SGG 3643

Computer Programming III

Starting with HTML

Ivin Amri Musliman

Web References

- HTMLSource, http://www.yourhtmlsource.com/
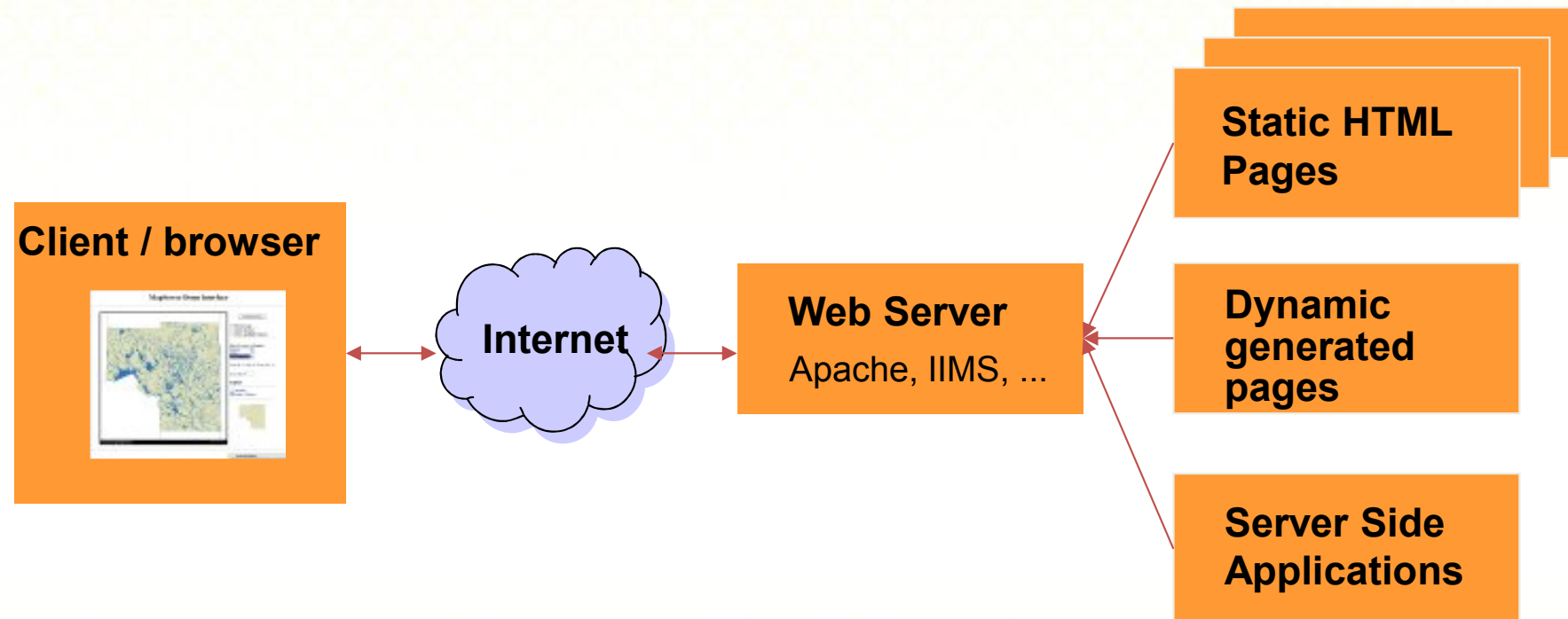- W3schools, http://www.w3schools.com/html/default.asp

Discussion Forums

- comp.infosystems.www.authoring.html: A USENET newsgroup where HTML authoring issues are discussed. "How To" questions should be addressed here. Note that many issues related to forms and CGI, image maps, transparent gifs, etc. are covered in the WWW FAQ.

- www-html@w3.org: A technical discussion list. If you have a proposal for a change to HTML/XHTML

- w3c-translators@w3.org: This is a mailing list for people working on translations of W3C specifications

How does the WWW work?

Client / browser

Internet

Web Server
Apache, IIMS, ...

Static HTML Pages

Dynamic generated pages

Server Side Applications

How does the WWW work?

