Information Management Resource Kit

Module on Management of Electronic Documents

UNIT 2. FORMATS FOR ELECTRONIC DOCUMENTS AND IMAGES

LESSON 3. PRESENTATIONAL MARK-UP: HTML

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.

Objectives

At the end of this lesson, you will be able to:

- understand the main **features of HTML**, and
- create a simple **HTML document**.

What is HTML?

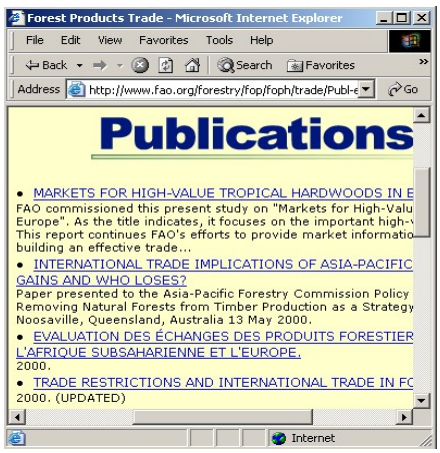
Tim Berners-Lee

The **Hypertext Mark-up Language** (HTML) is a mark-up language designed for the presentation of information on the **World Wide Web**, using a web browser.

HTML evolved from the need to share documents on the Internet and render them meaningfully on different browser platforms.

It was originally created by **Tim Berners-Lee** when he was working on the first concepts of the Web at CERN in the late 1980s.

What is HTML?

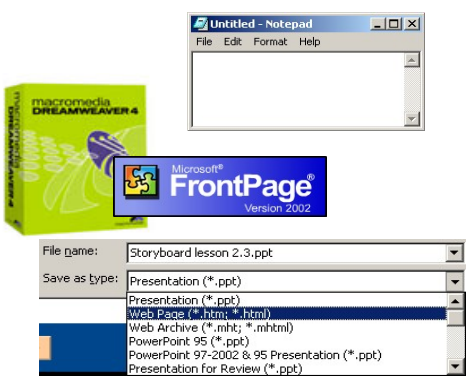


HTML contains information that defines:

- **basic presentation** of a document (headers, paragraphs, lists and tables),
- **hyperlinks**, and
- **multimedia** information.

Using HTML, you have the basic mark-up to create the documents you want to publish on the Web.

What do you need?



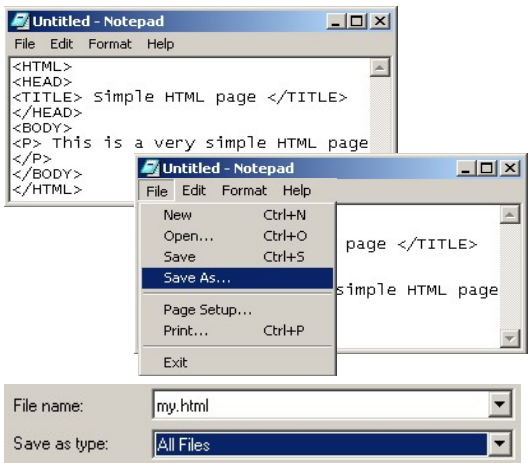
What do you need to create an HTML document?

Simple HTML documents can be created easily using **any text editor**.

There are also many **HTML authoring** packages available for creating more complex pages or complete web sites.

Many applications and software packages can also generate HTML documents, either using a '**Save As HTML**' feature or by exporting information as HTML.

How to create an HTML document



Let's now consider a simple **text editor**, such as Notepad.

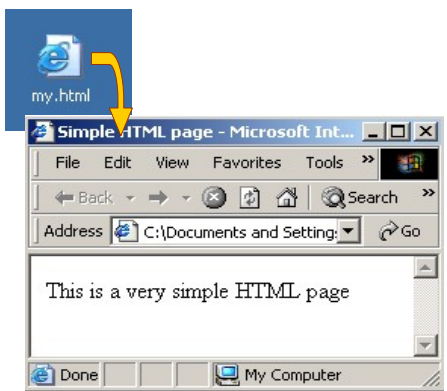
It can be used to write the content in HTML, which makes it very easy to visualize the content in a web page.

You just have to:

• select **Save As** from the **File** menu, and

• select **All files** from the **Save as type** drop down list and **name** the file using the HTML format: e.g. "my.html".

