TEXT ENCODING: AN INTRODUCTION

A tool for research and pedagogy

November 14, 2018

Olivia Wikle
Digital Initiatives Librarian
omwikle@uidaho.edu
Text Encoding

- The process of adding marks (or codes or tags) to text to say specific things about specific parts of the text.

- Text encoding is a way of "marking up" a text.
Whatever our <b>souls</b> are made of, <i>his</i> and <i>mine</i> are the same.
But I, being poor, have only my dreams;
I have spread my dreams under your feet;
Tread softly because you tread on my dreams.
Text Encoding

```html
<!DOCTYPE html>
<html>
  <body>
    <h1>Heading</h1>
    <p>Paragraph.</p>
  </body>
</html>
```
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK AND WEARY
OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE IN ODDED NEARLY NAPPING SUDDENLY THERE CAME A TAPPING
GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TIS SOME VISITOR I MUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AHD DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER, AND EACH SEPARATED DYING EMBER WROUGHT IT
S GHOST UPON THE FLOOR, EAGERLY I WISHED THE MORROW VAINLY I
HAD SOUGHT TO BORROW FROM MY BOOKS SURCEASE OF SORROWS
OR Row FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS THEREFORE EVERMORE
ONCE UPON A MIDNIGHT DREARY
WHILE I PONDERED WEAK AND WEARY
OVER MANY A QUAIN'T AND CURIOUS
VOLUME OF FORGOTTEN LORE
WHILE INODED NEARLY NAPPING
SUDDENLY THERE CAME A TAPPING
GAS OF SOMEONE GENTLY RAPPING
RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR I MUTTERED
TAPPING AT MY CHAMBER DOOR
ONLY THIS AND NOTHING MORE
AND DISTINCTLY I REMEMBER
IT WAS IN THE BLEAK DECEMBER
AND EACH SEPARATED DYING EMER
WEROUGHT IT GHOST UPON THE FLOOR
EAGERLY I WISHED THE MORROW
VAINLY I HAD SOUGHT TO BORROW
FROM MY BOOKS
SURE A SORROW FOR
THE LOST "LENORE"
FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME "LENORE"
NAMELESS HEREFOREVERMOR
ONCE UPON A MIDSUMMER NIGHT DREARY WHILE I PONDERED WEAK AND WEAHEY
OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN Lore
WHILE IN ODDED NEARLY NAPPING SUDDENLY THERE CAME A TAPPLEN
GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR I MUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE. AND DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER, WHEN THE WIND WAS LIKE AN UNHOOLED WIND
SINGING A SONG OF Woe, AND THE SNOW MAN ICOLED WITH THE SNOW
FLOWER. I SAW A BARREN TRAVELLER STANDING IN THE NIGHT,
AND CALLING UPON THE TRAVELLER TO RETURN. I MET A LONE AND
SAD MAN, AND HE CAME TO ME WITH A STRANGE LOOK IN HIS EYES.
I SAW HIM STRUGGLE, AND THEN HE VANISHED. I REMEMBERED THE
PLAY AND HOW THE ROSE OF THE EVENING DIELED, AND THE PHANTOM
OF THE NIGHT.
ONCE UPON A MIDNIGHT DREARY

