xml – html – css
structure and form

xml
– for communicating structured data
– general language for labelled trees
html
– an xml format for text markup
css
– a stylesheet language
– for presenting html

Friday, 1 October 2010

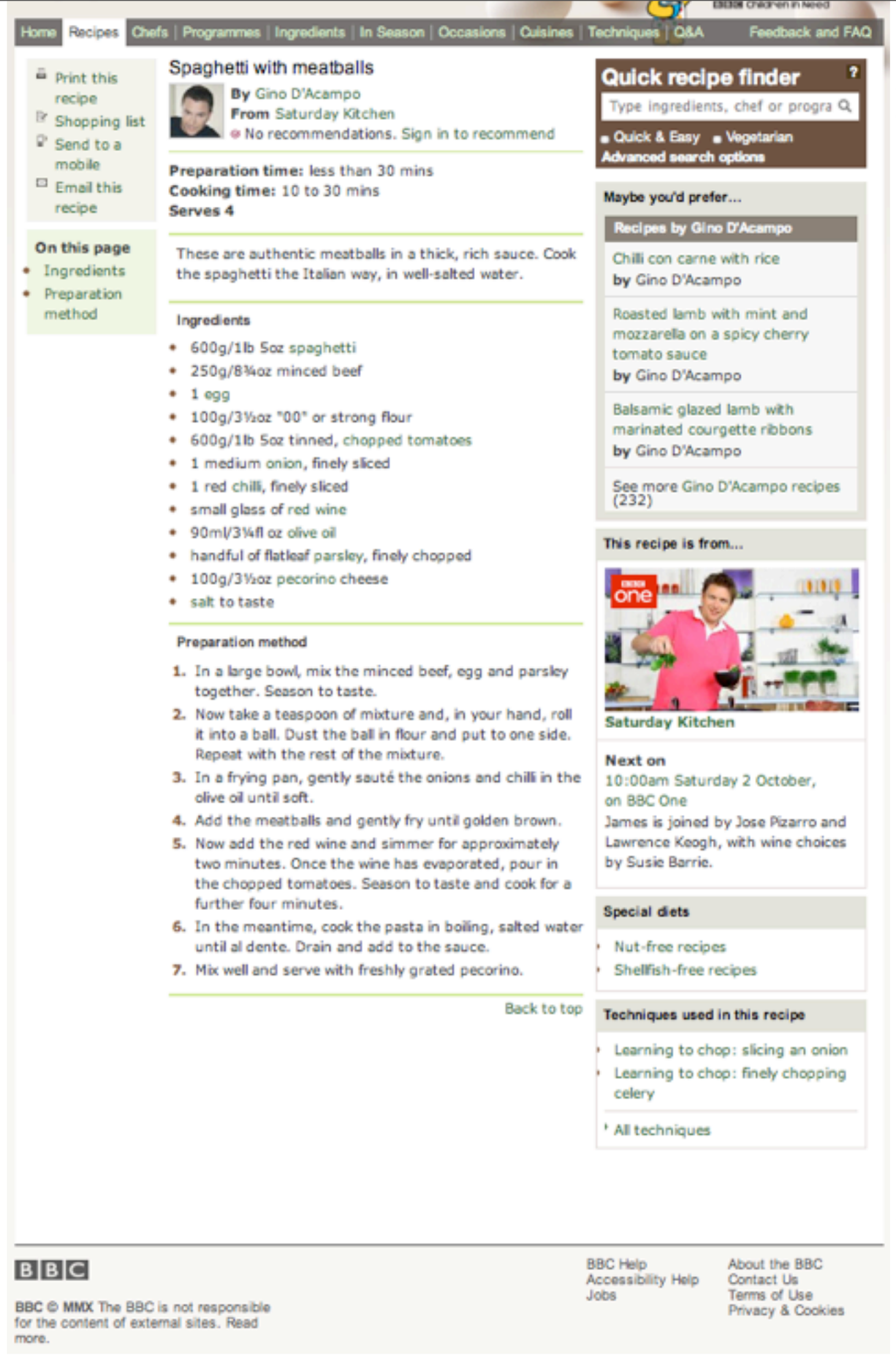
Computers use a variety of languages or protocols for communication.

We also use languages for communicating with computers.

XML is a general language for describing labelled trees.

HTML (hyper text markup language) is a language for describing structured documents – a special version of XML used for describing web pages.

CSS (cascading style sheets) is another language for describing the “style” used to present a document.




```
<?xml version="1.0" encoding="utf-8"?>
<collection>
  <recipe>
    <title>Spaghetti with meatballs</title>
    <description>
       A classic vegetarian dish.
    </description>

    <ingredient>spaghetti
        <grams>600</grams><ounces>21</ounces>
    </ingredient>
    <ingredient>minced beef
        <grams>250</grams><ounces>8.75</ounces>
    </ingredient>
    <ingredient>egg<number>1</number></ingredient>
    ...
    <method>
      <step>
        In a large bowl, mix the minced beef,
        egg and parsley together. Season to taste.
      </step>
      <step>
         Now take a teaspoon of mixture and, in your
         hand, roll it into a ball. Dust the ball in
         flour and put to one side. Repeat with the
         rest of the mixture.
      </step>
      ...
    </method>
    <time>
      <prep>less than 30 mins</prep>
      <cook>10 to 30 mins</cook>
    </time>
    <servings>4</servings>
    <image>
      <file>1.jpg</file><title>Beef Mince</title>
    </image>
    <by>Gino D'Acampo</by>
    <from>Saturday Kitchen</from>
  </recipe>
  ...
</collection>
```