How to create an HTML document

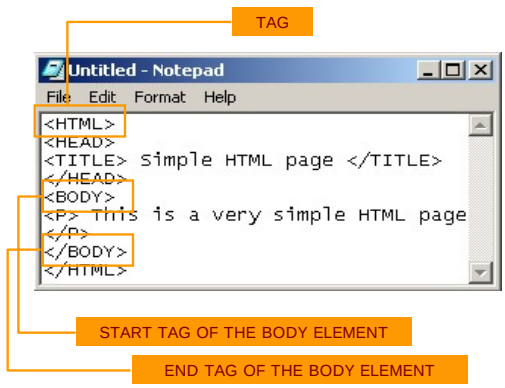


You can save the .html file on your computer.

In this example, if we click on **my.html**, we can see the result of what we just created.

Now that we have seen how to create an HTML page, let's try to understand how to write in HTML.

How to do - Basics



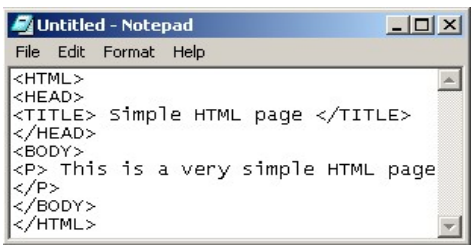
Considering the same example, you can see that the mark-up in an HTML document consists of **tags** which are delimited by opening and closing **angle brackets < >**.

Tags represent elements in the document that will be displayed in a web browser.

The name of the element appears in the **start tag** and in the matching **end tag**, where it has an additional forward slash '/' in front of it.

Attention should be given to the **tag nesting**: for example, the tag </TITLE> is before the tag </HEAD>, as the TITLE element is contained in the HEAD element.

How to do - Basics



Elements which have start and end tags can contain either text, other elements or a mixture of text and elements.

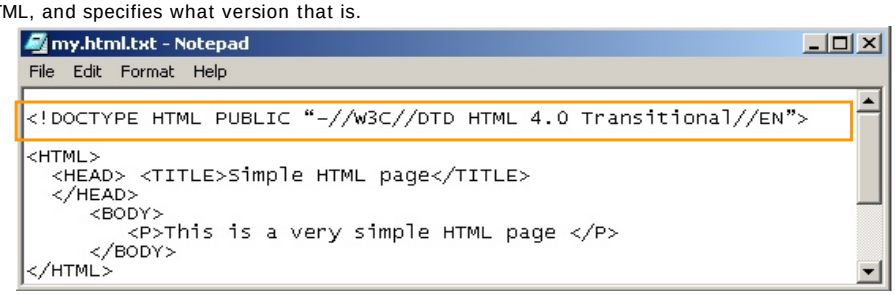
In this example, could you determine the relationship between the various elements?

- The <BODY> element contains a <HEAD> and a <HTML> element.
- The <HTML> element contains a <HEAD> and a <BODY> element.
- The <HEAD> element contains a <HTML> and a <BODY> element.

Click on the answer of your choice

How to do - Basics

As you can see in the example, another piece has been added at the very top: the **document type declaration (DTD)**. This displays nothing on screen, it tells the browser what version of HTML you are writing in.
More specifically, **<!DOCTYPE ...>** declares that this document conforms to a specific version of HTML, and specifies what version that is.

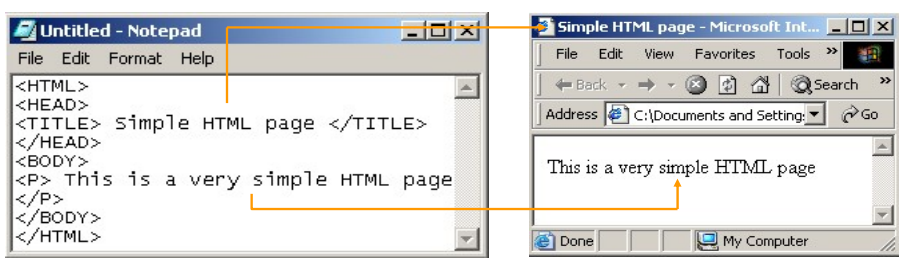


DTD is not a requirement, but it should be included at the top of every web document to be consistent with standards published by W3C (the World Wide Web Consortium). On the W3C website you can find information on:

how to write the document type declaration
(www.w3.org/TR/REC-html40/struct/global.html#h-7.2)