- Web information is stored in documents called **Web pages**.

- Web pages are files stored on computers called **Web servers**.

- Computers reading the Web pages are called **Web clients**.

- Web clients view the pages with a program called a **Web browser**.

- Popular browsers are, among others, **Internet Explorer, Netscape Navigator, Opera.**

How does the browser fetch the pages?

- A browser fetches a Web page from a server by sending **a request**.

- A request is a standard HTTP request containing **a page address**.

- A page address looks like this: http://www.xyz.com/page.html.

- The server sends the requested document back to the browser.

How does the browser display the pages?

- All Web pages contain **instructions for display**
- The browser displays the page by **reading these instructions**.
- The most common display instructions are called **HTML tags**.
- HTML tags look like this:

<p>This is a Paragraph</p>.

markup

What is "Markup"?

Markup

Heading,
bold, large

italic

Paragraph, left oriented

Source: Dynamic Web-Publishing

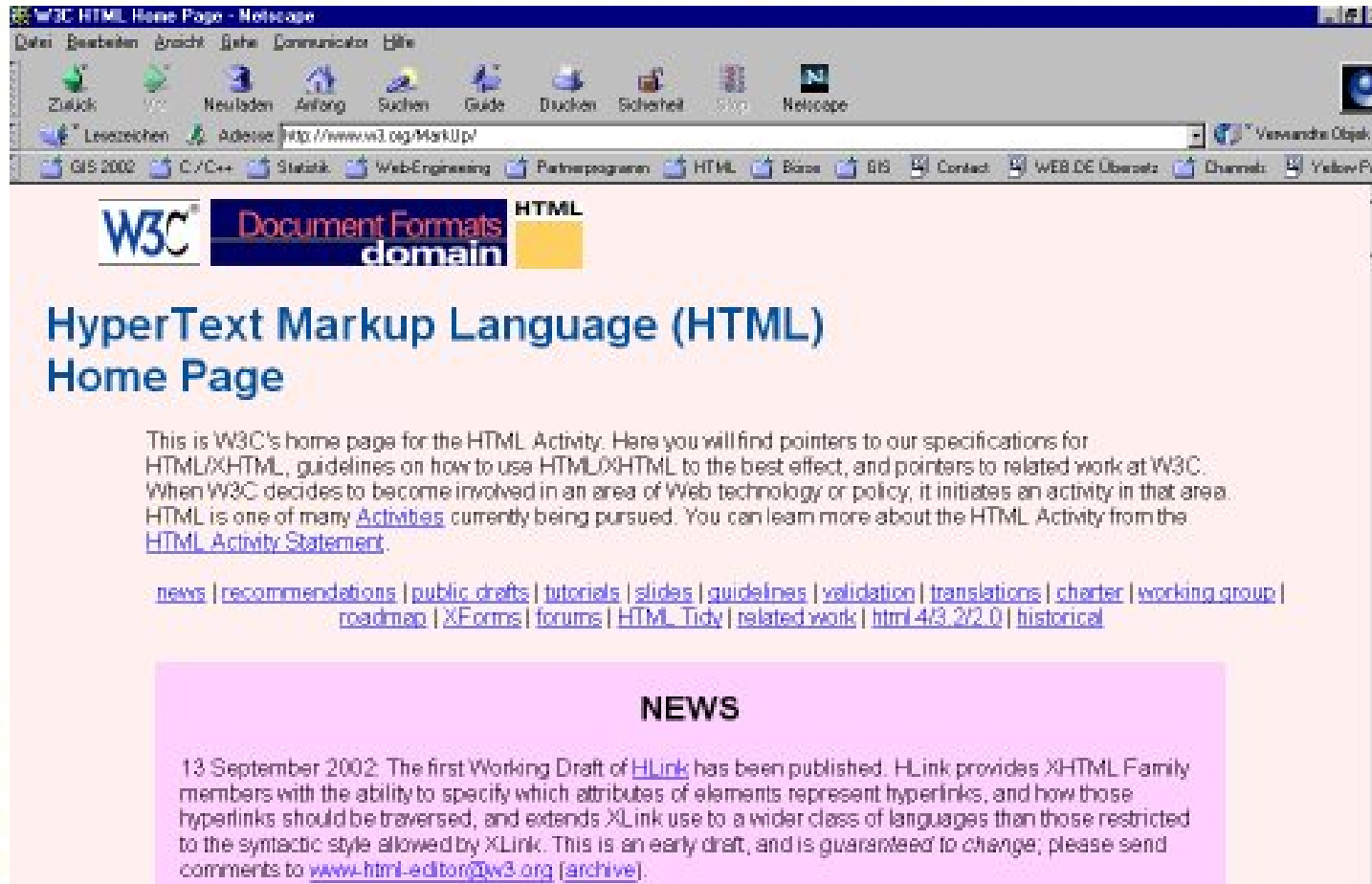# Hypertext Markup Language (HTML)