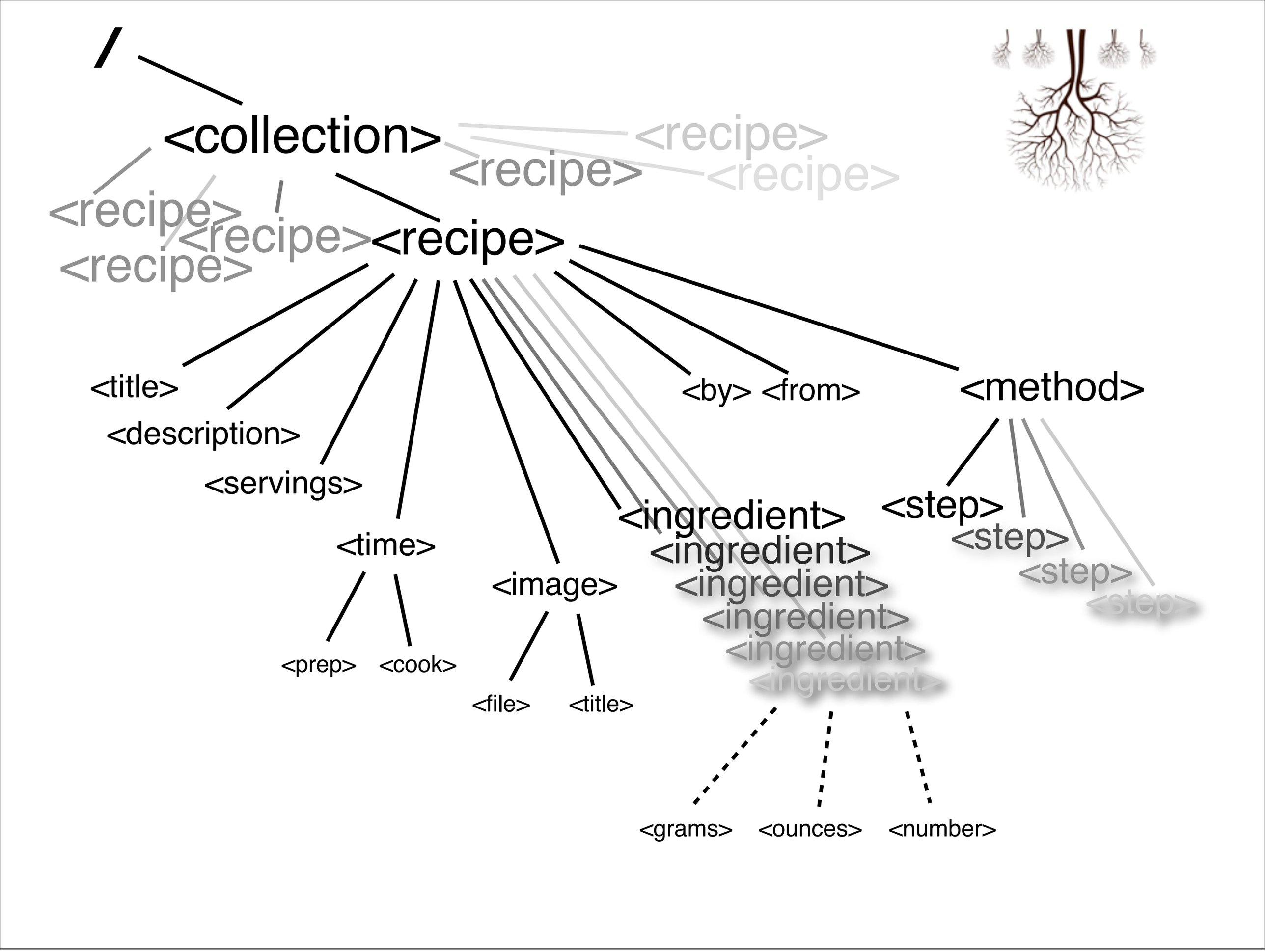
http://www.bbc.co.uk/food/recipes/database/spaghettiwithmeatbal_72227.shtml

Friday, 1 October 2010

We represent the information in the recipe using XML
The first line is just "housekeeping" – it says that we are using a particular version of xml and also specifies how the characters in this file are encoded as bits and bytes (we will talk more about utf-8 later).

Here, we don't use html tags – we just make up tags that make sense to us to describe the structure of the recipe.

XML is flexible markup language. You can use any set of tags that you want. You just need to match "opening" and "closing" tags.