WHILE I PONDERED WEAK AND WEARY

OVER MANY A QUIET AND CURIOUS VOLUME OF FORGOTTEN LORE

WHILE INODDED NEARLY NAPPING SUDDENLY THERE CAME AT A TAPPING

GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR

TIS SOME VISITOR I Muttered TAPPING AT MY CHAMBER DOOR

ONLY THIS AND NOTHING MORE I DISTINCTLY REMEMBER IT WAS IN THE

BLEAK DECEMBER ANDEACH SEPARATE Dying EMBER WROUGHT IT

GHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY

I HAD SOUGHT TO BORROW FROM MY BOOKS SURCEASE OF SORROWS

ORROW FOR THE LOST LENORE

FOR THE RARE AND RADIANT MAIDEN

WHOM THE ANGELS NAME LENORE NAMELESS HEREFOREVERMORE
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK AND WEARY
OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE IN ODD ED NEARLY NAPPING, SUDDENLY THERE CAME AT APPIN
GAS OF SOMEONE, GENTLY RAPPING, RAPPING AT MY CHAMBER DOOR
TISSOMEVISITORIMUTTEREDTAPPINGATMYCHAMBERDOORONLY
THIS AND NOTHING MORE AHD DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPAREATED DYINGEMBER WROUGHT IT
SGHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY I
HAD SOUGHT TO BORROW FROM MY BOOKS SUR CEASE OF SORROWS
ORROW FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS HERE FOREVERMOR
ONCE UPON A MIDNIGHT DREARY

WHILE I PONDERED WEAK AND WEARY

OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE

WHILE I NODDED NEARLY NAPPING

SUDDENLY THERE CAME A TAPPING

GAS OF SOMEONE GENTLY RAPPING

RAPPING AT MY CHAMBER DOOR

TIS SOME VISITOR IMUTTERED TAPPING AT MY CHAMBER DOOR

ONLY THIS AND NOTHING MORE AH!