- no programming language, but **markup** language

- Standardized by a semi-official standard organization (http://www.w3c.org/)

- easy legible: ASCII

- independent from hardware & software

The official Home Page

- http://www.w3.org/MarkUp/

Current State

- XHTML^{TM} 2.0 W3C Recommendation 26 January 2000, revised 1 August 2002
  http://www.w3.org/TR/xhtml1/

- HTML 4.1 *Specification*, http://www.w3.org/TR/html4/

Implementing HTML

- First: HTML base structure.

- Proper usage of entities.

- Use line breaks and empty lines for optimal human readability.
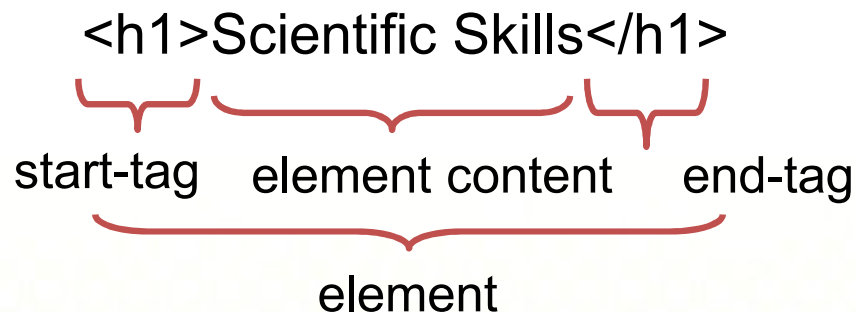
Document structure

- Each HTML document consists of HTML elements.

- HTML elements are defined by HTML tags.

HTML-Tags

- Tags are used for text markup.
- HTML tags have a name and are – like XML tags - surrounded by the **two characters < and >**
- The surrounding characters are called **angle brackets**
- Tags correspond with different elements of the page (paragraphs, lists, special formatted text pieces, images, etc.).

<h1>Scientific Skills</h1>

start-tag   element content   end-tag

element

Structure of Tags

- Tags normally come in pairs, like <h1> ... </h1>

- The first tag in a pair is the start tag, the second tag is the end tag.

- The text between the start and end tags is the element content.

- Empty Tags like <hr /> (horizontal rule) can be written as <hr></hr>.

  Recommended: <hr /> (like in XML).

```
<p>This is a paragraph
<p>This is another paragraph
```

→

```
<p>This is a paragraph</p>
<p>This is another paragraph</p>
```

HTML tags II

- HTML tags are **not case sensitive,** <b> means the same as <B>.

- *However: Lower case letters are recommended.*

- If you want to prepare yourself for the next generations of HTML you should start using lowercase tags. The World Wide Web Consortium (W3C) recommends lowercase tags in their HTML 4 recommendation, and XHTML (the next generation HTML) *demands* lowercase tags.

Tag attributes

- Tags can have attributes.
- Attributes can provide additional information about the HTML elements on your page.
- Attributes always come in name/value pairs: `name="value"`
- Attributes are always added to the *start tag* of an HTML element.