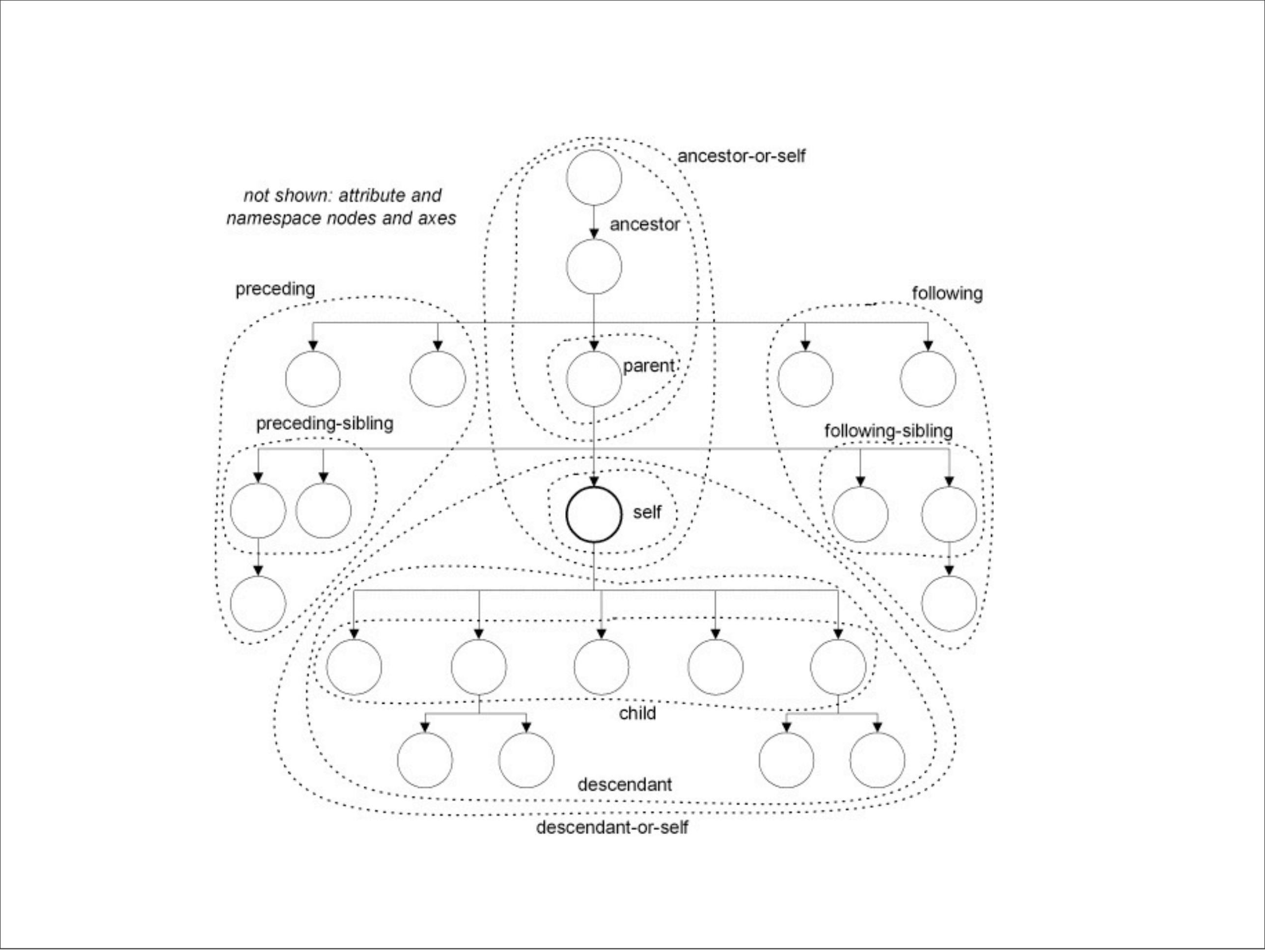


XML organises information in labelled trees. We draw them upside-down.

We will markup text with these structures using xml and html, control the appearance of web pages using css

Normally web pages are written in html, which uses a standard set of tags that browsers understand.

However, we can also write stylesheets for xml that tell the browser how to present tags that we have invented to describe the structure we are interested in.

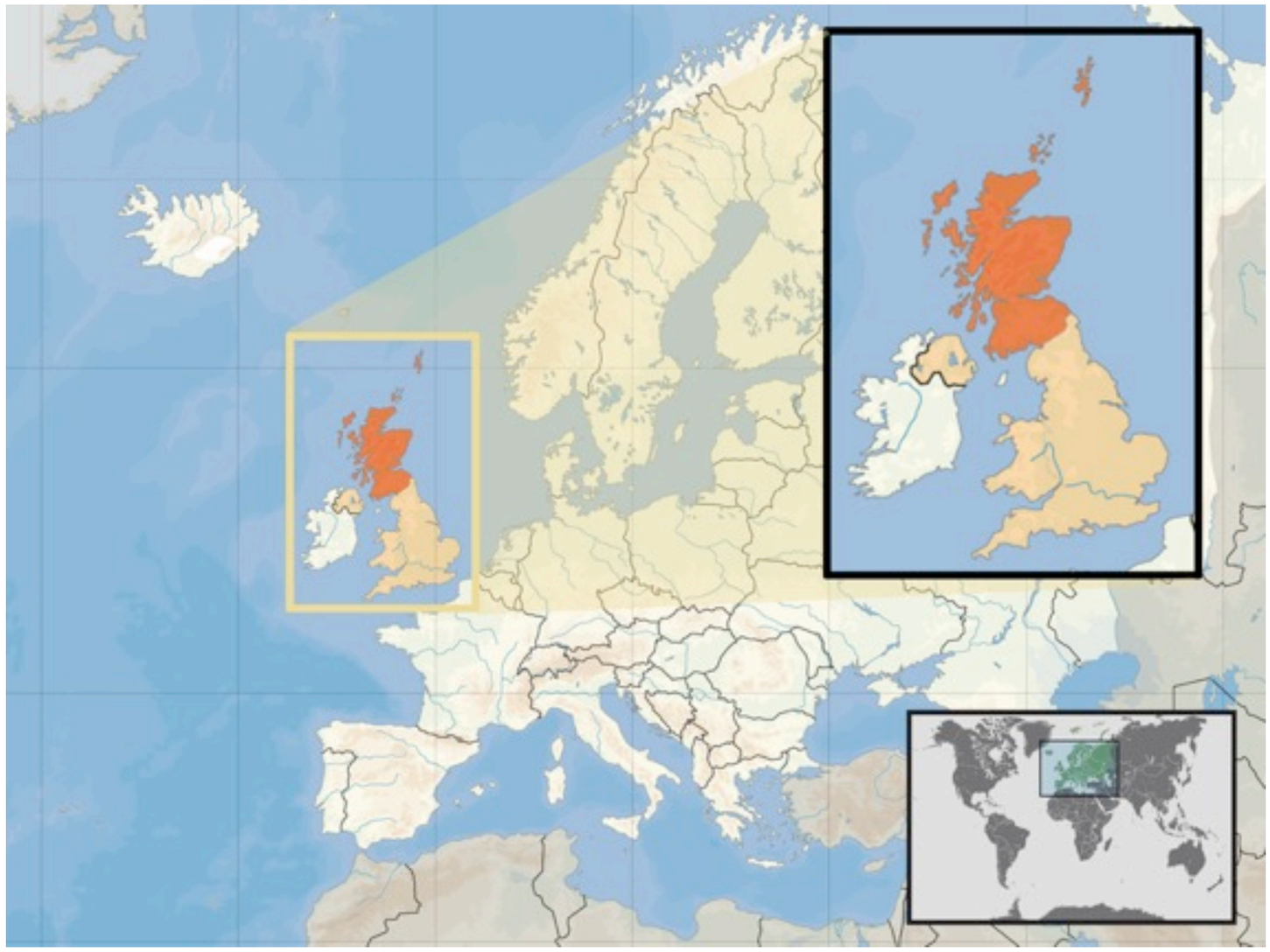


Friday, 1 October 2010

When we talk about trees, we often use the language of family trees – parent, children, sibling, descendant, ancestor.

```
<world>
  <europe>
    <fr>
      ...
    </fr>
    <de>
      ...
    </de>
    <uk>
      <eng>
        ...
      </eng>
      <ni>
        ...
      </ni>
      <scot>
        ...
      </scot>
      <wales>
        ...
      </wales>
    </uk>
  </europe>
    ...
</world>
```