How to do - Basics

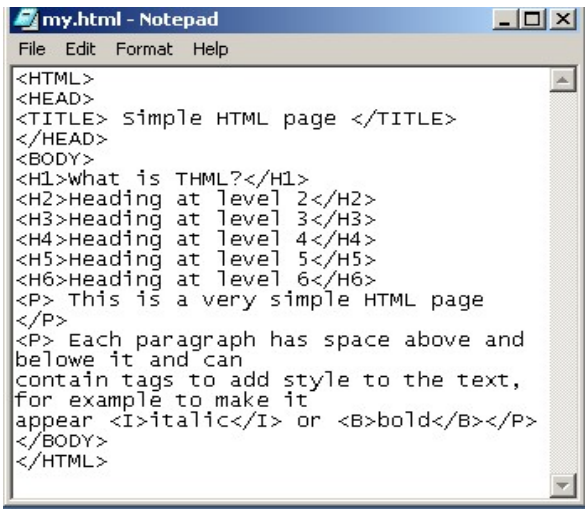
The <HEAD> element contains a **<TITLE>** which is not displayed in the main text of the document but is used to display the title in the top border of the browser window. The **<BODY>** element contains the main content of the HTML document, which is displayed in the main web browser window.



In our example the <BODY> only contains a short paragraph of text inside a <P> element.

Just a tip before proceeding: HTML elements (for HTML 4) are **not case sensitive**, so something scripted as <hTmL> will work just as <html> or <HTML> would. However, using the same case across documents is good practice.

How to do – Simple Layout



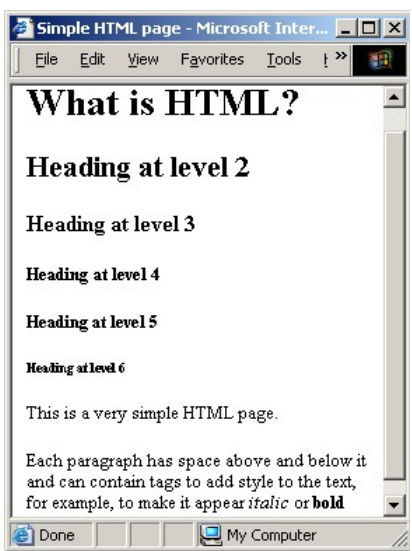
Basic HTML mark-up is used to lay out text in the <BODY> of the page with **headers**, **paragraphs of text** and some **simple formatting** of text within paragraphs. There are **six levels** of heading denoted by the mark-up **<H1> to <H6>**. The text (title) to appear in the header is placed between the opening and closing tags of the header.

Our example also shows the use of **<I>** and **<B>** tags to make the browser render text in *italic* or **bold typeface**.

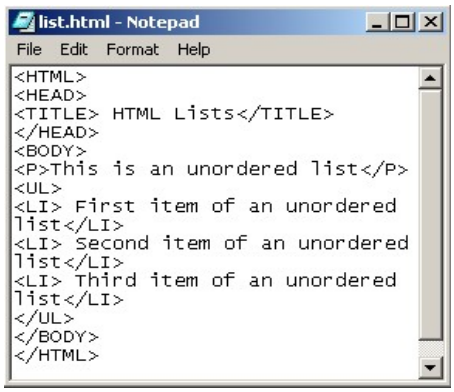
Click here to see the results as an HTML page.

How to do – Simple Layout

RESULT



How to do – Lists



Click here to see the results as an HTML page.

The two most common types of lists are unordered and ordered lists.

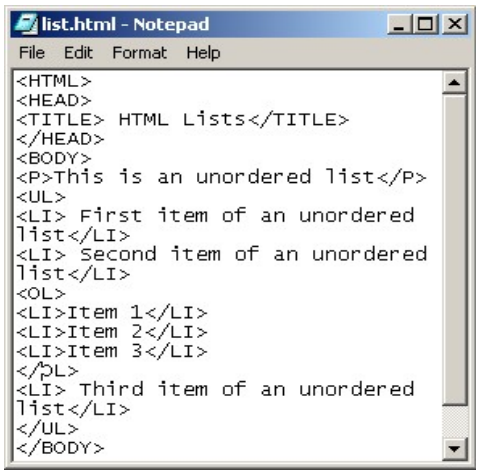
UNORDERED LISTS

They are denoted by the **<UL>** element. Each item in the list is contained in an **<LI>** element.
When displayed in the browser the <LI> elements are laid out one above the other in the list, with a bullet character in front of each one (we can also change the character displayed, for example to a square).
The <LI> element can contain text and almost any of the other formatting and layout elements available in HTML, including other lists.

