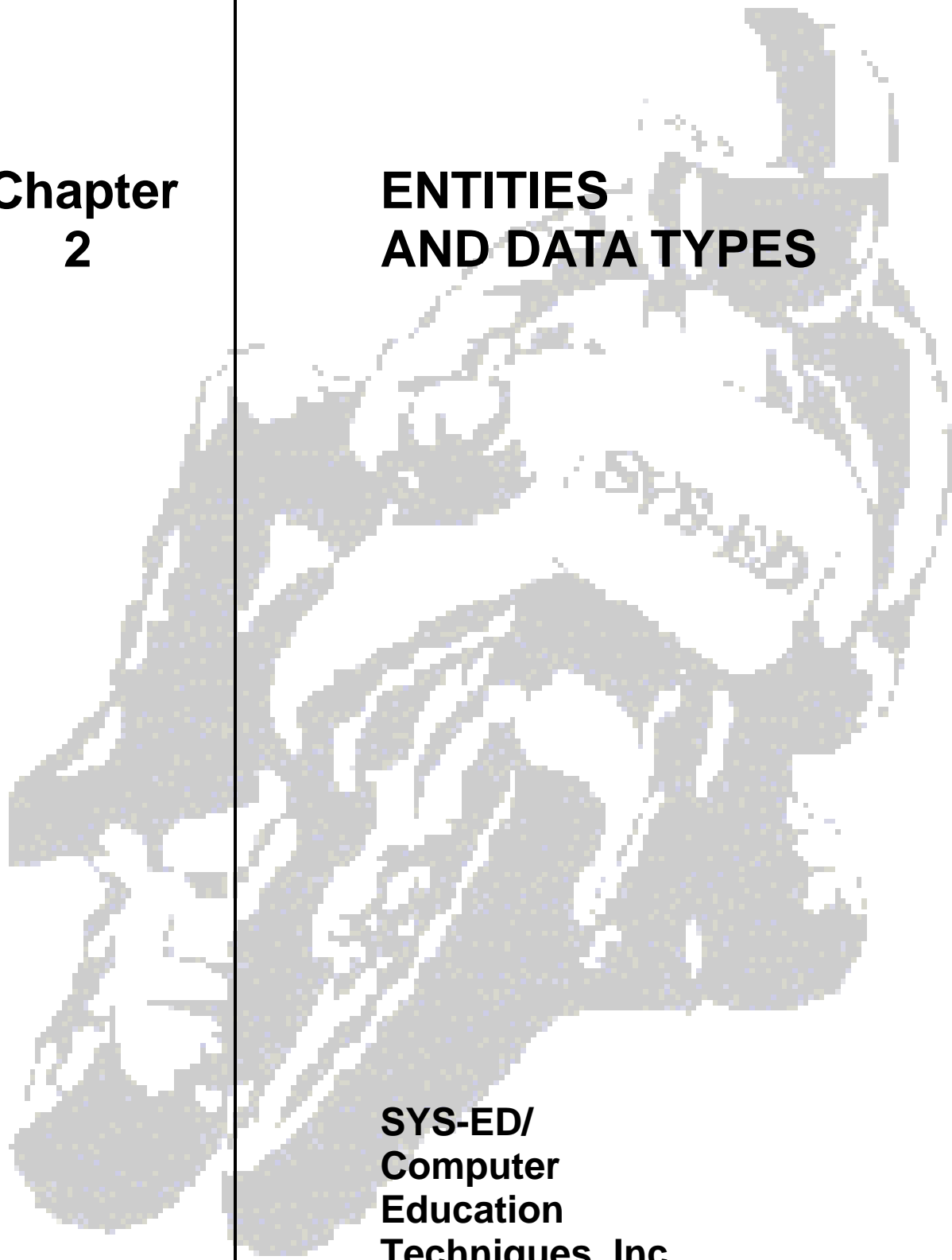


**Chapter
2**

**ENTITIES
AND DATA TYPES**

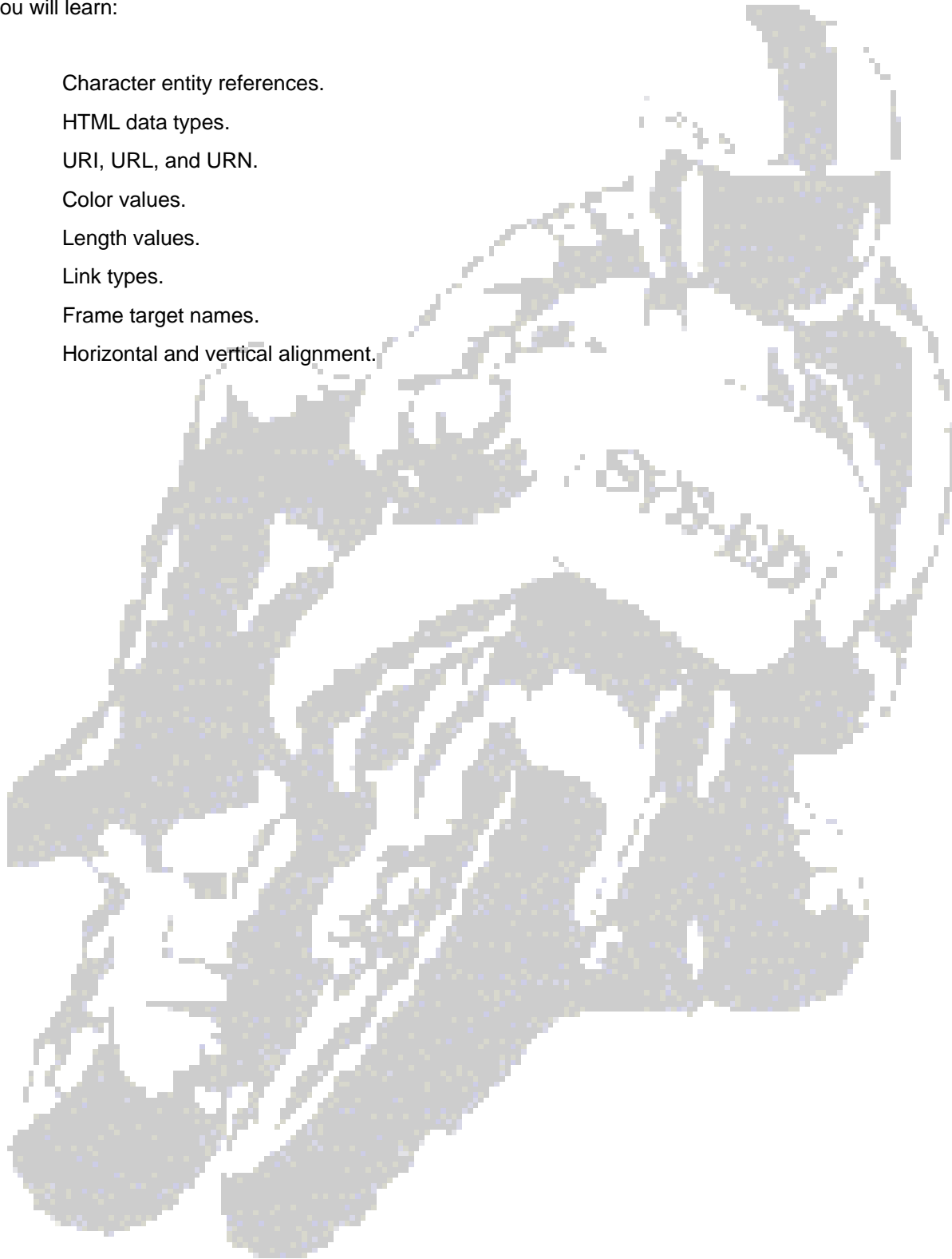


**SYS-ED/
Computer
Education
Techniques, Inc.**

Objectives

You will learn:

- Character entity references.
- HTML data types.
- URI, URL, and URN.
- Color values.
- Length values.
- Link types.
- Frame target names.
- Horizontal and vertical alignment.