nested hierarchy

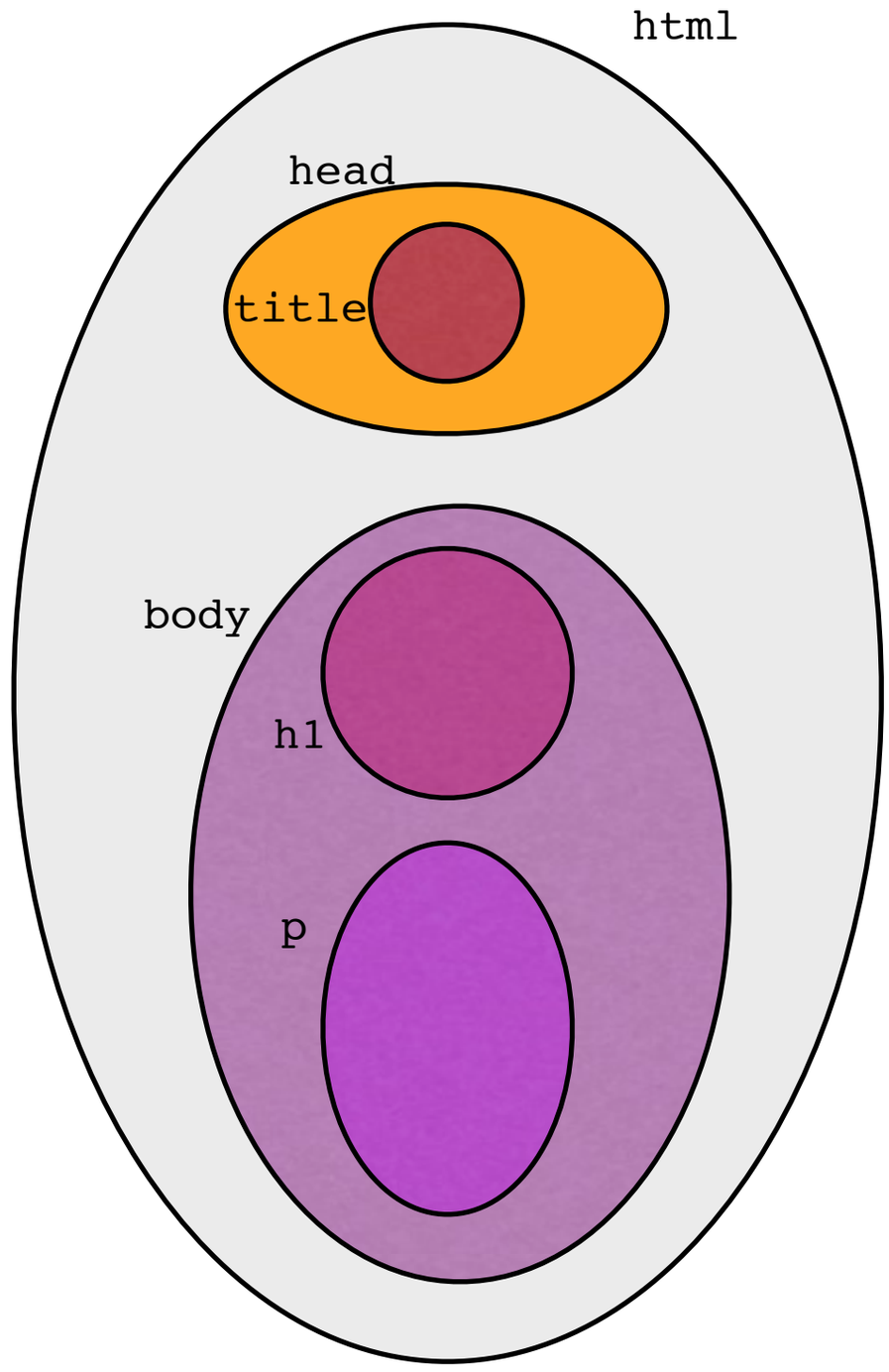


xml can be used to represent any nested hierarchy.

The various social and political divisions form a nested hierarchy.

html5

```
<!DOCTYPE html>
<html lang="en">
    <head>
        <title>Swapping Songs</title>
    </head>
    <body>
        <h1>Swapping Songs</h1>
        <p>Tonight I swapped some of
        the songs I wrote with some friends, who
        gave me some of the songs they wrote.
        I love sharing my music.</p>
    </body>
</html>
```



Swapping Songs

Tonight I swapped some of the songs I wrote with some friends, who gave me some of the songs they wrote. I love sharing my music.

Friday, 1 October 2010

Here is an html example (shown rendered as a “web page” below).

We can represent the structured hierarchy of the web page by a “map”. Note that the circles do not overlap – that’s what we mean by a nested hierarchy. The same structure is represented by the html tags being properly nested.

It is easyier to see whether your xml is properly nested is you use indentation to keep track of the depth of each tag in the tree.

some html5 tags

```
a section
<section> content </section>

headings h1 - h2 - ... - h6
<h1> content </h1>

an address
<address> streetaddress </address>

paragraph
<p> content </p>

blockquote
<blockquote> content </blockquote>

an image
<img src ="URL" alt="title"/>

line break
<br />
```

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE html>
<html>
    <head>
        <title> content </title>
    </head>
    <body>
        content
    </body>
</html>
```


html uses a special vocabulary of tags

Here are some common html tags.

For this course, we will be using xml with the html5 dialect for our web pages.
The second line is more housekeeping.

css

cascading style sheets

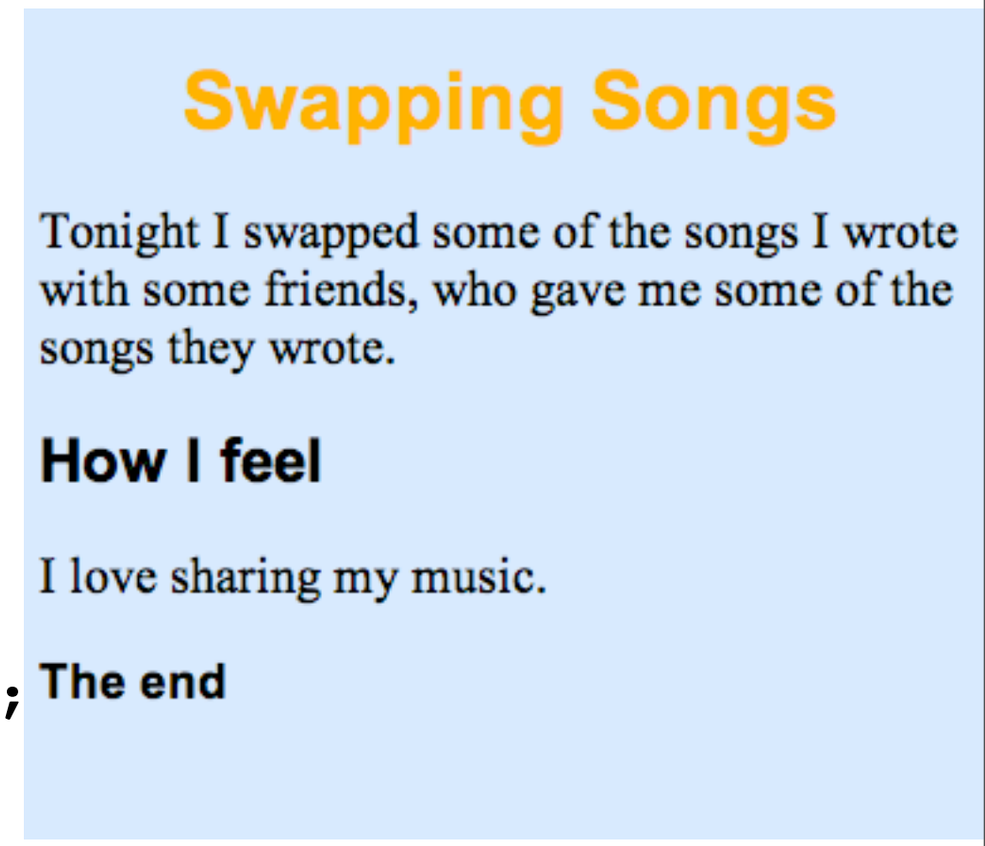
```
body{
     background-color:#d0e4fe;
}