DISTINCTLY I REMEMBER IT WAS IN THE

BLEAK DECEMBER, ANDEACH SEPARATED

DYINGember wrought it

S GHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW

VAINLY I HAD SOUGHT TO BORROW FROM MY BOOKS

SUREASE OF SORROWS OR ROW FOR THE LOST LENORE

FORTHERRARE AND RADIANT MAIDEN

WHOM THE ANGELS NAME LENORE

NAMELESS HEREFOREVERMORE
ONCE UPON A MIDNIGHT DREAMY
WHILE I PONDERED WEAK AND WEARY
OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE IN ODDED NEARLY NAPPING SUDDENLY THERE CAME AT APPIN
GAS OF SOMEONE: GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR I MUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AND DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATED DYINGEMBER WROUGHT IT
S GHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY
I HAD SOUGHT TO BORROW FROM MY BOOKS SURCEASE OF SORROWS
ORROW FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS HEREFOREVERMORE
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK AND WEARY
OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE INODDED NEARLY NAPPING SUDDENLY THERE CAME A TAPPING
GASO OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TIS SOME VISITOR IMUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AHD DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATED DYINGEMBER WROUGHT IT
SGHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY
I HAD SOUGHT TO BORROW FROM MY BOOKS SURCEASE OF SORROWS
ORROW FOR THE LOST LENORE FOR OTHER RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS HEREFOREVERMORE
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK AND WEA
RY OVER MANY A QUAIN T AND CURI OUS VOLUME OF FORGOTTEN LORE
WHILE I NODDED NEARLY NAPPING SUDDENLY THERE CAME A TAPPIN
GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR I MUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AH DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATED DYINGEMBER WROUGHT IT
SGHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY I
HAD SOUGHT TO BORROW FROM MY BOOKS SUR CEASE OF SORROWS
OR ROW FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS HEREFOR EVERMORE
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK AND WEARY
OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE INODDED NEARLY NAPPINGSUDDENLY THERECAMEATAPPIN
GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOMEVISITORIMUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MOREAH DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER ANDEACH SEPARETED DYINGEMBER WROUGHT IT
S GHOST UPON THE FLOOR EAGERLY I WISHED THEMORROW VAINLY I
HAD SOUGHT TO BORROW FROM MY BOOKS SURCEASE OF SORROWS
ORROW FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESSHERE FOREVERMORE
ONCE UPON A MIDNIGHT DREARY
WHILE I PONDERED WEAK AND WEA\nRY OVER MANY A QUAIN T AND CURI OUS VOLUME OF FORGOTTEN LORE
WHILE INODDED NEARLY NAPPING SUDDENLY THERE CAME AT A PPIN
GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR IMUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AH DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATED DYING EMBER WROUGHT IT
UNGHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY I
HAD SOUGHT TO BORROW FROM MY BOOKS SURREASE OF SORROWS
ORROW FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS HEREFOREVERMORE
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK AND WEA
RY OVER MANY A QUAIN'T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE I NODDED NEARLY NAPPING SUDDENLY THERE CAME A TAPPIN
G AS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR IMUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AHD DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATED DYINGEMBER WROUGHT IT
SGHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY I
HAD SOUGHT TO BORROW FROM MY BOOKS SUR CEASE OF SORROWS
ORROW FOR THE LOST LENORE FOR THE RARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS THEREFOREVERMOR
ONCE UPON A MIDNIGHT DREARY
WHILE I PONDERED WEAK AND WEA
RY OVER MANY A QUAIN T AND CURIOUS VOLUME OF FORGOTTEN LORE
WHILE I NODDED NEARLY NAPPING SUDDENLY THERE CAME A TAPPING
GAS OF SOMEONE GENTLY RAPPING RAPPING AT MY CHAMBER DOOR
TISSOME VISITOR IMUTTERED TAPPING AT MY CHAMBER DOOR ONLY
THIS AND NOTHING MORE AH DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATED DYING EMBER WROUGHT IT
GHOST UPON THE FLOOR EAGERLY I WISHED THE MORROW VAINLY
I HAD SOUGHT TO BORROW FROM MY BOOKS SURCEASE OF SORROWS
ORROW FOR THE LOST LENORE FOR THEIR ARE AND RADIANT MAIDEN
WHOM THE ANGELS NAME LENORE NAMELESS HEREFOREVERMORE
ONCE UPON A MIDNIGHT DREARY WHILE I PONDERED WEAK
AND WEARY OVER MANY A QUAIN'T AND CURIOUS VOLUME OF
FORGOTTEN LORE WHILE I NODDED NEARLY NAPPING
SUDDENLY THERE CAME A TAPPING AS OF SOMEONE GENTLY
RAPPING RAPPING AT MY CHAMBER DOOR TIS SOME VISITOR I
MUTTERED TAPPING AT MY CHAMBER DOOR ONLY THIS AND
NOTHING MORE AH DISTINCTLY I REMEMBER IT WAS IN THE
BLEAK DECEMBER AND EACH SEPARATE DYING EMBER
WROUGHT ITS GHOST UPON THE FLOOR EAGERLY I WISHED THE
MORROW VAINLY I HAD SOUGHT TO BORROW FROM MY BOOKS
SURCEASE OF SORROW SORROW FOR THE LOST LENORE FOR
THE RARE AND RADIANT MAIDEN WHOM THE ANGELS NAME
LENORE NAMELESS HERE FOREVERMORE
Once upon a midnight dreary while I pondered weak and weary
Over many a quaint and curious volume of forgotten lore While I
nodded nearly napping suddenly there came a tapping As of
some one gently rapping rapping at my chamber door Tis some
visitor I muttered tapping at my chamber door Only this and
nothing more Ah distinctly I remember it was in the bleak
December And each separate dying ember wrought its ghost
upon the floor Eagerly I wished the morrow vainly I had sought to
borrow From my books surcease of sorrow sorrow for the lost
Lenore For the rare and radiant maiden whom the angels name
Lenore Nameless here for evermore
Once upon a midnight dreary, while I pondered, weak and weary,
Over many a quaint and curious volume of forgotten lore— While I
nodded, nearly napping, suddenly there came a tapping, As of
some one gently rapping, rapping at my chamber door. “’Tis
some visitor,” I muttered, “tapping at my chamber door— Only
this and nothing more.” Ah, distinctly I remember it was in the
bleak December; And each separate dying ember wrought its
ghost upon the floor. Eagerly I wished the morrow;—vainly I had
sought to borrow From my books surcease of sorrow—sorrow for
the lost Lenore— For the rare and radiant maiden whom the
angels name Lenore— Nameless here for evermore.
Once upon a midnight dreary, while I pondered, weak and weary,

Over many a quaint and curious volume of forgotten lore—

While I nodded, nearly napping, suddenly there came a tapping,

As of some one gently rapping, rapping at my chamber door.

