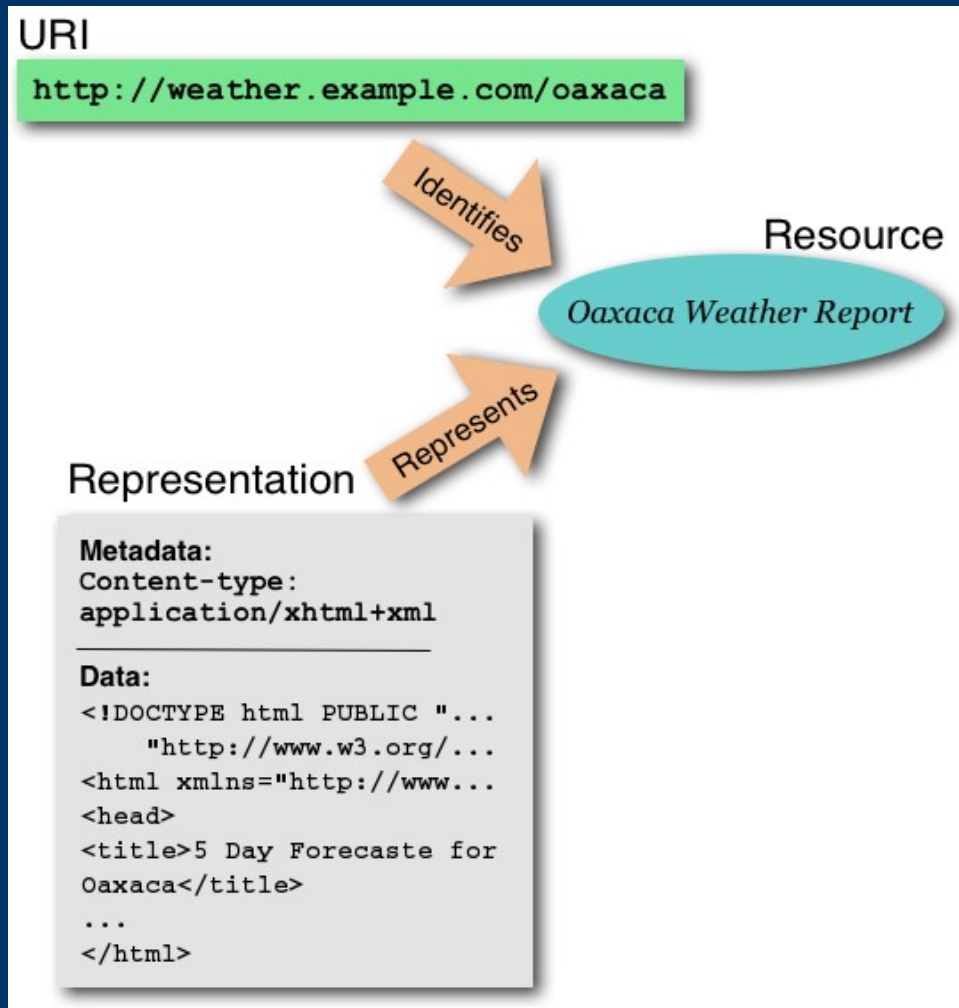


Data Formats and Resource Representations for the Web



Resources in the Web



- We “access” resources in the Web.
- Resources may be “written” or “represented” using web-based content/data formats:
 - Text
 - Images
 - Etc..

Definition: Web Page

- A “page” or document or content that we access in the web.
- Usually an HTML page.
 - It may be any of the following:
 - Text
 - XML-based content formats
 - Graphics
 - Multimedia
 - Even binary-encoded files.