1 Character Entity References

A character entity reference is a SGML construct that references a character of the document character set.

HTML 4.0 supports several sets of character entity references:

ISO 8859-1 (Latin-1) characters	This specification now covers the entire set of Latin characters, such as ; © ; and ® .
Symbols, mathematical symbols, and Greek letters	These characters are represented by glyphs in the Adobe font "Symbol".
Markup-significant and internationalization characters	These are used for bidirectional text.

1.1 HTML Character Entities

Some characters have a special meaning in HTML, such as the less than sign (<) which defines the start of an HTML element. In order for the browser to display these characters entities will need to be inserted in the HTML source.

A character entity has three parts an:

• ampersand (&)	• entity name or a # and an entity number	• a semicolon (;).
-----------------	---	--------------------

In order to display a less than sign in an HTML document it will be necessary to write: < or <.

- The advantage of using a name instead of a number is that a name is easier to remember.
- The disadvantage is that not all browsers support the newest entity names, while the support for entity numbers is very good in almost all browsers.

Entities are case sensitive.

1.2 Non-breaking Space

The most common character entity in HTML is the non-breaking space.

HTML will truncate spaces in text. If there are 10 spaces in the text, HTML will remove 9 of them. In order to add spaces to the text, the character entity will need to be used.

Example:

```
SYS-ED is the licensed computer school owned and operated by
Computer Education Techniques, Inc.
```

The text in this HTML file will be displayed with "Computer Education Techniques, Inc." as one piece. It will not be broken into multiple lines.

1.3 Common Entities

Character	Description	Entity by Name	Entity by Number
	non-breaking space	 	
<	less than	<	<
>	greater than	>	>
&	Ampersand	&	&
"	quotation mark	"	"
'	apostrophe	'	'
¢	Cent	¢	¢
£	Pound	£	£
¥	Yen	¥	¥
§	Section	§	§
©	Copyright	©	©
®	registered trademark	®	®
×	Multiplication	×	×
÷	Division	÷	÷

2 HTML Data Types

2.1 Case Information

Each attribute definition includes information about the case-sensitivity of its values. The case information is presented with the following keys:

Keys	Explanation
CS	The value is case-sensitive. User agents interpret "a" and "A" differently.
CI	The value is case-insensitive. User agents interpret "a" and "A" as the same.
CN	The value is not subject to case changes. Because it is a number or a character from the document character set.
CA	The element or attribute definition itself gives case information.
CT	Consult the type definition for details about case-sensitivity.

If an attribute value is a list, the keys apply to every value in the list, unless otherwise indicated.

2.2 Text Strings

A number of attributes, %Text; in the DTD, accept text that is meant to be "human readable".

3 URI, URL, and URN

The web includes objects which can be accessed using an extendable number of protocols.

3.1 URI: Uniform Resource Identifier

This document defines a way to encapsulate a name in any registered name space, and label it with the name space, producing a member of the universal set. An encoded and labeled member of this set is known as a Universal Resource Identifier, or URI.

The universal syntax allows access of objects available using existing protocols. The specification of the URI syntax does not imply anything about the properties of names and addresses in the various name spaces which are mapped onto the set of URI strings.

The properties follow from the specifications of the protocols and the associated usage conventions for each scheme.

3.2 URL and URN

For existing Internet access protocols, it is necessary in most cases to define the encoding of the access algorithm into an address.

There is currently a movement to define a space of more persistent names than any URLs. The URI syntax and URL forms have been in widespread since 1990.

4 Colors

The attribute value type "color" (%Color;) refers to color definitions. A color value may either be a hexadecimal number - prefixed by a hash mark - or one of the following sixteen color names.

The color names are case-insensitive.