“'Tis some visitor,” I muttered, “tapping at my chamber door—

Only this and nothing more.”

Ah, distinctly I remember it was in the bleak December;

And each separate dying ember wrought its ghost upon the floor.

Eagerly I wished the morrow;—vainly I had sought to borrow

From my books surcease of sorrow—sorrow for the lost Lenore—

For the rare and radiant maiden whom the angels name Lenore—

Nameless here for evermore.
The Raven
By: Edgar Allan Poe

Once upon a midnight dreary, while I pondered, weak and weary,
Over many a quaint and curious volume of forgotten lore—

While I nodded, nearly napping, suddenly there came a tapping,
As of some one gently rapping, rapping at my chamber door.
“'Tis some visitor,” I muttered, “tapping at my chamber door—
Only this and nothing more.”

Ah, distinctly I remember it was in the bleak December;
And each separate dying ember wrought its ghost upon the floor.
Eagerly I wished the morrow;—vainly I had sought to borrow
From my books surcease of sorrow—sorrow for the lost Lenore—
For the rare and radiant maiden whom the angels name Lenore—
Nameless here for evermore.
MARKUP IS NOT NEW
Electronic Markup

- Standard Generalized Markup Language (SGML)
  - HTML
  - XML
XML

- XML = Extensible Markup Language
  - Both human- and machine-readable
  - International standard, non-proprietary
  - Hardware and software independent

- In XML, tags label portions of text:

  ```xml
  <p>Here he again fell faintly back. Again his mind wandered: but he rallied, and less obscurely proceeded.</p>
  ```

- Tags
  - Elements
  - Attributes
Elements

- Elements are the basic, named structural units of an XML document

```xml
<title>The Odyssey</title>
```
Attributes

- Attributes are name-value pairs (name="value") associated with elements

```xml
<q who="#dumbledore">
    I sometimes find, and I am sure you know the feeling, that I simply have too many thoughts and memories crammed into my mind.
</q>
```
Attributes

attribute ("the type attribute")

<name type="person">Art Garfunkel</name>

attribute name  attribute value
Anatomy of an XML Tag

```
<greeting xml:lang="en">Hello, world!</greeting>
```
Other Features of XML Documents

- Tags are nested
  - Each XML document has one "root" tag

```xml
<text>
  <body>
    <lg>
      <l>But I, being poor, have only my dreams;</l>
      <l>I have spread my dreams under your feet;</l>
      <l>Tread softly because you tread on my dreams.</l>
    </lg>
  </body>
</text>
```
Other Features of XML Documents

- XML Declaration:
  ```xml
  <?xml version="1.0" encoding="utf-8"?>
  ```

- Comments:
  ```xml
  <!-- XML and HTML share the same syntax for comments -->
  ```

- Entities:
  ```xml
  &lt; &gt; &quot; &apos; &amp;
  ```
<?xml version="1.0" encoding="utf-8"?>

<!-- this novel is by Fanny Fern -->

<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started...</p>
    <p>Ruth was not sighing because she was about to leave her father's roof...</p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding...</p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings...</p>
  </chapter>
</book>
<?xml version="1.0" encoding="utf-8"?>

<!-- this novel is by Fanny Fern -->

<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started...</p>
    <p>Ruth was not sighing because she was about to leave her father's roof...</p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding...</p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings...</p>
  </chapter>
</book>
<?xml version="1.0" encoding="utf-8"?>
<!-- this novel is by Fanny Fern -->
<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started... </p>
    <p>Ruth was not sighing because she was about to leave her father's roof... </p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding... </p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings... </p>
  </chapter>
</book>
<?xml version="1.0" encoding="utf-8"?>

<!-- this novel is by Fanny Fern -->

<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started...</p>
    <p>Ruth was not sighing because she was about to leave her father's roof...</p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding...</p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings...</p>
  </chapter>
</book>
<xml version="1.0" encoding="utf-8"?>
<!-- this novel is by Fanny Fern -->
<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started... </p>
    <p>Ruth was not sighing because she was about to leave her father's roof... </p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding... </p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings... </p>
  </chapter>
</book>
<?xml version="1.0" encoding="utf-8"?>

<!-- this novel is by Fanny Fern -->

<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started...</p>
    <p>Ruth was not sighing because she was about to leave her father's roof...</p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding...</p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings...</p>
  </chapter>
</book>
XML in itself is not a vocabulary...

