2.12 Hypertext and XML Data Structures

Hypertext:

- A mechanism for representing information structure.
- It differs from traditional information storage data structures in format and use.
- Hypertext is stored in Hypertext Markup Language (HTML) and eXtensible Markup Language (XML).
- HTML and XML provide detailed descriptions for subsets of text similar to the zoning that increase search accuracy and improve display of hit results.

- Hypertext Structure
  - used in the Internet environment
  - requires electronic media storage for the item.
  - Hypertext allows one item to reference another item via an imbedded pointer.
  - Each separate item is called a node.
  - Reference pointer is called a link.
  - Each node is displayed by a viewer that is defined for the file type associated with the node.
- Hypertext Markup Language (HTML)
  - defines the internal structure for information exchange across the World Wide Web on the Internet.
  - A document is composed of the text of the item along with HTML tags that describe how to display the document.
  - Tags are formatting or structural keywords contained between less-than, greater than symbols (e.g., <title>, <strong>).
  - The HTML tag associated with hypertext linkages is <a href= …#NAME /a>
  - where "a" and "/a" are an anchor start tag and anchor end tag denoting the text that the user can activate.
  - "href" is the hypertext reference containing either a file name if the referenced item is on this node or an address(URL) and a file name if it is on another node.
  - "#NAME" defines a destination point other than the top of the item to go to.

  - The URL has three components:

- access method the client used to retrieve the item

- Internet address of the server where the item is stored

- address of the item at the server

- Figure 4.14 shows an example of a segment of a HTML document.

```
<CENTER>
<IMG SC="/images/home_iglo.jpg" WIDTH=468 HEIGHT=107
BORDER=0 ALT="WELCOME TO NETSCAPE><BR>
<P>
<DL>
<A HREF="/comprod/mirror/index.html">
<DD>
The   beta   testing   is   over:   please   read   our   report   <A
HREF="http://www.charm.net/doc/charm/report/theme.html">   and   your
can            find            more            references            at
HREF="http://www.charm.net/doc/charm/results/tests.html">
```

Figure 4.14 Example of Segment of HTML

- Hypertext is a non-sequential directed graph structure, where each node contains its own information.

- The author assumes, the reader can follow the linked data as easily as following the sequential presentation.

- A node may have several outgoing links, each of which is then associated with some smaller part of the node called an anchor.

- When an anchor is activated, the associated link is followed to the destination node, thus navigating the hypertext network.

- Conventional items are read sequentially by a user.

- In a hypertext environment, the user "navigates" through the node network by following links.

- Hypertext references are used to include information that is other than text (e.g., graphics, audio, photograph, video) in a text item.

- Dynamic HTML
    - Combination of the latest HTML tags, style sheets and programming that help to create WEB pages that are more animated and responsive to user interaction.

- Supports features such as object-oriented view of a WEB page and its elements, cascading style sheets, programming that can address most page elements add dynamic fonts.
- Allows the specification of style sheets in a cascading fashion.

XML:

- The extensible Markup Language (XML) is a standard data structure on the WEB.
- Objective is to extend HTML with semantic information.
- The logical data structure within XML is defined by a Data Type Description (DTD)
- The user can create any tags needed to describe and manipulate their structure.
- The following is a simple example of XML tagging:

```
<company>Widgets Inc.</company>
<city>Boston</city>
<state>Mass</state>
<product>widgets</product>
```

- Resource Description Format (RDF)
  - used to represent properties of WEB resources such as images, documents and relationships between them.
  - This will include the Platform for Internet Content Selection (PICS) for attaching labels to material for content filtering (e.g., unsuitable for children).
- Hypertext links for XML
  - These are defined in the Xlink (XML Linking Language) and Xpoint (XML Pointer language) specifications.
  - allow different types of links to locations within a document and external to the document.
  - allow an application to know if a link is positioning reference within an item or link to another document.
  - help in determining what needs to be retrieved to define the total item.
- XML Style Sheet Linking
  - define how to display items on a particular style sheet and handle cascading stylesheets.
  - allow designers to limit what is displayed to the user and allow expansion to the whole item if desired.