4.1 Color Names and sRGB Values

Color Name	Color Value
Black	#000000
Green	#008000
Silver	#C0C0C0
Lime	#00FF00
Gray	#808080
Olive	#808000
White	#FFFFFF
Yellow	#FFFF00
Maroon	#800000
Navy	#000080
Red	#FF0000
Blue	#0000FF
Purple	#800080
Teal	#008080
Fuchsia	#FF00FF
Aqua	#00FFFF

The color values "#0000FF" and "Blue" both refer to the color purple.

4.2 Guidelines for Using Colors

Although colors can add significant amounts of information to documents and make them more readable, the following guidelines need to be evaluated when including color in documents:

- The use of HTML elements and attributes for specifying color is deprecated. Style sheets should be used instead of color attributes.
- Color combinations that cause problems for people with color blindness should be avoided.
- If a background image or background color is used, then various text colors should be used to create contrast.
- Colors specified with the BODY and FONT elements and bgcolor on tables look different on different platforms. The effects should be checked on different types of machines.

5 Lengths

HTML specifies three types of length values for attributes:

Pixels	The value is an integer that represents the number of pixels of the screen.
Length	The value may be measured either in pixels or a percentage of the available horizontal or vertical space.
MultiLength	The value may be a length; or a relative length. A relative length has the form "i*", where "i" is an integer.

Length values are case-neutral.

6 Link Types

Link types refers to a space-separated list of link types. White space characters are not permitted within link types.

The link types are case-insensitive. User agents, search engines, etc. may interpret these link types in a variety of ways.

Alternate	Designates substitute versions for the document in which the link occurs. When used together with the lang attribute, it implies a translated version of the document.
Stylesheet	Refers to an external style sheet. This is used together with the link type "Alternate" for user-selectable alternate style sheets.
Start	Refers to the first document in a collection of documents. This link type tells search engines which document is considered by the author to be the starting point of the collection.
Next	Refers to the next document in a linear sequence of documents. User agents may choose to preload the "next" document, to reduce the perceived load time.
Prev	Refers to the previous document in an ordered series of documents. Some user agents also support the synonym "Previous".
Contents	Refers to a document serving as a table of contents. Some user agents also support the synonym ToC, from "Table of Contents".
Index	Refers to a document providing an index for the current document.
Glossary	Refers to a document providing a glossary of terms that pertain to the current document.
Copyright	Refers to a copyright statement for the current document.
Chapter	Refers to a document serving as a chapter in a collection of documents.
Section	Refers to a document serving as a section in a collection of documents.
Subsection	Refers to a document serving as a subsection in a collection of documents.
Appendix	Refers to a document serving as an appendix in a collection of documents.
Help	Refers to a document offering help.
Bookmark	Refers to a bookmark. A bookmark is a link to a key entry point within an extended document.

7 Frame Target Names

Frame target names must begin with an alphabetic character: a-z, A-Z. User agents should ignore all other target names.

The following target names are reserved and have special meanings.

_blank	The user agent should load the designated document in a new, unnamed window.
_self	The user agent should load the document in the same frame as the element that refers to this target.
_parent	The user agent should load the document into the immediate FRAMESET parent of the current frame. This value is equivalent to _self if the current frame has no parent.
_top	The user agent should load the document into the full, original window (thus canceling all other frames). This value is equivalent to _self if the current frame has no parent.

8 Horizontal and Vertical Alignment

8.1 Horizontal Alignment – The align Attribute

align Attribute Value	Description
Left	Left-flush data/object. This is usually the default value.
Center	Center data/object.
Right	Right-flush data/object.
Justify	Double-justify text.
Char	Align text around a specific character. If a user agent doesn't support character alignment, behavior in the presence of this value is unspecified.

8.2 Vertical Alignment – The valign Attribute

This attribute specifies the vertical position of data/object.

valign Attribute Value	Description
top	Data/object is flushed with the top of the cell.
middle	Data/object is centered vertically. This is usually the default value.
bottom	Data/object is flushed with the bottom.
baseline	In a table, all cells in the same row as a cell whose valign attribute has this value should have their textual data positioned so that the first text line occurs on a baseline common to all cells in the row. This constraint does not apply to subsequent text lines in these cells.