd, mark, meter, rp, rt, ruby, s, samp, small, span, strong, sub, sup, time, var
- Others
  - a, area, b, base, bdi, bdo, br, button, command, datalist, hr, input, keygen, label, link, meta, optgroup, option, output, param, progress, select, source, style, textarea, track, u, wbr

HTML Elements Category

- Structure elements
  - html, head, body, ...
  - section, article, ...
- Paragraph elements
  - block elements
  - h1, h2, ul, ol, table, ...
- Phrase elements
  - inline elements
  - text elements
  - i, b, em, strong, ...



HTML is for content

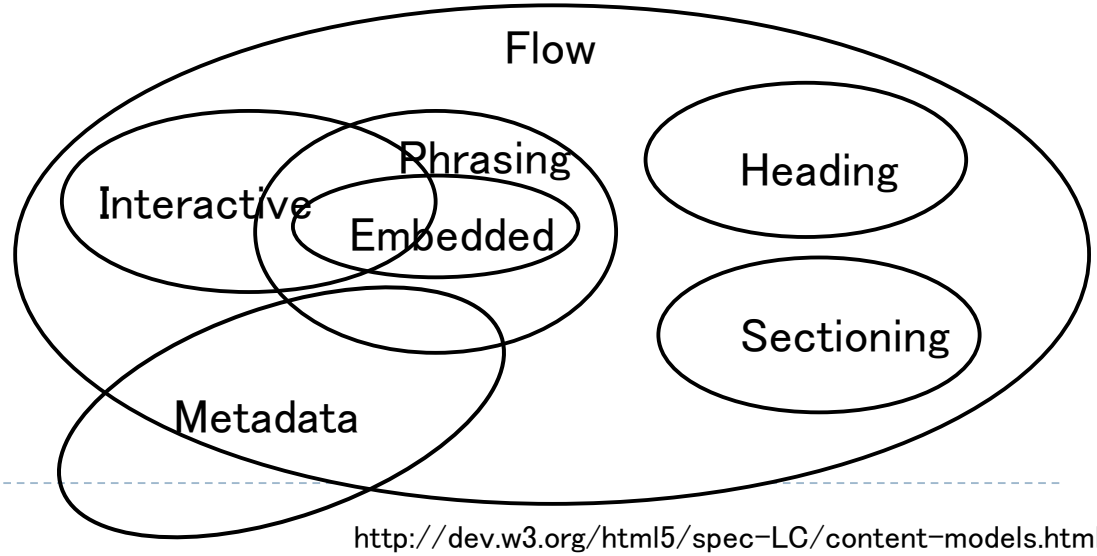
- Elements without content meaning
  - center
  - font
- Attributes without content meaning
  - align attribute
  - bgcolor attribute
  - type attribute for ol and ul
- Elements with content meaning
  - em
  - strong
  - cite
  - dfn
  - code
  - samp
  - kbd
  - var
  - abbr

Think why you want to have bold or italic fonts

- Emphasize
  - This is important.
  - This is important.
  - This is very important.
- First appearance, definition
  - HTML is the language of describing Web pages.
  - <dfn>HTML</dfn> is the language of describing Web pages.
- Input/output of computer programs
  - The assignment statement <code>x=x+1</code> increases the value of <var>x</var> by one.
- Abbreviation
  - At Keio University <abbr title="Shonan Fujisawa Campus">SFC</abbr>, students can choose various subjects to study.

HTML5 Content Model

- Metadata content
  - Elements for specifying CSS, Javascript, ...
  - Specify relationship with other documents
- Flow content
  - Elements for body
- Sectioning content
  - Define scope of headings and footers
- Heading content
  - Header of a section
- Phrasing content
  - text
  - mark-up of intra-paragraph level
- Embedded content
  - Embed non-text contents
  - Import another resource
- Interactive content
  - Specify user interaction