Tag Attributes - Example

- This tag defines the body element of your HTML page: <body>. With an added bgcolor attribute, you can tell the browser that the background color of your page should be red, like this:

  `<body bgcolor="red"> ... </body>`

- This tag defines an HTML table: <table>. With an added border attribute, you can tell the browser that the table should have no borders:

  `<table border="0"> ... </table>`

Quote styles

- Attribute values should always be enclosed in quotes. Double style quotes are the most common, but single style quotes are also allowed.

- In some rare situations, like when the attribute value itself contains quotes, it is necessary to use single quotes:
  name='John "Shotgun" Nelson'

- Note: This nesting is often necessary in JavaScript on transferring parameters to functions..

Nesting of elements

- Think of elements as being like boxes (http://www.w3.org/TR/REC-CSS2/box.html). A box can have a box inside it, or can be inside a box, but it can't be inside a box and outside it as well. Neither can your elements.

This is valid (X)HTML:

****The Lord Of The Rings is a ****fantastic**** story.****

This is not:

****The Lord Of The Rings is a ****fantastic**** story.****

The <em> starts outside the <strong> but finishes inside it. The tags are not properly nested.

Blocks

- Many box positions and sizes are calculated with respect to the edges of a rectangular box called a containing block. In general, generated boxes act as containing blocks for descendant boxes; we say that a box "establishes" the containing block for its descendants.

- Block boxes: have space around them and elements that come after start below them instead of next to them.

- Inline boxes: are displayed next to each other such as a <strong> element within a paragraph, the flow of text is not broken

In the beginning, there was the box....

- A HTML document is a series of (rectangular) boxes, stacked either one inside the other or one after the other.
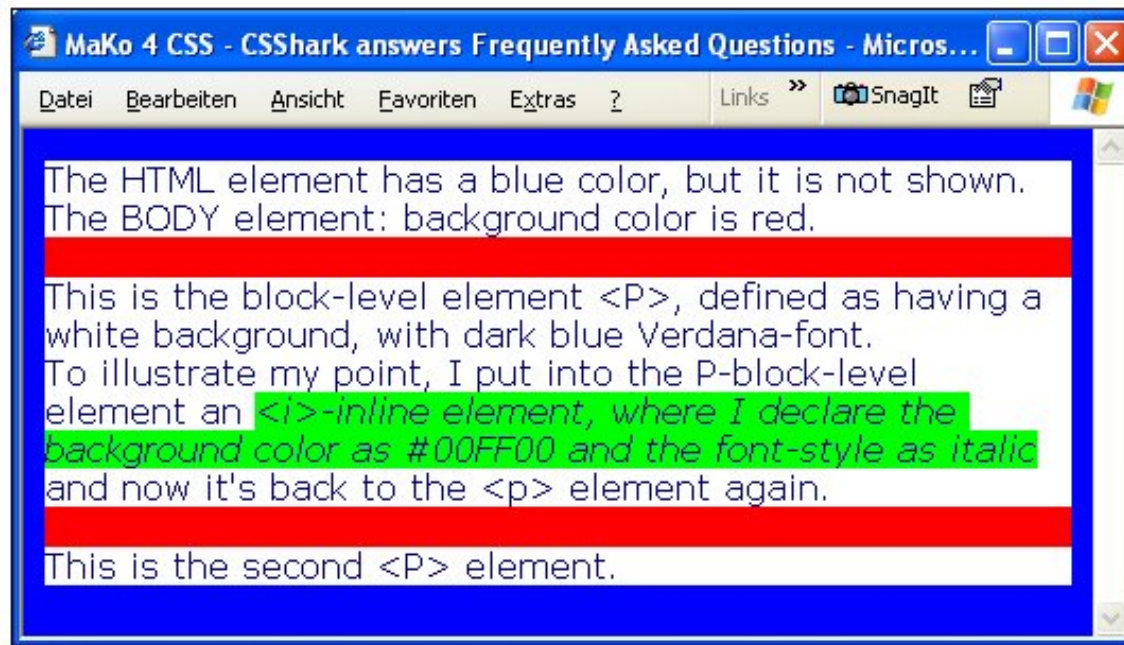
- This model also shows the difference between block-level elements and inline elements - a block-level element begins and ends a line, the inline element does not and is placed within another element.