It provides rules (syntax) for structuring encoding
Basic Syntactic Rules of XML

■ Start tags must have corresponding closing tags:
  \(<\text{name}>Harry</\text{name}>\)

■ An XML document must have one and only one root element

■ Attribute values must be quoted with matching single or double quotation marks:
  \(<\text{note} \text{type}="#pencil" \text{place}=\text{margin}>...</\text{note}>\)
Basic Syntactic Rules of XML

- Elements must be properly nested:
  
  ```xml
  <p>That is a <b><i>very</i></b> interesting problem.</p>
  <p>That is a <b><i>very</i></b> interesting problem.</p>
  ```

- Special characters must be escaped:
  
  ```xml
  &lt; for < and &amp; for &
  ```
When a document follows XML’s syntactic rules, that document is considered “well-formed”
Is this XML document well-formed?

```xml
<?xml version="1.0" encoding="utf-8"?>

<memo date="2003-10-24">
    <to><name="person">Ben</name></to>
    <body>Lock down the negative zone!</body>
</memo>
```
<?xml version="1.0" encoding="utf-8"?>

<memo date="2003-10-24">
  <to>
    <name type="person">Ben</name>
  </to>
  <body>
    Lock down the negative zone!
  </body>
</memo>
Encoding Languages

- Comic Book Markup Language (CBML): http://dcl.slis.indiana.edu/cbml/
<q> (quoted) contains material which is distinguished from the surrounding text using quotation marks or a similar mark: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages from sources.

<said> (speech or thought) indicates passages thought or spoken aloud, whether explicitly indicated in the source or not.

@direct may be used to indicate whether the quoted matter is regarded as direct or indirect speech.

@aloud may be used to indicate whether the quoted matter is regarded as having been vocalized or signed.

<quote> (quotation) contains a phrase or passage attributed by the narrator or author to some agency external to the document.

att.global.source provides an attribute used by elements to point to an external source.

@source specifies the source from which some aspect of this element is drawn.

<cit> (cited quotation) contains a quotation from some other document, together with a bibliographic reference.

one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.

<mentioned> marks words or phrases mentioned, not used.

<soCalled> contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility,
Connecting a document to a vocabulary

```xml
<?xml version="1.0" encoding="utf-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">

<memo date="2003-10-24">
  <to><name type="person">Ben</name></to>
  <body>Lock down the negative zone!</body>
</memo>

</TEI>
```
Schemas

- A language for expressing constraints about XML documents
  - *A schema declares the elements and attributes and the allowable content for those elements and attributes.*
<table>
<thead>
<tr>
<th>Apples</th>
<th>Bananas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Name:** African Coffee Table
- **Width:** 80
- **Length:** 120
<h:table>
  <h:tr>
    <h:td>Apples</h:td>
    <h:td>Bananas</h:td>
  </h:tr>
</h:table>

<f:table>
  <f:name>African Coffee Table</f:name>
  <f:width>80</f:width>
  <f:length>120</f:length>
</f:table>
Valid XML Document:

- **Correct Vocabulary**: all elements and attributes actually exist in the specified encoding language

- **Correct Grammar**: elements are used in the right place, in the right order
Different Types of Encoding:

- **Presentational**
  - *Concerned with formal properties of a text*

- **Descriptive**
  - *Concerned with content of a text*
  - *Describes what the text is, not how it should be formatted*
This is heathen Greek to you still? Your *lapis philosophicus*?

She sat down to read Hardy’s *Jude the Obscure*.

A pretty common case, I believe; in all *vehement* deatings.
This is heathen Greek to you still? Your <foreign xml:lang="la">lapis philosophicus</foreign>?