h1{
     color:orange;
     text-align:center;
}

p{
     font-family:"Times New Roman";
     font-size:20px;
}
```

Friday, 1 October 2010

css is a stylesheet language for describing the presentation of an html file

html5 + css

```
<!DOCTYPE html>
<html lang="en">
    <head>
        <title>Swapping Songs</title>
        <style type="text/css">
            body{
                background-color:#d0e4fe;
            }
            h1{
                color:orange;
                text-align:center;
            }
            p{
                font-family:"Times New Roman";
                font-size:20px;
            }
        </style>
    </head>
    <body>
        <h1>Swapping Songs</h1>
        <p>Tonight I swapped some of
        the songs I wrote with some friends, who
        gave me some of the songs they wrote.</p>
        <h2>How I feel</h2>
        <p>I love sharing my music.</p>
        <h3>The end</h3>
    </body>
</html>
```



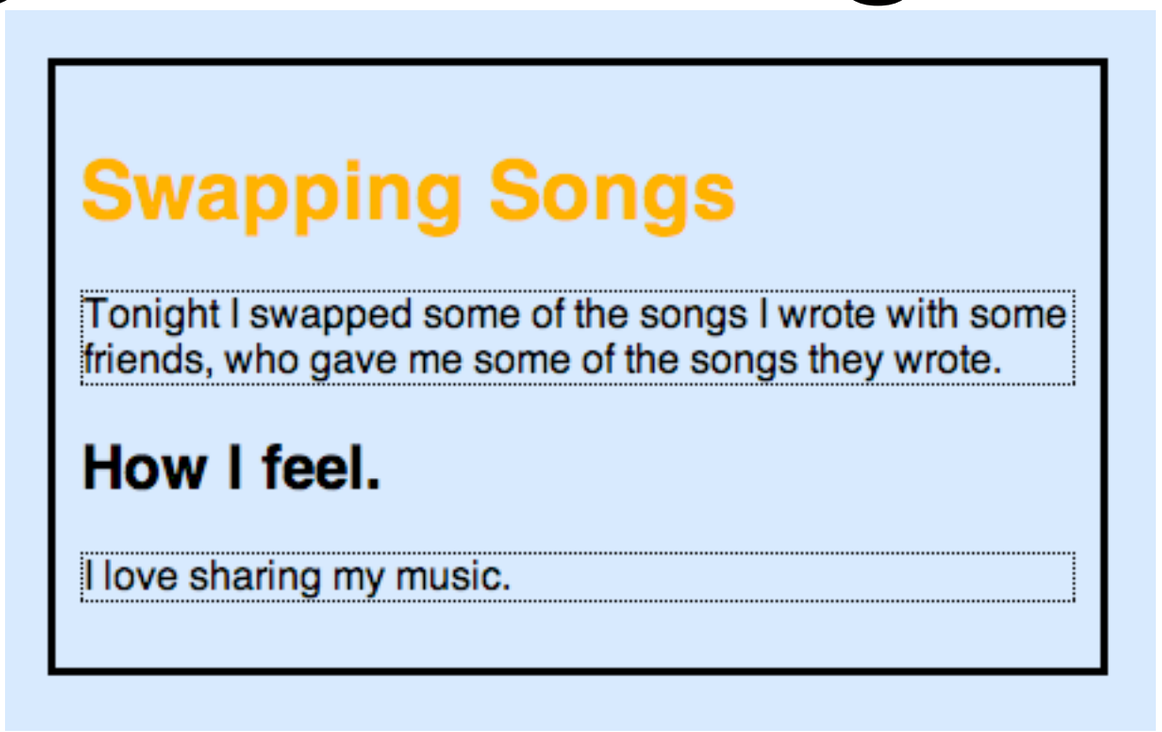
We can add a <style> node to our html tree to specify the presentation of our page.

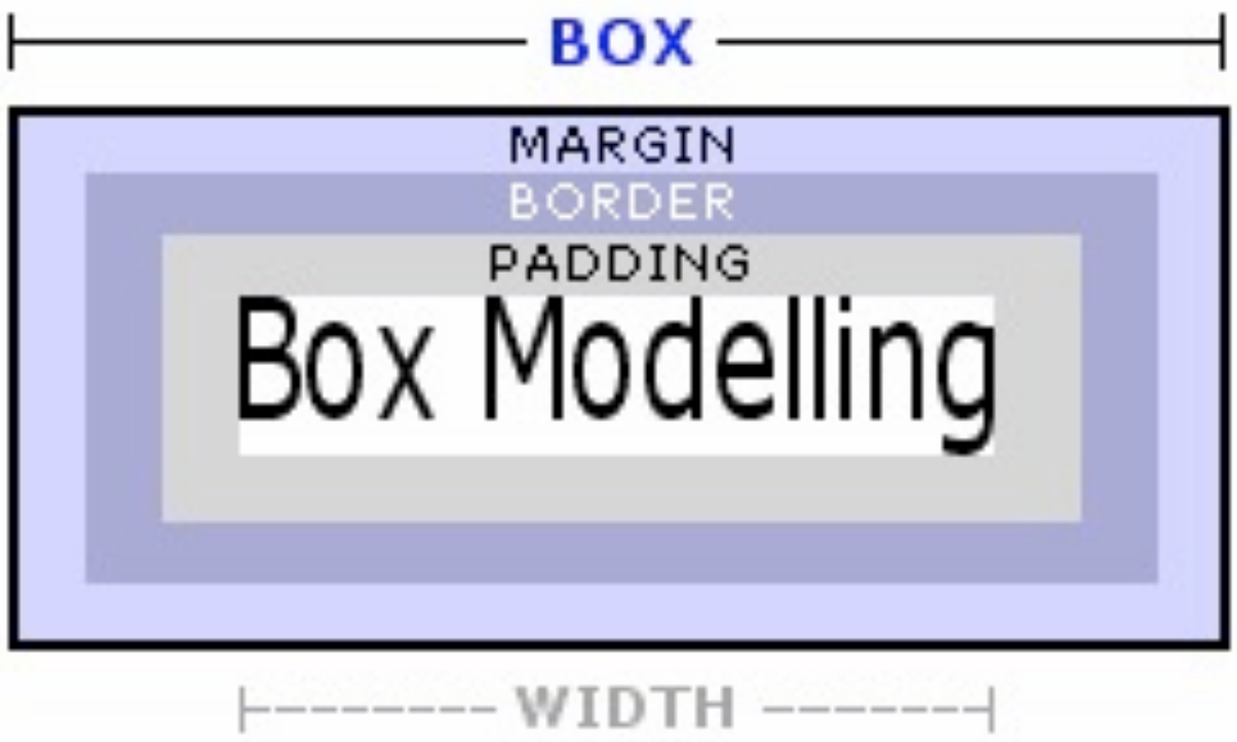
Colours (always spelled “color” in css) can be given by name or by number (more about the numbers later).

Size here is given in pixels (px) but it can also be given in points (pt), or centimetres (cm). Except for zero lengths, every length needs a unit.

CSS padding border margin

