A Short Introduction to XML: Mark Up

Eugenia Sokolinski
What is it all about


<table>
<thead>
<tr>
<th>Article</th>
<th>Publication</th>
<th>Issue/Volume</th>
<th>Year</th>
<th>Page Numbers</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islam in Etiopia</td>
<td>Gori Alessandro</td>
<td>Rassegna di studi etiopici</td>
<td>37</td>
<td>1993-45-87</td>
<td>http://....</td>
</tr>
<tr>
<td>“Paleografia quale scienza dello spirito”: Once more on the Gǝʿǝz</td>
<td>Bausi Alessandro</td>
<td>J. B. Quenzer, ed., Exploring Written Artefacts: Objects, Methods, and Concepts</td>
<td>1</td>
<td>2021-33-33</td>
<td></td>
</tr>
</tbody>
</table>
Part 1
Things about things (or Words about words)

Metadata
Metalanguage
Markup
Markup language
Tag
Explaining things: metadata

**Metadata**: data (information) about things (or other data)
- e.g. A book in a library has a description in the catalogue = metadata
- e.g. People and companies have short descriptions (address, phone number) in a telephone book = metadata
- e.g. Persons, places, objects, events are described in an encyclopaedia = metadata
- e.g. Words are described in a dictionary = metadata
- e.g. A manuscript in a repository, or a digital copy of a manuscript in a computer has a description in a catalogue (printed or digital) = metadata
- e.g. A text has been copied many times, with changes, and a collation/critical edition have been produced to describe it = metadata
- can be descriptive (title/name, content/profession), administrative (access, rights, use), structural (composition)
Explaining things: metalanguage

**Metalanguage**: a language (words or symbols = terms) to describe another language
- e.g. grammatical categories sets (noun, verb, case, etc.) used to describe a human language
- e.g. programming language, that is a set of terms to get a computer produce the results you want
Explaining things: markup

**Markup**: is needed to structure text to make it understandable (to humans and to computers, but computers of course need more formal logic)

```
Feature - Researching to Add Richness

Sometimes you're writing nothing but a Blistering Hot love scene, or a character study, or a drabble. In those cases, you're canon is all you'll want or need, but there are other times a Yuletide, a Big Bang, your own original novel-in-progress, that sprawling Au epic that's been taking up all your spare imagination for weeks when you're going to want the sort of little details that add texture and depth to a story. You maybe blessed with a library that includes just the reference books you want or a local library with sympathetic librarians. If not, there's always the Interwebs ... but sifting treasure from trash can be a challenge.

Let's take a look at some ways to make the World Wide Web cough up the details that you need to write the story that you want.

With examples involving the historical novel Frontier Wolf.
```
Explaining things: markup language

**Markup language**: a set of rules governing what markup information may be included in a document and how it is combined with the content of the document.

<table>
<thead>
<tr>
<th>Marking</th>
<th>Location</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>margin</td>
<td>Align text vertically or horizontally, depending on the guidelines drawn in the text.</td>
<td></td>
</tr>
<tr>
<td>margin</td>
<td>Use boldface type for the marked text.</td>
<td></td>
</tr>
<tr>
<td>inline</td>
<td>Use boldface type on the text underlined with the zigzags.</td>
<td></td>
</tr>
<tr>
<td>inline</td>
<td>Capitalize the marked text.</td>
<td></td>
</tr>
<tr>
<td>inline</td>
<td>Capitalize the circled text (alternate version).</td>
<td></td>
</tr>
<tr>
<td>margin</td>
<td>Capitalize the indicated text.</td>
<td></td>
</tr>
<tr>
<td>margin and inline</td>
<td>Insert a comma.</td>
<td></td>
</tr>
<tr>
<td>margin and inline</td>
<td>Delete the marked text. When used inline, the longer &quot;tail&quot; of the symbol crosses through the text to be deleted.</td>
<td></td>
</tr>
<tr>
<td>margin and inline</td>
<td>Delete the marked text and close up the characters on either side to fill the space left.</td>
<td></td>
</tr>
<tr>
<td>margin</td>
<td>Insert an em dash (the insertion of an en dash is handled similarly).</td>
<td></td>
</tr>
<tr>
<td>inline</td>
<td>Change the hyphen to an em dash (the change of a hyphen to an en dash is handled similarly).</td>
<td></td>
</tr>
<tr>
<td>inline</td>
<td>Insert an em dash. Note the caret marking the insertion point, a notation that can be used for other insertions as well, such as a single character or a word.</td>
<td></td>
</tr>
<tr>
<td>margin</td>
<td>Use italic type for the marked text.</td>
<td></td>
</tr>
<tr>
<td>inline</td>
<td>Italicize the text that's underlined.</td>
<td></td>
</tr>
</tbody>
</table>
Explaining things: markup language