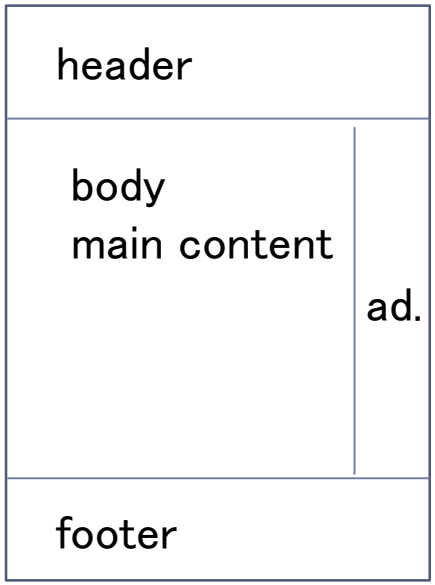


http://dev.w3.org/html5/spec-LC/content-models.html

Standard body structure of HTML5

- HTML5 has elements for representing standard pages.

```
<body>
  <header>
    header
  </header>
  <article>
    body
    main content
  </article>
  <aside>
    advertisement
  </aside>
  <footer>
    footer
  </footer>
</body>
```



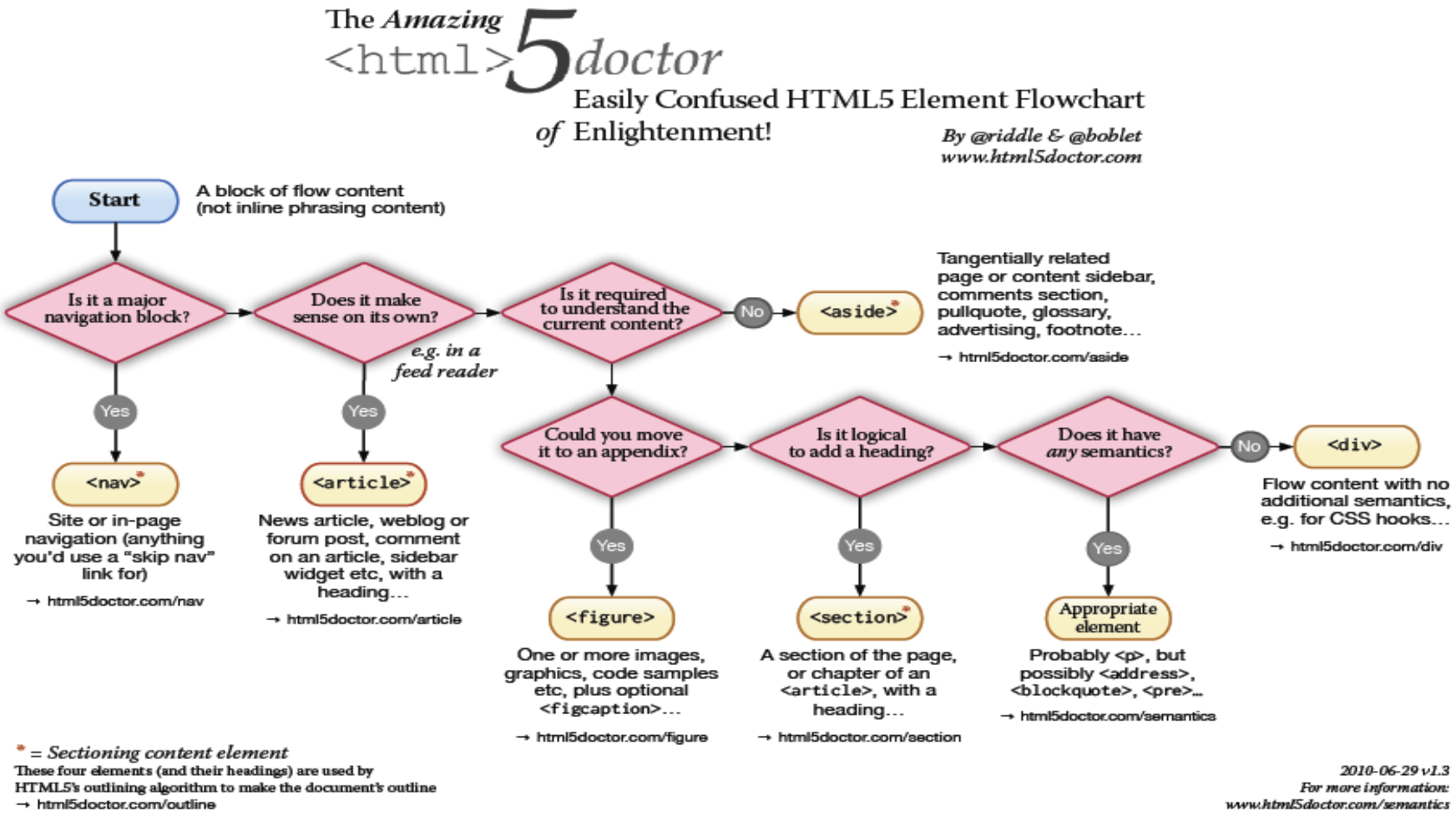
article may contain header and footer

HTML5 Sectioning Elements

- div
  - a division
  - no specific meaning
- section
  - a general section
  - a group of paragraphs
  - may have header and footer
- article
  - a self-contained content
  - can be separated out for reuse
- nav
  - a section containing navigation
- aside
  - a section separate from the main content

New for HTML5

Which sectioning elements to use?



http://html5doctor.com/wp-content/uploads/HTML5Doctor-sectioning-flowchart.pdf

h1-6 and section

- Before HTML5
  - h1 is the highest level
  - h6 is the lowest level

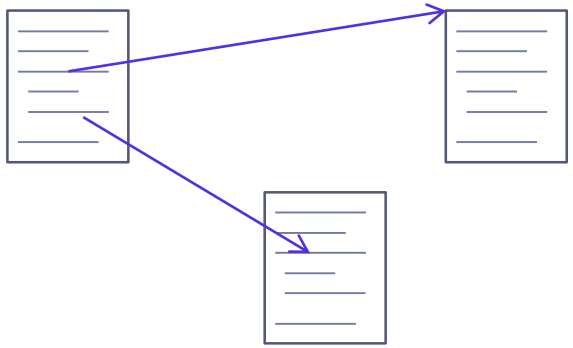
```
<h1>Chapter 1</h1>
head part of chapter 1
<h2>Section 1.1</h2>
content of section 1.1
<h2>Section 1.2</h2>
content of section 1.2
<h3>Subsection 1.2.1</h3>
subsection inside section 1.2
<h1>Chapter 2</h1>
head part of chapter 2
<h2>Section 2.1</h2>
content of section 2.1
```

- HTML5
  - nesting of section defines the level
  - can always use h1

```
<section>
<h1>Chapter 1</h1>
head part of chapter 1
  <section>
  <h1>Section 1.1</h1>
  content of section 1.1
  </section>