She sat down to read Hardy’s <title>Jude the Obscure</title>.

A pretty common case, I believe; in all <emph>vehement</emph> debatings.
<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. Ruth started...</p>
    <p>Ruth was not sighing because she was about to leave her father's roof...</p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in Ruth to have no one but the family at the wedding...</p>
    <p>"Well, I hope Harry will be happy," said Ruth's mother-in-law, old Mrs. Hall, as she untied her cap-strings...</p>
  </chapter>
</book>
Ruth Hall: A Domestic Tale of the Present Time

The old church clock rang solemnly out on the midnight air. Ruth started...

Ruth was not sighing because she was about to leave her father’s roof...

It was so odd in Ruth to have no one but the family at the wedding...

"Well, I hope Harry will be happy," said Ruth’s mother-in-law, old Mrs. Hall, as she untied her cap-strings...
<xml version="1.0" encoding="utf-8"?>
<!-- this novel is by Fanny Fern -->

<book>
  <title>Ruth Hall: A Domestic Tale of the Present Time</title>
  <chapter>
    <title>Chapter I.</title>
    <p>The old church clock rang solemnly out on the midnight air. <name>Ruth</name> started... </p>
    <p><name>Ruth</name> was not sighing because she was about to leave her father's roof... </p>
  </chapter>
  <chapter>
    <title>Chapter II.</title>
    <p>It was so odd in <name>Ruth</name> to have no one but the family at the wedding... </p>
    <p>"Well, I hope <name>Harry</name> will be happy," said <name>Ruth's</name> mother-in-law, old <name>Mrs. Hall</name>, as she untied her cap-strings... </p>
  </chapter>
</book>
Transforming XML
Uses of Text Encoding

■ Digital Scholarly Editions
■ Research (text analysis, data visualization)
■ Teaching
Uses of Text Encoding: Digital Scholarly Editions

- Goals and Motivations:
  - Textual scholarship
  - Preservation
  - Access
  - Interpretation & reading
  - Computational manipulation and analysis
Uses of Text Encoding: Digital Scholarly Editions

- William Blake Archive: http://www.blakearchive.org
- Walt Whitman Archive: https://whitmanarchive.org
- Dante Gabriel Rossetti Archive: http://www.rossettiarchive.org
- Mark Twain Project Online: http://www.marktwainproject.org
- Petrarchive: http://dcl.slis.indiana.edu/petrarchive/
Ah, not this granite, dead and cold!
Far from its base and shaft expanding—
round zones circling, comprehending;
Thou, Wchi smallcaps
ASHINGTON, art all the worlds, the continent's entire—
alone, America;
Europe's as well, in every part, castle of lord or laborer's cot,
On frozen North, or sultry South—
his tent—
Old Asia's there with venerable smile, seated amid her ruins;
Greets the antique the hero new? 'tis but the same—
same—
The indomitable heart and arm-proofs of never-broken line,
Courage, alertness, patience, faith, the same—
Wherever sails a ship, or house is built on land,
Through teeming cities' streets, indoors or out,
Now, or to come, past where patriot will,
Wherever Freedom, poised by Tolerations, swayed by Law,
Stands or is rising thy true monument.
Digital Edition: Petrarchive

Petrarchive
An edition of Petrarch’s songbook
Rerum vulgarium fragmenta
Edited by H. Wayne Storey, John A. Walsh and Isabella Magri

Petrarchive: http://dcl.slis.indiana.edu/petrarchive/visindex_2up.php
Digital Edition: Petrarchive
Ardente uertute ornata et calda.

Alma gentil chui tante carte uergo.

O sol gia donestate intero albergo.

Torre in alto valor fondata et salda.

O fiāma / o rose sparse in dolce falda

Di iuua nueva / ì chio mi specchio ḟ tergo.

O piacer. onde lali al bel uiso ergo /

Che luce soura quanti il sol ne scald.

Del uostro nome se mie rime intese

Fossin si lunghe / aurei pien Tyle ḅ battro.