What is an HTML File?

- HTML stands for **H**yper **T**ext **M**arkup **L**anguage
- An HTML file is a text file containing HTML elements with **markup tags.**
- The markup tags tell the Web browser **how to display** the page.
- An HTML file must have an **htm** or **html** file extension.
- An HTML file can be created using a **simple text editor.**

A first try…

- Use your favorite editor to type in the following lines.
- Store the file with the extension .html.
- Open this file in your favorite browser.

```
<html>
<head>
<title>Title of page</title>
</head>
<body>
<p>This is my first homepage. <b>This text is bold</b></p>
</body>
</html>
```

HTM or HTML Extension?

- When you save an HTML file, you can use either the `.htm` or the `.html` extension.

- Modern software should support four letter extensions.

- I recommend to use `.html` for all your files.

General HTML document structure

Year 2000

```
<!doctype html public "-//W3C//DTD HTML 4.0 //EN">
<html>
<head>
<title>No Title</title>
</head>
<body>
Here comes the contents
</b
</h
```
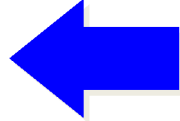
Year 2002

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 2.0//EN"
"http://www.w3.org/TR/xhtml2/DTD/xhtml2.dtd">
<html xmlns="http://www.w3.org/2002/06/xhtml2">
<head>
<title>A study of population dynamics</title>
          ... other head elements...
</head>
<body>
          ... document body...
</body>
</html>
```

Fundamental HTML 4.01 Structure

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
        "http://www.w3.org/TR/html4/strict.dtd">
<html>
<head><title>A simple page</title>
<meta name="description" content="Sample for basic structure of a HTML page">
<meta name="author" content="Ivin Amri">
<meta name="keywords" content="Web, Internet, HTML">
<meta name="generator" content="HTML EDITOR">
<meta http-equiv="Content-Type" content="text/html; charset=iso-8559-1">
 <META HTTP-EQUIV="Refresh"
        CONTENT="10; URL=http://psz.utm.my/"></head>
<body>
        ... document body...
</body>
</html>
```

HTML document type definitions

- The valid tags for markup are declared in document type definitions (DTD). For HTML 4.01, it exist in the following form:

  – strict.dtd

  – loose.dtd

  – frameset.dtd -> discussed later (frames)

- Details: http://www.w3.org/TR/REC-html40/

strict.dtd

- The standard DTD (document type definition), for clean HTML 4-Code.

- This DTD does not contain any deprecated (old) elements and uses stylesheets for formatting purposes.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML
4.01//EN""http://www.w3.org/TR/html4/strict.dtd">
```

strict.dtd II

- Authors should use the Strict DTD unless they need the presentation control for user agents that don't (adequately) support style sheets.

- However, actual browsers support style sheets.

loose.dtd

- DTD for a phase of transition.

- Transitional DTD, which includes additional presentation attributes and elements that W3C expects to phase out as support for style sheets matures ("deprecated elements).

- Authors should use the Strict DTD when possible, but may use the Transitional DTD when support for presentation attribute and elements is required.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
```

The <head> Section

```
...
<head>
<title>A simple page</title>
<meta name="description" content=„Sample for basic structure of a HTML
page.">
<meta name="author" content="Franz-Josef Behr">
<meta name="keywords" content="Web-Engineering,HTML">
<meta name="generator" content="Ulli Meybohms HTML EDITOR">
<meta http-equiv="Content-Type" content="text/html; charset=iso-8559-
1">
 <META HTTP-EQUIV="Refresh"
      CONTENT="10; URL=http://new_website.com/">
</head>
…
```

The <head> Section II

- This is usually a quite short section

- Never put any of the text that should appear on the page itself here

The <head> Section: <title>

- The title appears on the stripe at the very top of the window

- One cannot use character styles (<b>, <u>, <i>, etc) within the <title>...</title> block

- Information in the title can be used by your bookmarking/hotlist and also by search engines

- -> Try to make your titles specific and concise!