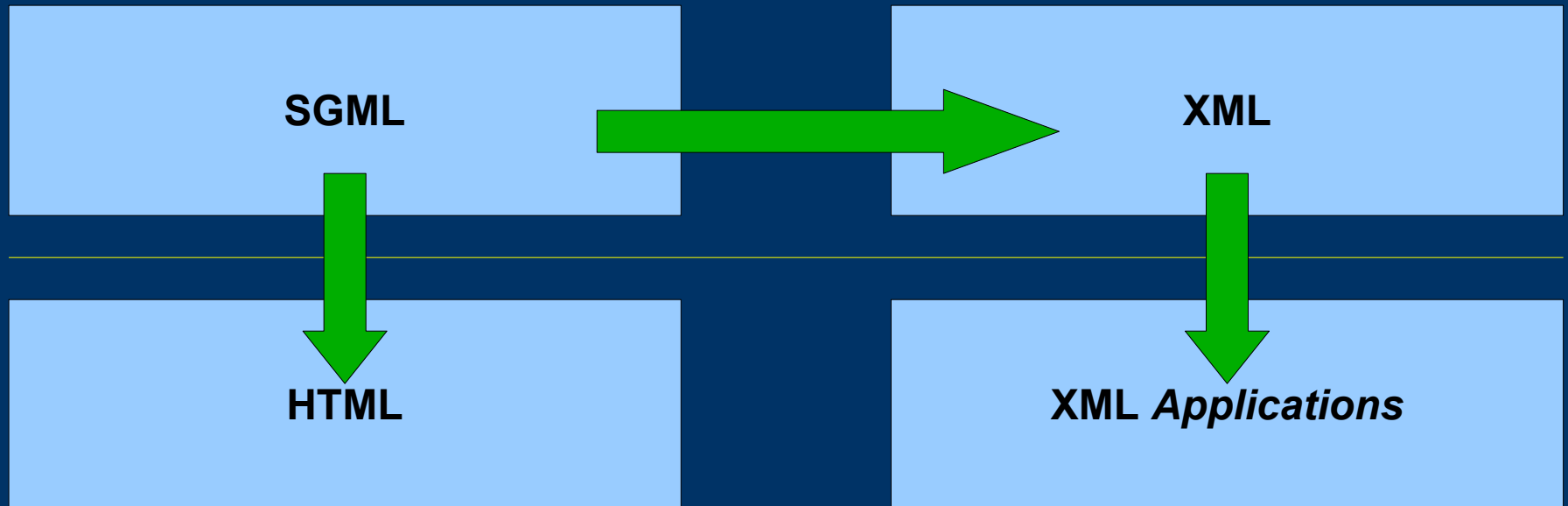
HTML: Hypertext Markup Language

- Basic web programmer's markup language.
- Last Known Version: 4.01*
- Uses “tags” to markup data or sections in a web page.
- An application of SGML (*Standard Generalized Markup Language*)
 - SGML is a metalanguage.

HTML 4.01 to XHTML 1.0

- Development of HTML as a language was “halted” in favor of its XML-based incarnation.
- XHTML : Extensible Hypertext Markup Language
 - Reformulation of HTML in XML
 - XHTML is an XML application.

Metalanguages



XML : Extensible Markup Language

- The meta-language for creating XML-based documents.
 - As a meta language, it provides rules as to how these application languages are created.
- The term extensible means:
 - You can create your own document type.
 - You can create your own tags.
 - You may add new elements/tags to existing document types/applications for your own use.

XML : Data document-based

- XML focuses on data
 - Especially data that can be arranged into “document-form”
 - XML is simplified SGML
 - Designed for the exchange of data in the Internet
 - XML is for self-descriptive data.
 - The data is described by tags. Much like in HTML.*
 - XML document elements can be arranged as a tree-structure.
 - XML is text-based and human-readable.
 - Separate Lecture on XML to follow...
-
-