La tana el nilo . athlante. olīpo. ḕ calpe.

Poi che portar nol posso in tutte ḕ quattro

Parti del mōdo / Vdrallo il bel paese.

Chappenin parte. el mar circonda ḕ lalpe.
Calpe.


Chappenin parte. El mar circonda. Lalpe.

Alma gentil chui tante carte uergo.
Torre in alto valor fondeta et salda.

Di uiua neue / Ì chio mi specchion tergo.
Che luce soura quanti il sol ne scaldà.

Fossin si lunge / aurei pien Tyle 7 battro.
Poi che portar nol posso in tutte 7 quattro
Uses of Text Encoding: Research

Dates
<date when="1837-04-05">April the 5th, 1837</date>

People
<persName>Joe Vandal</persName>

Places
<placeName xml:id="pal">The Palouse</placeName>
Uses of Text Encoding: Research

- Text Encoding is Never Simple:
  - Text encoding is not simple data entry: it is part of research
  - Text encoding is not neutral or objective
  - Text encoding is a strategic representation of the text

- Flanders, Julia. "Overview of Descriptive Markup and the TEI." *Intensive Introduction to TEI*, GSLIS, University of Illinois at Urbana-Champaign, 2011.
  http://wwp.northeastern.edu/outreach/seminars/_current/presentations/overview/overview.xml
Research: Comic Book Readership Archive
Dear Mary, Sal, & Jim:

Naturally, the question of the month is: What's going to happen in SPIDER-MAN #200? Whatever it is, it's sure to be a landmark. For the first time in years, I'm totally in the dark about what's going on — despite all the advance publicity! Rather than try to guess, I'm going to focus on an issue which many people will forget in the rush... SPIDER-MAN #196. First of all, note that it's well-crafted, hardly dull, and enhanced by excellent artwork (those underwater scenes make me wonder if Sal and Jim wouldn't be the perfect team for the new SUB-MARINER book you keep talking about). Yet, at the same time, what really happened this issue? Mysterio may as well has shot Spider-Man with that dart last issue! But there's the catch: because no mysteries were cleared up in issue #196, because Spider-Man had taken another frustrating when he found out that the trap had endured was all an illusion). SPIDER-MAN #196 will result in nearly twice as good a story in issue #200. In the great Marvel tradition, you've made the reader empathize with our heroes, so that his discouragement at being unable to catch Mysterio becomes the reader's as well. Very subtle, but a big plus. #200 did more than just mark time until #200.

Before I forget, I'd like to compliment John Byrne and Terry Austin on this year's SPIDER-MAN ANNUAL. It was great to see Doctor Octopus back. He was the nice old guy who courted Aunt May, but once more rotten to the core. It was the best annual for Spider-Man since 1965, and you deserve a lot of praise for it.

Matt Kaufman
383 West Pennsylvania
Urbana, IL 61801

reason why this cover looked so good was that there was a good Inker embellishing Keith's pencils, namely Pablo Marcos. Keith's (and Sal's) interior work has been so good to say the least, and Jim's inks don't really work all that well with either of those artists. All I can see is that either you give the book a good Inker like Pablo — or maybe Bob Pod and the really looked good over Keith on the cover (or the cover of the newest annual) — or finally make John Byrne drop the X-lein so that he can take over this book (he also looked great in the annual). Unless something phenomenal happens — like a miracle — next year it's a real shame to think this book will be as exciting as it has been this year... and considering that this book is Marvel's mainstay, I can only reiterate: it's a shame.

Kurt Olinson
1194
Cortland College
Water Valley, ME 6190

It would appear that we have a bit of a disagreement here. Mr. Kaufman enjoyed SPIDER-MAN #196. Mr. Olinson thought it up. Kurt doesn't think there's been enough. And both gentlemen disagree with each other on the quality of issue #199's art (though they both seem to have liked the art on this year's annual). One issue, two very different opinions. But these, that's what makes horses race... and comicbooks, it would seem.

Dear Matt,

SPIDER-MAN #198 was definitely one of the most exciting issues I have read in a long time. Mysterio had to

Research: Comic Book Readership Archive
Why was text encoding helpful?

