**HTML Syntax Rules**

You should ensure your HTML is **well-formed.**

Only certain tags are permitted.

Only certain attributes and values are permitted.

Often, only content is permitted;

* e.g. <p> elements cannot contain other <p> elements

Nesting must be done properly. Otherwise, there is difficulty building the tree.

**Two Sets of Syntax Rules!**

**XML syntax** uses a very strict set of rules.

**HTML syntax** allows you to break the rules*… in certain cases.*

**XML Syntax**

Tags must be lowercase.

Each tag must have an end tag.

… etc.

**Breaking the Rules**

What does your browser do if your web page is not well-formed?

* Browsers (almost never gives error messages. They do their best to build the tree and display the page

**HTML Validation**

If browsers don't give error messages, how do you know if your page is well-formed or not?

* You can **validate** your page: <https://html5.validator.nu/>

**Character Sets**

A **character set** is a collection of characters, e.g. ASCII or Unicode.

A **coded character set** assigns a unique number to each distinct character.

A **character encoding** refers to the way the numbers are converted to bytes for storage and transmission.

Graphical user interface, text

Description automatically generated with medium confidence

Browsers need to know which character encoding was used to create your web page.

**A Better Solution?**

The Apache web server can be configurated so that when it serves a text file, it converts it to e.g. UTF-8.

**Reserved Characters**

Some characters have a special meaning in HTML.

To display them in a web page, you may need to use their **character references:**

A picture containing graphical user interface

Description automatically generated

Suppose you want to display a character that is not part of your character set or not easy to type on a keyboard, then you can yse a character reference:

Background pattern

Description automatically generated with low confidence