```
body {
      margin: 20px;
      padding: 10px;
      font-family: Helvetica, Arial, sans-serif;
      border-style: solid;
      background-color:#d0e4fe;
}
p {
    border-style:dotted;
    border-width:1px;
}
h1{
      color:orange;
}
h2 {
    font-size: 28px;
    line-height: 44px;
    padding: 22px;
}
h3 {
    font-size: 18px;
    line-height: 22px;
    padding: 11px ;
}
```





Friday, 1 October 2010

The layout of a web page uses a box model: each node occupies a box around which we can add any or all of padding, a border, and a margin.

tables

```
<table>
  <caption>Some cookery books</caption>
  <thead>
    <tr>
      <th>isbn</th> <th>title</th>
      <th>author</th> <th>pubID</th> <th>pages</th>
    </tr>
  </thead>
  <tr>
    <td>029785593X</td> <td>From Nature To Plate </td>
    <td>Tom Kitchin</td> <td>7642</td> <td>272</td>
  </tr>
  <tr>
    <td>955904609</td><td> Cookbook</td>
    <td> Martin Wishart </td><td>3556</td><td>256</td>
  </tr>
  <tr class="centre">
    <td>...</td><td>...</td>
    <td>...</td><td>...</td>
    <td>...</td>
  </tr>
</table>
```

Some cookery books

isbn	title	author	pubID	pages
029785593X	From Nature To Plate	Tom Kitchin	7642	272
955904609	Cookbook	Martin Wishart	3556	256
...	...	...	...	...

Tables are a common way of presenting information.

HTML markup for tables makes rows more important than columns.

Wine calories table:

Wine	Wine Calories 115ml	Large Glass
Alcohol-free Wine	37 calories	74 Cals
Champagne	96 calories	192 Cals
Dry Red Wine	83 calories	166 Cals
Dry White Wine	77 calories	154 Cals
Rose	82 calories	164 Cals
Sparkling	92 calories	184 Cals
Sweet Red Wine	100 calories	200 Cals
Sweet White Wine	103 calories	206 Cals
Fortified Wines	**Wine Calories**	**Large Glass**
Bianco Vermouth	167 calories	334 Cals
Ginger Wine	190 calories	380 Cals
Martini Bianco	150 calories	300 Cals
Martini Extra Dry	150 calories	300 Cals
Martini Rose	180 calories	360 Cals
Martini Rosso	192 calories	384 Cals
Port	170 calories	340 Cals
Sherry average	140 calories	280 Cals

All values correct at time of testing, values for wine calories may vary between different sized glasses!

http://www.weightlossforall.com/

Friday, 1 October 2010

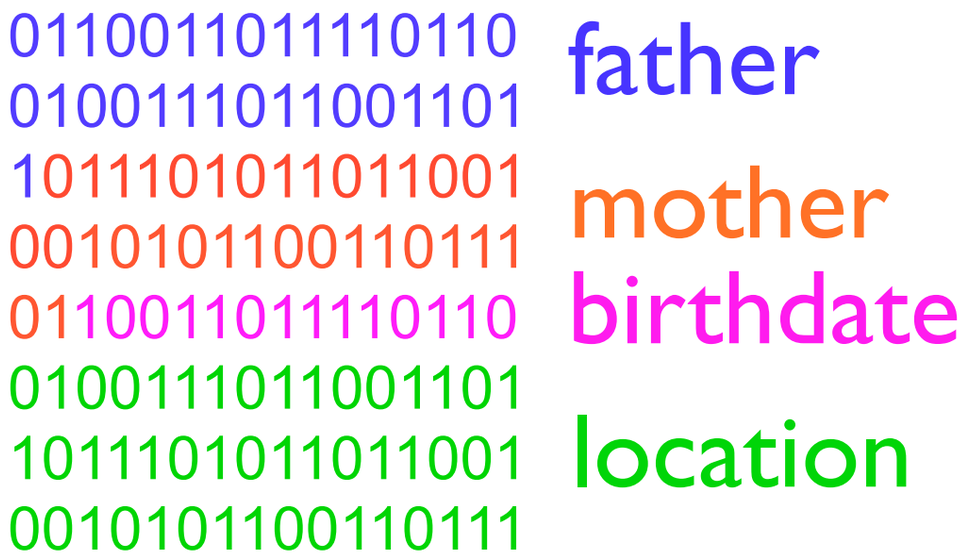
Here is an example

BREADS & CEREALS	Portion size *	per 100 grams (3.5 oz)	energy content
Bagel (1 average)	140 cals (45g)	310 cals	Medium
Biscuit digestives	86 cals (per biscuit)	480 cals	High
Jaffa cake	48 cals (per biscuit)	370 cals	Med-High
Bread white (thick slice)	96 cals (1 slice 40g)	240 cals	Medium
Bread wholemeal (thick)	88 cals (1 slice 40g)	220 cals	Low-med
Chapatis	250 cals	300 cals	Medium
Cornflakes	130 cals (35g)	370 cals	Med-High
Crackerbread	17 cals per slice	325 cals	Low Calories
Cream crackers	35 cals (per cracker)	440 cals	Low / portion