- Multiple names (Spider-Man, Spidey, Web Head, Peter Parker, Petey; Mary Jane, MJ; Doctor Octopus, Otto Octavius, Doc Ock)
- Separating incoming letters from replies and other editorial content
- Distinguishing Spider-Man, the character, from Spider-Man, the comic book
Research: Comic Book Readership Archive

- <div type="group">
  - <div type="letter">
    - <div type="reply">
      - <persName key="spider-man">Spidey’s</persName>
      - <persName key="mj">M.J.'s</persName>
      - <persName key="len_wein">Len Wein</persName>
      - <persName key="steve_ditko">Steve Ditko</persName>
    </div>
    - <bibl>
      <title>Spidey</title>
      <biblScope unit="issue">#130</biblScope>,
      <biblScope unit="page">page 23</biblScope>,
      <biblScope unit="panel">panel 3</biblScope>
    </bibl>
  </div>
</div>
Dear <orgName>Marvel</orgName>,

With #issue #174, "The Amazing <persName>Spider-Man" improved 100%!

I always did like the <persName>Punisher</persName>, and when <persName>Hitman</persName> made his appearance in <persName>Spider-Man</persName>, I hoped that the two would meet.

Len forgot all about melodrama in this issue and concentrated on fast-moving action with just the right amount of plot. It was great.

Now a few notes to Len:

It's about time that Pete and Mary Jane either got together or split. I'm sick and tired of "Pete runs off, Mary gets mad, Pete explains, Mary accepts, but stays mad." It's boring!

Since Len hasn't talked to Pete since #168, are we through with her?

Maybe you should introduce a new girl to be Pete's new love? Not a Gwen and not a Betty and not a Liz and not an MJ. Somebody totally new.

Do you realize the entire supporting cast has dwindled to nothing. Let's get some new blood!

Final thought: Let's have some real continuity between "Marvel Team-Up" and this book. Like a three-book cross-over?

Just keep up the good work Len!

Tony R. Rose

Box 14A Dooley Rt.
Boles, AR 72926
Uses of Text Encoding: Teaching
Uses of Text Encoding: Teaching

El Eternauta,
- Héctor G. Oesterheld,
Solano López
Uses of Text Encoding: Teaching

<text>
  <front>
    <div type="titlePage" xml:id="eg_0001">
      <head>el eternauta</head>
      <cbml:caption rendition="#uc #center">héctor g. oesterheld / solano lópez</cbml:caption>
    </div>
  </front>
</text>
Operación Bolívar,
- Edgar Clement
Narrator is seen outside a large colonial, baroque door. May be an intervened photograph of this entrance.

Panels 2 through 6 are contained inside large panel #1. Caption for panel #3 is located below the image, while the one for #4 is inside the panel, next to the image. Captions for panels #5-6 start inside the box for #5 but cross into the space for #6.
"... one of the oldest strains of digital humanities, text encoding, is fundamentally an exercise in close reading"

"... the exercise defamiliarizes the act of reading, reveals its continuity with digital text mining, and offers insights that may not solely exist in either realm"

Activity: Creating Well-Formed XML

- Design your own meaningful tags (elements and attributes) to describe the content of these documents:
**French Cranberry Chicken**

1 cup Soup Of Success French Crème Dressing (prepared)
1 cup whole cranberry sauce
1 package dry onion soup mix
1 1/2 pounds of boneless chicken breast

Place chicken in a 9 X 13 dish.
Mix dressing, cranberry sauce, and soup mix and pour over chicken.
Bake at 350 degrees for 45 minutes.
Serve over rice and enjoy.
XML Validator

Encode your texts here: https://tinyurl.com/y8hskjgf

Check for well-formedness here: https://codebeautify.org/xmlvalidator
Resources

- Women Writers Project: Resources for Teaching and Learning Text Encoding. [http://wwp.northeastern.edu/outreach/resources.html](http://wwp.northeastern.edu/outreach/resources.html)