How to do – Lists

RESULT

How to do – Lists



Click here to see the results as an HTML page.

In this example the second item of the unordered list contains an ordered list.

ORDERED LISTS

They are denoted by the **<OL>** element. The ordered list can contain as many or as few items as we like. When an ordered list is displayed in the browser, each item is laid out prefixed with a number that shows its position in the list.

Roman numbers are used by default, but we can alter the style of the numbering.

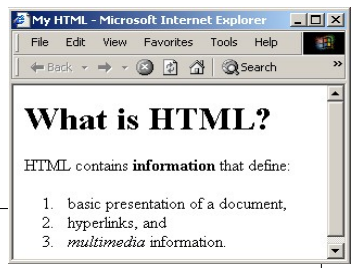
How to do – Lists

RESULT

How to do – Lists

Can you complete the HTML code for this page?

Complete the HTML code by typing the correct characters in the empty fields. Then click on the Confirm button.

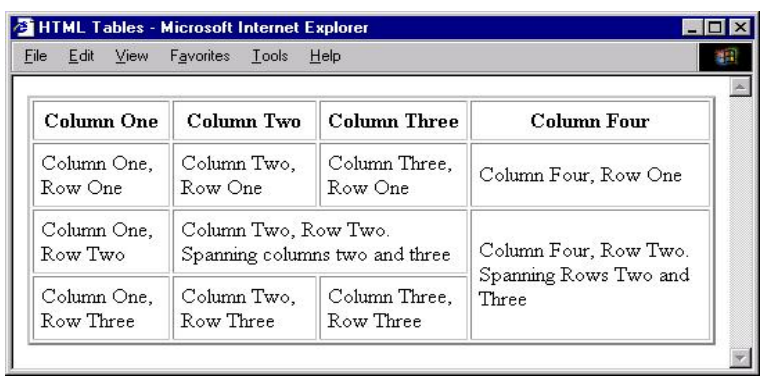



```
<HTML>
        <HEAD><----->My HTML</-----></---->
        <BODY>
      <-->What is HTML?</-->
                <P>HTML contains <->information</-> that defines:</P>
                <-->
                        <LI>basic presentation of a document,</LI>
                        <LI>hyperlinks, and</LI>
                        <LI><->multimedia</-> information.</LI>
                 </-->

        </---->
</HTML>
```

How to do – Tables

Let's have a look at this table. There are columns, rows, and borders of specific widths. How do we create this table in HTML?



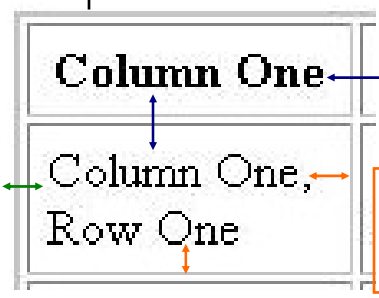
Column One	Column Two	Column Three	Column Four
Column One, Row One	Column Two, Row One	Column Three, Row One	Column Four, Row One
Column One, Row Two	Column Two, Row Two. Spanning columns two and three		Column Four, Row Two. Spanning Rows Two and Three
Column One, Row Three	Column Two, Row Three	Column Three, Row Three	

How to do – Tables

Tables are marked up in HTML using the **<TABLE>** element.