**Markup language**: a set of rules governing what markup information may be included in a document and how it is combined with the content of the document

Feature - Researching to Add Richness

Sometimes, you're writing nothing but a blistering hot love scene, or a character study, or a drabble. In those cases, your canon is all you'll want or need. But there are other times — Yuletide, a Big Bang, your own original novel-in-progress, that sprawling AU epic that's been taking up all your spare imagination for weeks — when you're going to want the sort of details that add texture and depth to a story.

You may be blessed with a home library that includes just the reference books you want or a local library with sympathetic librarians. If not, there's always the Intarwebs ... but sifting treasure from trash can be a challenge.

Let's take a look at some ways to make the World Wide Web cough up the details that you need to write the story that you want.

*With examples involving the historical novel Frontier Wolf.*
Explaining things: markup

Markup: is needed to structure text to make it understandable (to humans and to computers, but computers of course need more formal logic).

Markups:

Feature - Researching to Add Richness

Sometimes you’re writing nothing but a blistering hot love scene, or a character study, or a drabble. In those cases, you’re canon is all you’ll want or need, but there are other times, Yuletide, a Big Bang, your own original novel-in-progress, that sprawling Au epic that’s been taking up all your spare imagination for weeks when you’re going to want the sort of little details that add texture and depth to a story. You may be blessed with a library that includes just the reference books you want or a local library with sympathetic librarians. If not, there’s always the Interweb... but sifting treasure from trash can be a challenge.

Let’s take a look at some ways to make the world wide web cough up the details that you need to write the story that you want.

With examples involving the historical novel Frontier Wolf.
Explaining things: markup

**Markup**: is needed to structure text to computers, but computers of course need...

Note: the markup example on the right (from the Encyclopaedia Aethiopica) uses the markup convention of the Encyclopaedia Aethiopica. Other books may use e.g. italics for the main title, quotation marks for foreign words, indentation for subtitle, ↑ or bold type for cross-references, etc., etc., etc.

When we speak of books, the markup conventions are usually specific for each publication project (single book, series), publishing house and must be consistent so that the human eye can easily understand what is what = parse the description.
Explaining things: using markup

**Presentational markup**: used for display, as in the examples above. This is the markup invisibly applied the text editors you use. (And you may be surprised to know that Word, OpenOffice and even Photoshop actually apply their markup in the same way, only that you do not see it from the graphic user interface – but it is always there.)

- **Bold**: Manuscripts
- **Plain**: The term M. (from Lat. **Italic**: manu scriptum) ...

**Manuscripts**
The term M. (from Lat. *manu scriptum*) ...

**Descriptive markup**: assigns labels to describe the actual content; such markup can be used both for displaying the data but also for looking for something in particular.

- **Title**: Manuscripts
- **Paragraph**: The term M. (from Lat. **Foreign**: manu scriptum) ...

**Manuscripts**
The term M. (from Lat. *manu scriptum*) ...

Ask computer: look for all **Foreign** elements
Answer: *manu scriptum*
Markup languages

1. Assign descriptions to each word = label (sticky note) = tag

2. Use a predefined vocabulary = set of tags

3. Can be understood (processed) by humans and computers

4. Must be logical and consistent