  <section>
  <h1>Section 1.2</h1>
  content of section 1.2
  <section>
    <h1>Subsection 1.2.1</h1>
    subsection of section 1.2
    </section>
  </section>
</section>
<section>
  <h1>Chapter 2</h1>
  head part of chapter 2
  <section>
    <h1>Section 2.1</h1>
    content of section 2.1
  </section>
</section>
```

Hyperlink of HTML

- HTML = Hypertext Markup Language
  - Hyperlink is the most important part of HTML.
- Hyperlink
  - Relate documents
  - Use URL (Uniform Resource Locator) to point other documents.
- Hyperlink elements
  - a
  - link
  - img
  - object



a element vs link element

- a element
  - Can only be used inside body
  - anchor element
  - Link a part of a document to another document.

```
<p><a href="http://www.keio.ac.jp">Keio University</a> is
the oldest university in Japan.</p>
```

- link element
  - Can only be used inside head
  - Link documents
  - Link stylesheets, javascripts, ...
  - rel for normal link
  - rev for reverse link

```
<head>
...other head information...
<title>Chapter 5</title>
<link rel="prev" href="chapter4.html" />
<link rel="next" href="chapter6.html" />
<link rel="stylesheet" href="book.css" />
</head>
```

Summary

- Web documents
  - paper documents vs online documents
  - structured documents
  - SGML: Standard General Markup Language
- HTML
  - SGML application
  - HTML structure
  - separation of content and presentation
  - hyperlinks
- Questions
  - Is HTML a good (perfect?) language for Web pages?
  - What is missing (or unnecessary) in HTML?