Crumpets	93 cals (per crumpet)	198 cals	Low-Med
Flapjacks basic fruit mix	320 cals	500 cals	High
Macaroni (boiled)	238 cals (250g)	95 cals	Low calorie
Muesli	195 cals (50g)	390 cals	Med-high
Naan bread (normal)	300 cals (small plate size)	320 cals	Medium
Noodles (boiled)	175 cals (250g)	70 cals	Low calorie
Pasta (normal boiled)	330 cals (300g)	110 cals	Low calorie
Pasta (wholemeal boiled)	315 cals (300g)	105 cals	Low calorie
Porridge oats (with water)	193 cals (350g)	55 cals	Low calorie
Potatoes** (boiled)	210 cals (300g)	70 cals	Low calorie
Potatoes** (roast)	420 cals (300g)	140 cals	Medium
Rice (white boiled)	420 cals (300g)	140 cals	Low calorie
Rice (egg-fried)	500 cals	200 cals	High in portion
Rice (Brown)	405 cals (300g)	135 cals	Low calorie
Rice cakes	28 Cals = 1 slice	373 Cals	Medium
Ryvita Multi grain	37 Cals per slice	331 Cals	Medium
Ryvita + seed & Oats	180 Cals 4 slices	362 Cals	Medium
Spaghetti (boiled)	303 cals (300g)	101 cals	Low calorie

Friday, 1 October 2010

and another

information is structured data



Information is stored as bits, but we normally think about how it is organised at a higher level.

The organisation of data is one way of representing knowledge.

Relational Data

A **table** of **data**. Each **entry** in a **cell** is a **datum**

Books

isbn	title	author	pubID	pages
029785593X	From Nature To Plate	Tom Kitchin	7642	272
955904609	Cookbook	Martin Wishart	3556	256
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

When we think of tables as a way of organising data then rows and colums are equally important – but different.

Relational Data

Rows represent the things we're interested in.

We call each row a **record**.

Books

isbn	title	author	pubID	pages
029785593X	From Nature To Plate	Tom Kitchin	7642	272
955904609	Cookbook	Martin Wishart	3556	256
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

A **table** of similar records is called a **relation**.

Each row, or record, represents an item (in this example, a book).

Relational Data

Columns represent properties or attributes.
Each of these is a **field**.

Books

isbn	title	author	pubID	pages
029785593X	From Nature To Plate	Tom Kitchin	7642	272
955904609	Cookbook	Martin Wishart	3556	256
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Each record in the relation has the same format.

Each column represents a property of the item, or ‘field’.

We have the same fields (the same information) for every item.

Relational data

Publishers

ID	name	address
7,642	Weidenfeld & Nicolson	London
3,556	Mr Max Publishing	Edinburgh
...	...	...
...	...	...
...	...	...

A typical database has many relations.
An ID or **key** field uniquely identifies a record.

Friday, 1 October 2010

Some fields have the special property that the value in that field uniquely identifies the item. We call this a key.