XHTML

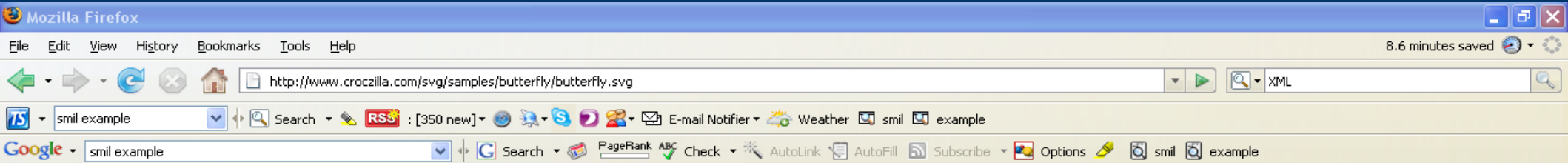
- A much stricter form of HTML.
- Follows the rigid rules of XML
 - ...which results to better and faster processing
 - E,g, rules:
 - One root element (which is `<html>`).
 - Required elements: DOCTYPE, html, head, title, body
 - Tags are always “closed” or a proper singleton tag (e.g. `<p>...</p>` `<hr/>`)
 - No overlap of tags.
 - Lower case tag names and attribute names.
 - Attributes in the form `attr="value"` or `attr='value'`.

XHTML

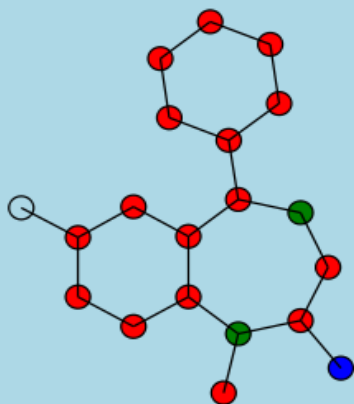
- In XHTML (as well as other XML-based documents), other document types may be combined or embedded. E.g. XHTML+SVG, XHTML+MathML, etc.
- In the new Modularized XHTML, you create your own XHTML subset for your special needs.

Invasion of XML-based formats

- TREND: New data formats are applications of XML.
 - Standard and simple API for XML parsing.
 - XML is backward compatible as well as future compatible.
 - Entities that exchange data/documents in the Internet now prefer XML-based formats over proprietary/binary formats.
 - Hype perhaps.



Scalable Vector Graphics (SVG) Example



Source of: http://www.croczilla.com/svg/samples/diazepam/diazepam.xml - Mozilla Firefox

```
<?xml version="1.0" ?>
<?xml-stylesheet href="cml.css" type="text/css"?>
<!-- original source: http://www.xml-cml.org/ -->
<cmlidoc>
<molecule convention="MDLMol" id="diazepam">
  <date day="22" month="11" year="1995">
  </date>
  <atomArray>
    <atom id="a1">
      <string builtin="elementType">C</string>
      <float builtin="x2">1.4278</float>
      <float builtin="y2">-0.202</float>
    </atom>
    <atom id="a2">
      <string builtin="elementType">C</string>
      <float builtin="x2">0.2204</float>
      <float builtin="y2">0.7483</float>
    </atom>
    <atom id="a3">
      <string builtin="elementType">C</string>
      <float builtin="x2">1.1983</float>
      <float builtin="y2">-1.7262</float>
    </atom>
    <atom id="a4">
      <string builtin="elementType">N</string>
      <float builtin="x2">2.9383</float>
      <float builtin="y2">0.1377</float>
    </atom>
    <atom id="a5">
      <string builtin="elementType">C</string>
      <float builtin="x2">0.2204</float>
      <float builtin="y2">2.3093</float>
    </atom>
    <atom id="a6">
      <string builtin="elementType">C</string>
      <float builtin="x2">-1.111</float>
      <float builtin="y2">-0.0138</float>
    </atom>
    <atom id="a7">
      <string builtin="elementType">C</string>
      <float builtin="x2">2.4195</float>
      <float builtin="y2">-2.6766</float>
    </atom>
    <atom id="a8">
      <string builtin="elementType">C</string>
```

Chemical Markup Language (CML)