This element can have a number of **attributes** which are used to control how the table is displayed:

RULES determines which sides of the cells in the table will have a line drawn along its border (in our case, all of them)



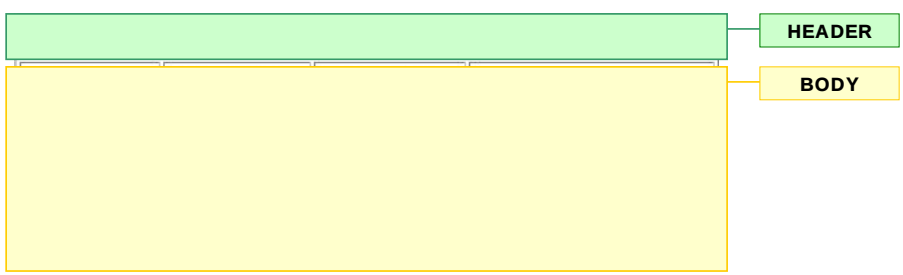
CELL SPACING sets the distance between the cells in the table

BORDER sets the width of the border around the outside of the table

CELL PADDING sets the space between the border of the cell and its content

How to do – Tables

The <TABLE> element can contain a header, a body and a footer. These are denoted by the elements <**THEAD**>, <**TBODY**> and <**TFOOT**>. Our example uses the <THEAD> and <TBODY> tags.



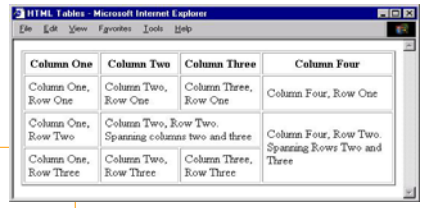
If we wanted to include information at the foot of our table (information which was repeated even when the table broke across multiple pages when printed) we could use a <TFOOT> element as well.

How to do – Tables

Now, let's look at the HTML code of our table.

```
<HTML>
  <HEAD><TITLE>HTML Tables</TITLE></HEAD>
  <BODY>
    <TABLE BORDER="2" RULES="ALL" CELLSPACING="2" CELLPADDING="5">
      <THEAD>
        <TR><TH>Column One</TH><TH>Column Two</TH>
            <TH>Column Three</TH><TH>Column Four</TH></TR>
      </THEAD>
      <TBODY>
        <TR><TD>Column One, Row One</TD><TD>Column Two, Row One</TD>
            <TD>Column Three, Row One</TD><TD>Column Four, Row One</TD></TR>
        <TR><TD>Column One, Row Two</TD>
            <TD COLSPAN="2">Column Two, Row Two. Spanning columns two and three</TD>
            <TD ROWSPAN="2">Column Four, Row Two. Spanning Rows Two and Three</TD></TR>
        <TR><TD>Column One, Row Three</TD><TD>Column Two, Row Three</TD>
            <TD>Column Three, Row Three</TD></TR>
      </TBODY>
    </TABLE>
  </BODY>
</HTML>
```

Click here to review the results as an HTML page.



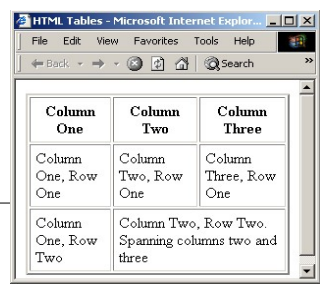
The <THEAD> elements can contain a number of rows, denoted by the <**TR**> element. Each row contains a set of cells denoted by <**TH**> elements.

The value of the **COLSPAN** attribute tells the browser the **number of columns** we want the cell to span (in our case, two). The value of the **ROWSPAN** attribute on <TD> tells the browser the **number of rows** we want the cell to span.

How to do – Tables

Can you complete the HTML code for this table?

Complete the HTML code by typing the correct characters in the empty fields. Then, click on the Confirm button.

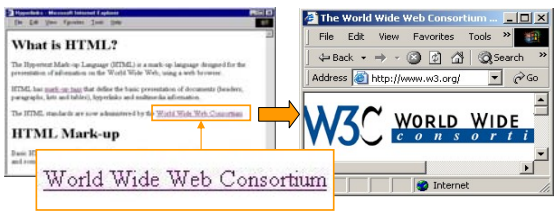



```
<HTML>
 <HEAD><TITLE>HTML Tables</TITLE></HEAD>
  <BODY>
   <TABLE BORDER="2" RULES="ALL" CELLSPACING="2" CELLPADDING="5">
     <----->
       <TR><--><TH>Column One</TH><TH>Column Two</TH><TH>Column Three</TH></TR>
     </THEAD>
        <TBODY>
             <TR><TD>Column One, Row One</TD>
                 <TD>Column Two, Row One</TD>
                  <TD>Column Three, Row One</TD>
           </-->
               <TR><TD>Column One, Row Two</TD>
                   <TD COLSPAN="2">Column ---, Row ---. Spanning Columns Two and Three</TD>
                </TR>
     </----->
    </TABLE>
   </BODY>
</HTML>
```

How to do – Hyperlinks

An important feature of HTML and web browsers is the ability to navigate information using **hyperlinks**.