The isbn field is a key for the Books table
The ID field is a key for the Publishers table

We can use a key field to let one table refer to (index into) another table.

Relational data

Publishers

ID	name	address
7,642	Weidenfeld & Nicolson	London
3,556	Mr Max Publishing	Edinburgh
...	...	...

Books

isbn	title	author	pubID	pages
029785593X	From Nature To Plate	Tom Kitchin	7642	272
955904609	Cookbook	Martin Wishart	3556	256
...	...	...	...	...

A typical database has many relations.

An ID or **key** field uniquely identifies a record.

Some fields have the special property that the value in that field uniquely identifies the item. We call this a key.

The isbn field is a key for the Books table
The ID field is a key for the Publishers table

We can use a key field to let one table refer to (index into) another table.

Relational data

- field
  - a property or attribute
- record
  - values for each field, for a given item
- relation or table
  - set of records, representing a set of items
- key
  - a field that uniquely identifies an item

Check that you understand the terminology.

Relational data

- field
  - a property or attribute
- record
  - values for each field, for a given item
- relation or table
  - set of records, representing a set of items
- key
  - a field that uniquely identifies an item

Check that you understand the terminology.

Relational data

- field
  - a property or attribute
- record
  - values for each field, for a given item

- relation or table
  - set of records, representing a set of items
- key
  - a field that uniquely identifies an item

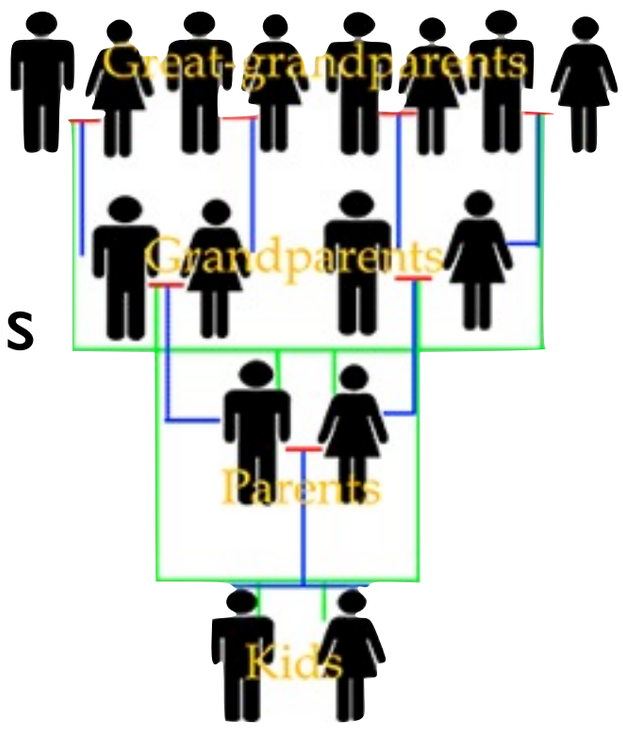
Check that you understand the terminology.

Relational data

- field
  - a property or attribute
- record
  - values for each field, for a given item

- relation or table
  - set of records, representing a set of items

- key
  - a field that uniquely identifies an item



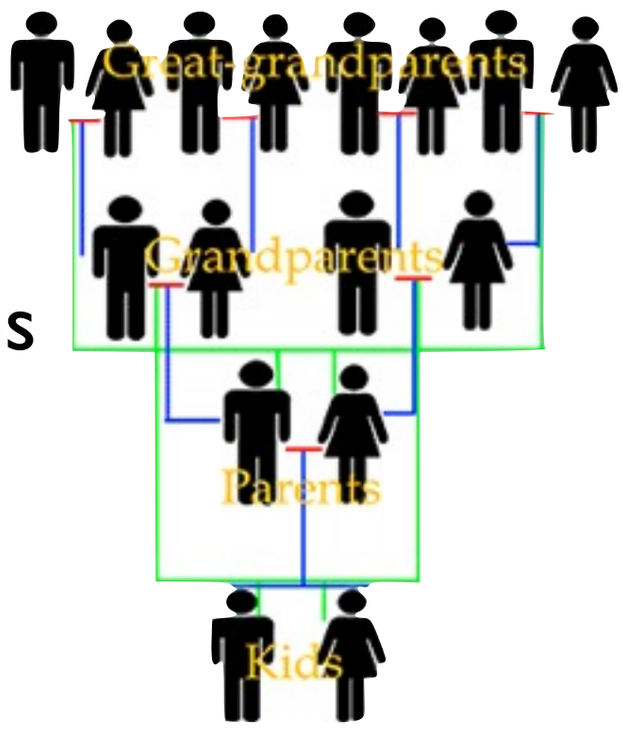
Check that you understand the terminology.

Relational data

- field
  - a property or attribute
- record
  - values for each field, for a given item

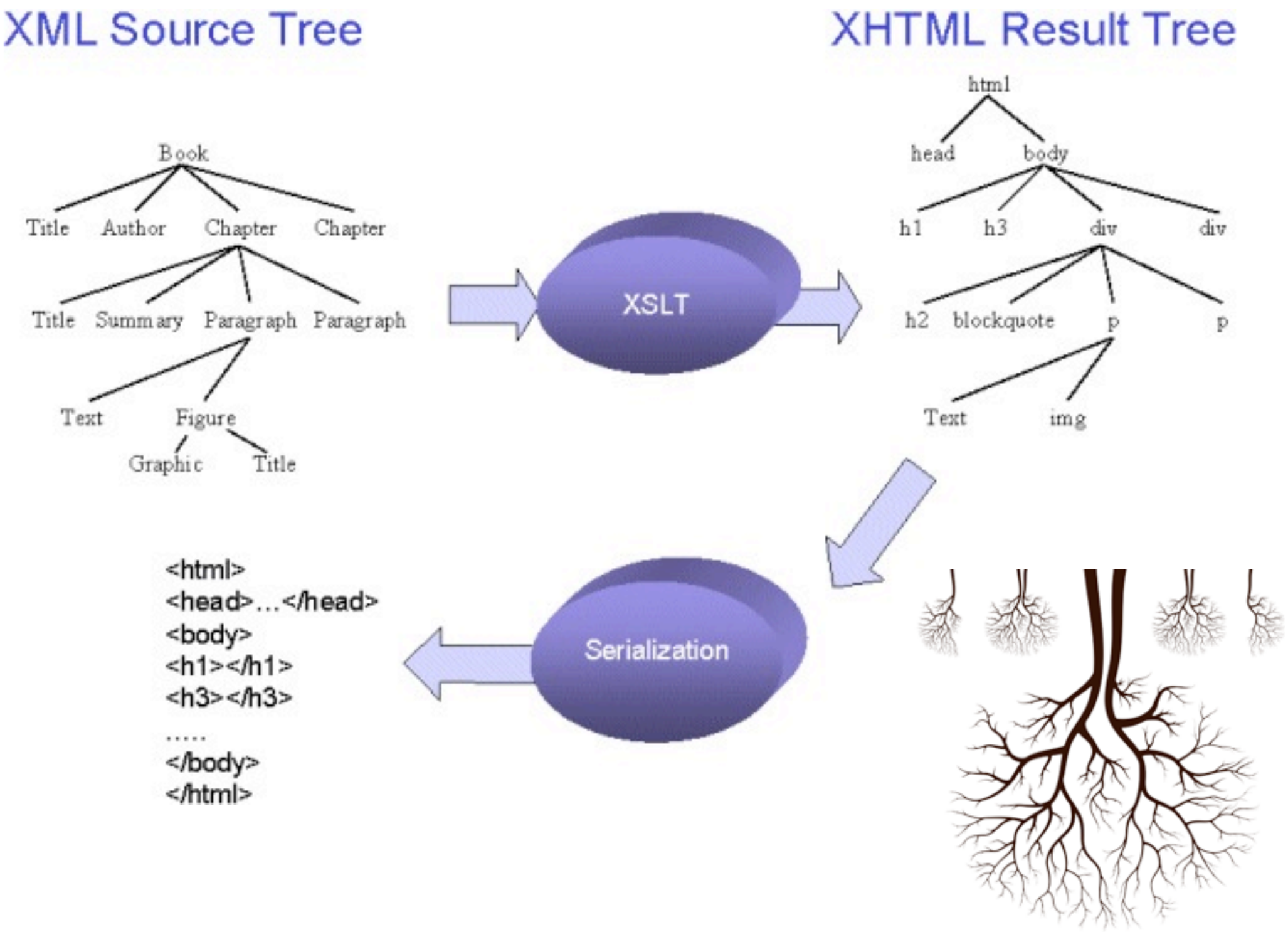
- relation or table
  - set of records, representing a set of items

- key
  - a field that uniquely identifies an item



Check that you understand the terminology.

http://nwalsh.com/docs/tutorials/xsl/xsl/slides.html



Friday, 1 October 2010

We started with a semantic view of a recipe, but then turned to html, then tables. Different ways of presenting information.

One of the things we'll do is to transform information from one representation into another.

For example, XSL lets us construct new trees out of old trees.

The basic idea is that we use xsl to extract parts of a tree and build a new tree.

We can use this to turn semantic xml into html, or to extract data from an xml or html page