MathML Basics - Mozilla Firefox

File Edit View History Bookmarks Tools Help

9.2 minutes saved

DOM Source of MathML - Mozilla Firefox

File Edit View Help

```

<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mrow>
    <mi>k</mi>
    <mo>=</mo>
    <mfrac>
      <mrow>
        <mfrac>
          <mrow>
            <msup>
              <mo>&part;</mo>
              <mn>2</mn>
            </msup>
            <mi>z</mi>
          </mrow>
          <mrow>

```

MathML Basics

Home | Projects | Docs | Jargon

Feedback | FAQ | Search

This document illustrates some basic MathML constructions. It is tailored to display correctly with just the Symbol font. You can view its source. However, MathML documents tend to be verbose and you might get lost trying to locate a MathML fragment with the usual view source. This demo has been made to illustrate the following aspects. You can right-click on any math fragment of interest throughout this document. The context menu won't show up. Rather, the math fragment will zoom, and if you right-click a second time, you will see the MathML WYSIWYG markup of the fragment, and if you right-click again a third time, the fragment will revert to its initial state. This tri-state mode is aimed at limiting conflicts with other agents that compete for the mouse.

With MathML, one can build sets such as (go on, right-click any of these equations to experiment the zoom) $\{0, 1, 2, 3, 4\}$ or $\{\lfloor \frac{a}{b} \rfloor a^2 + b^2 \leq 3\}$, write calculus $\frac{dy}{dx} = \frac{1}{y^2}$, form rather complicated expressions

$$\lim_{n \rightarrow N} (1 + \frac{1}{n})^n - e^N, k = \frac{\frac{\partial^2 x}{\partial x^2} \frac{\partial^2 x}{\partial y^2} - (\frac{\partial^2 x}{\partial x \partial y})^2}{(1 + (\frac{\partial x}{\partial x})^2 + (\frac{\partial x}{\partial y})^2)^2}, \text{ write vector equations } Y = aX + b, \text{ etc.}$$

Notice how the mathematics appear in the main flow of text and respond as you resize the window. You can also make displayed equations, such as the following ones:

$$x \xrightarrow{\text{maps to}} y = f_n(x) = \left(1 + \frac{1}{x^n}\right)^n$$

$$\int_a^b f(x) dx = \frac{b-a}{6} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right] - \frac{(b-a)^5}{4!5!} f^{(4)}(\eta), \quad a \leq \eta \leq b$$

$$|x| = \begin{cases} -x & \text{if } x < 0 \\ x & \text{otherwise} \end{cases}$$

example shows the i -th step of the multiplication of a matrix A by a vector x (notice how $a_{i1}, \dots, a_{in}, x_1$ are on the same

DOM Source of MathML - Mozilla Firefox

File Edit View Help

```

<math display="block" xmlns="http://www.w3.org/1998/Math/MathML">
  <mrow>
    <msubsup>
      <mo>&int;</mo>
      <mi>a</mi>
      <mi>b</mi>
    </msubsup>
    <mi>f</mi>
    <mrow>
      <mo>(</mo>
      <mi>x</mi>
      <mo>)</mo>
    </mrow>
    <mi>d</mi>
  </mrow>

```

i -th row

$$\begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

MathML combined in XHTML

Other Popular XML-Based Data Formats

- SMIL
 - Synchronized Multimedia Integration Language
 - GML
 - Geography(ic) Markup Language
 - XSL, XML Schema
 - Used with XML documents for stylesheet and structure spec
 - Data formats used by businesses (XBRL), news organizations, etc.
-
-

The Web Browser

- In charge of “rendering” content/data.
 - Thin client/browser
 - Content being rendered does not need additional software
 - I.e. If browser can “draw” the content without installing anything else.
 - Thick/Fat/Rich client/browser
 - Content requires additional software.
 - E.g. HTML containing Java applets, Flash requires installation of run time environments.
 - Hybrid
 - Processing performed at the client but requires data from the server.
-
-

Go Google

- <http://www.w3schools.com> for different tutorials.
- <http://www.w3.org> for technical papers/specifications of the different languages.