A hyperlink is a link between a source location in an electronic document and one or more target locations, either in the same document or in another document. In a typical hypertext system, the source of the hyperlink is displayed as a 'hotspot' on the screen, where a user can click to move to the document location specified as the target of the link.



In this example, "World Wide Web Consortium" is the text **displayed** in the browser **as a link** that can be clicked on with your mouse.

By clicking on this link, you will have access to the home page of the World Wide Web Consortium web site.

How to do – Hyperlinks

To define hyperlinks with and between documents, HTML uses a single element, **<A>**. Actually, the <A> element is used in two different roles: both as the **starting point** of a hyperlink and as the **anchor point** in a document which can be targeted by other links.

Let's have a look at these examples:

<A HREF =http://w3.org> text</A>

<H1><A NAME="S1"> text</A></H1>

<A HREF ="#S2"> text</A>

When we click on the link the browser uses the value of the **HREF** attribute to find the target of the link.

The value used in the HREF attribute is actually a Uniform Resource Locator (**URL**) which specifies the **address of any resource** on the Internet.

Click on each example to view the explanation.

How to do – Hyperlinks

<A HREF =http://w3.org> text</A>

<H1><A NAME="S1"> text</A></H1>

<A HREF ="#S2"> text</A>

Here we have put <A> elements inside each of the <H1> tags so our text can be the **target of a hyperlink**. The NAME attribute is used to define **an identifier for the anchor**. It's important to make the value of each NAME attribute **unique** in your document, otherwise the browser won't know which anchor is being referenced.

<A HREF =http://w3.org> text</A>

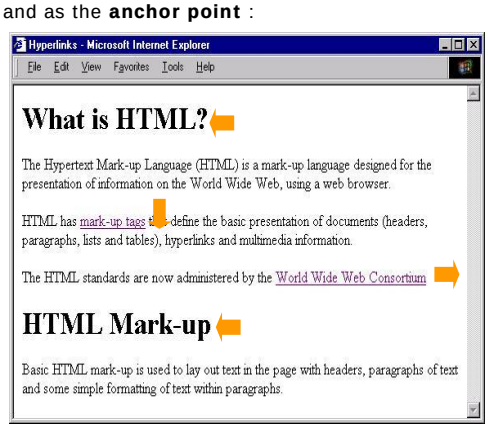
<H1><A NAME="S1"> text</A></H1>

<A HREF ="#S2"> text</A>

Here the value '#S2' indicates that the link points to an anchor **in the same document** (that's what the # means) with the value 'S2' for its NAME attribute. When we click on the link in the browser, the window scrolls to the position of the target anchor (in our example, the header titled 'HTML Mark-up').

How to do – Hyperlinks

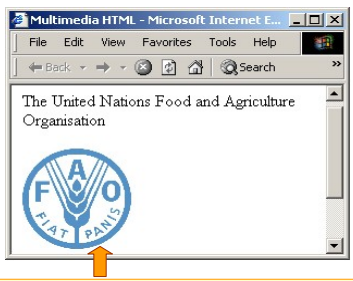
In this web page, for example, the <A> element is used both as the **starting point** of a hyperlink and as the **anchor point** :




```
<HTML>
  <HEAD><TITLE>Hyperlinks</TITLE></HEAD>
  <BODY>
    <H1><A NAME="S1">What is HTML?</A></H1>
    <P>The Hypertext Mark-up Language (HTML) is a mark-up language designed for the
       presentation of information on the World Wide Web, using a web browser.</P>
    <P>HTML has <A HREF="#S2">mark-up tags</A> that define the presentation of documents
       (headers, paragraphs, lists and tables), hyperlinks and multimedia information.</P>
    <P>The HTML standards are now administered by the <A HREF="http://w3.org">World Wide
       Web Consortium</A></P>
    <H1><A NAME="S2">HTML Mark-up</A></H1>
    <P>Basic HTML mark-up is used to lay out text in the page with headers, paragraphs of text
       and some simple formatting of text within paragraphs.</P>
  </BODY>
</HTML>
```

Click here to view the HTML source of this page.