The <head> Section: METADATA

- Metadata is "structured data about data"

- Describes the document more than its content

- HTML provides the META tag as a means of inserting two metadata into your HTML documents

  – Metadata about the document itself

  – Additional information to the HTTP header

Dublin Core Metadata

- Standard set of metadata elements established through consensus by an international, cross-disciplinary group of professionals

- Elements prefixed by DC. In META tag

  <META NAME="DC.Creator" CONTENT="Darryl Friesen">

Content		Intellectual Property	Instantiation
Coverage	Source	Contributor	Date
Description	Subject	Creator	Format
Type	Title	Publisher	Identifier
Relation		Rights	Language

Dublin Core: References

The Dublin Core

- – purl.org/dc

- – Nordic DC Metadata Creator
  http://www.lub.lu.se/cgi-bin/nmdc.pl

The <body> Section

- Embraces content of your page which will be displayed in the browser.

- All HTML elements for integration of text, images, hyperlinks etc. are placed between the <body> tags.

- The <body> tag itself can have some additional parameters (covered later),

```
<body>
<p>Your text comes here!</p>
</body>
```

Practice

- Implement your first HTML page using a standard editor.

- Check the result in your browser.

- Try definition of a background image

```
<body background="tile.gif">
```

- Try colors for fonts, background, links...

```
<body text="#000000" bgcolor="#FFFFFF"
link="#FF0000" alink="#FF0000"
vlink="#FF0000">
```

- Try page margins

```
<body leftmargin=30 topmargin=50>
```

Tools used to Create Web Pages

- Two extremes in creating web pages
  - Hand code all the text and html tags
  - Extremely difficult and opaque
    - No easy way to tell from the code how the page actually looks
    - Most efficient use of HTML
    - Reduces or eliminates unforeseen interactions between browsers and the html
      - Many web authoring tools insert nonstandard html code - this sometimes causes unexplained interactions
    - Some effects can only be achieved by hand coding
      - e.g., pop-up windows, …

Tools used to Create Web Pages II

- Two extremes in creating web pages
  - Use a web authoring tool to perform all the HTML markup
    - Many tools allow the user simply to type text and then automatically add the html tags
    - Depending on the sophistication of the authoring tool, it may also bundle features such as graphics, linking, animation and multimedia and allow the user to access them via menus and pull-down screens.
    - Chief advantages are
      - WYSIWYG
      - Novice users can immediately create a page

Free HTML-Editors

- HTML-Kit (http://www.chami.com/html-kit/) plus additional plug-ins

- NetScape Composer

- ASCII-Editor (i.e. UltraEdit, http://www.ultraedit.com)

- Ulli Meybohms German HTML Editor Phase 5 (http://www.meybohm.de/htmledit/index.html)

WYSIWYG HTML editors

- WYSIWYG = what you see is what you get
- There are far too many to mention, but here's a small selection
  - GoLive
  - DreamWeaver
  - Front Page
  - Pagemill (MacIntosh)
  - SoftQuad HoTMetal
  - …

FTP-Programs

- ftp programs are necessary to transfer the results to the server.

- ws-ftp, http://www.ispwitch.com

Graphic programs

- Graphic programs (and even standard office programs) nowadays export data to HTML:

- Corel Draw, etc.

- Adobe products like Photoshop, GoLive, …

- …

Strongly recommended: HTML Tidy

- … an open source utility for tidying up HTML developed by Dave Ragett.
- composed from an HTML parser and an HTML pretty printer.
- provides advice on how to make your pages more accessible to people with disabilities,
- can be used to convert HTML content into XML as XHTML.
- Tidy is W3C open source and available free.
- It has been successfully compiled on a large number of platforms, and is being integrated into many HTML authoring tools.
- Recently the maintenance of Tidy has been taken over by a group of dedicated volunteers on SourceForge, see:
  http://tidy.sourceforge.net/

Finally: Get Search Engines to Crawl Your Site!

- Google http://www.google.com/addurl.html

- Altavista http://doc.altavista.com/addurl/

- Yahoo http://docs.yahoo.com/info/suggest/

- Lycos http://home.lycos.com/addasite.html