Including images and multimedia



logo.jpg

<IMG SRC="logo.jpg" WIDTH="100"ALT="UN FAO Logo">

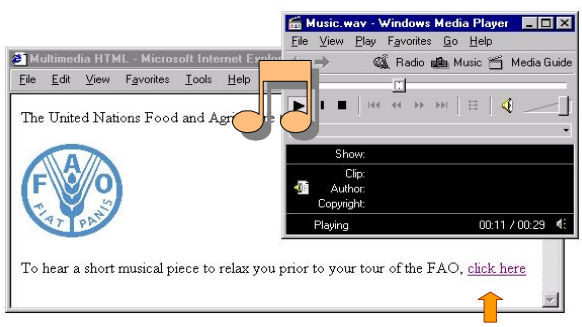
The **SRC** attribute contains the URL of the image file to be displayed

Images (graphics) can be included in an HTML document using the **<IMG>** element.

In this example, the **WIDTH** attribute has been used to specify the size of the graphic (the browser has set the width of the image and adjusted the height to keep the correct aspect ratio).

The **ALT** attribute contains a brief text which can be displayed as an alternative if the image cannot be displayed. It is also used to display a caption for the figure when the mouse is directly over it.

Including images and multimedia



Music.wav

<A HREF="music.wav">click here</A>

The <A> element, used for hyperlinks, can also be used to link to **multimedia** content. In our example, the **HREF** attribute contains the URL of the file containing the clip, and the **TYPE** attribute tells the browser the **MIME** type of the content.

The figure shows what happens when the document is loaded in a web browser and we click on the hyperlink to the audio clip. The browser launches Windows Media Player, an application which can play the audio clip for us.

Multipurpose Internet Mail Extensions (MIME) defines a list of recognized content types, for example "text/html", "image/png", "video/mpeg". The full list of content types is available from http://www.ietf.org/rfc/rfc2046.txt

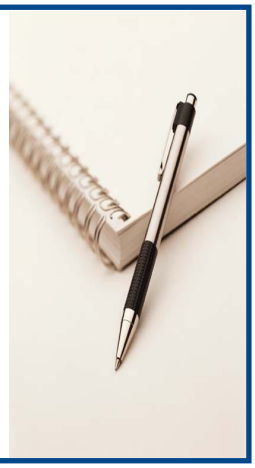
Summary

• HTML is an acronym, standing for **Hypertext Markup Language**. It is a language that can be transferred around the Internet and read by a Web Browser.

• Simple HTML documents can be created easily using **any text editor**.

• All content is defined by the **markup "tags"** of HTML, that are containers for whatever you put in the document.

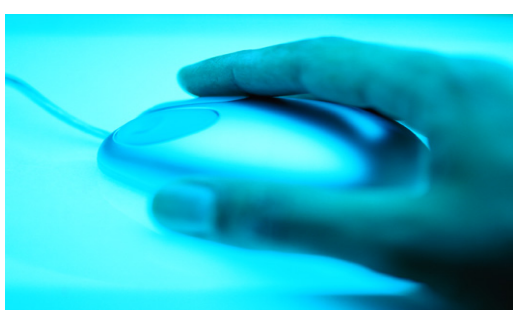
• Using HTML you can define **basic presentation** of a document (headers, paragraphs, lists and tables), **hyperlinks** and **multimedia** information.



Exercise

The following exercise will allow you to apply what you have learned to create an HTML document.

Good luck!



Exercise

Now it's your turn to create this HTML page! Click on the ? icon for help, if needed.

Write the HTML code in the box above.

Then click on View answer.

If you want to know more...

W3Schools Online Web Tutorials (www.w3schools.com), including HTML tutorials.

World Wide Web Consortium (www.w3.org). Open information standards for the Web, including the HTML specification.

W3C 10 Minutes Introduction to HTML (www.w3.org/MarkUp/Guide) - includes links to further information.

'Raggett on HTML 4' is published (1998) by Addison Wesley, ISBN 0-201-17805-2

'Beginning XHTML' is published (2000) by Wrox Press, ISBN 1-861003-43-9

The full list of recognized content types defined by MIME, available from the IETF website (http://www.ietf.org/rfc/rfc2046.txt)