CENTER FOR TEXTUAL STUDIES AND
DIGITAL HUMANITIES

Introduction to Digital Humanities Research & Computing

Fall Semester 2015

Week 10

Weekly Exercise - Example Solution 1

```
<?php
$march = 31;
$april = 30;
$july = 31;
$august = 31;
$december = 31;
$subtotal;

function addDays() {
global $march, $april, $july, $august, $subtotal;
$subtotal = $march + $april + $july + $august;
}

function minusDays($days) {
global $subtotal, $december;
$subtotal = $days - $december;
}

//run function to calculate adding days
addDays();

//output first subtotal to check addition function
echo '<p>first subtotal = '.$subtotal.'</p>';

//run function to calculate subtracting days
minusDays($subtotal);

//output result of calculations
echo '<p>final subtotal = '.$subtotal.'</p>';
?>
```

- how can we improve this solution?
- any limitations?
- abstraction issues?

Demo

Weekly Exercise - Example Solution 2

```
<?php
$subtotal;

function addDays() {
global $subtotal;
$march = 31;
$april = 30;
$july = 31;
$august = 31;
$subtotal = $march + $april + $july + $august;
}

function minusDays($days) {
global $subtotal;
$december = 31;
$subtotal = $days - $december;
}

/**run function to calculate adding days**/
addDays();

/**output first subtotal to check addition function**/
echo '<p>first subtotal = '.$subtotal.'</p>';

/**run function to calculate subtracting days**/
minusDays($subtotal);

/**output result of calculations**/
echo '<p>final subtotal = '.$subtotal.'</p>';
?>
```

- what's wrong with this solution?
- what are its limitations?
- abstraction?

Demo

Text Encoding

Overview

- practical value and importance to DH
- often considered within the context of critical editions, indices, concordances...
- can also be considered with the broader context of 'new media'...

Text Encoding

Markup

'information formally distinct from the character sequence of the digital transcription of a text, which serves to identify logical or physical features or to control later processing.'

- distinct from the text itself, serves to identify logical or physical features or to help with later processing
- unfamiliar expressions or codes
- considered within context of computer based typesetting and text processing
- 1960s to 80s typesetting and text processing offers foundation

Text Encoding

Markup - descriptive in nature

- encoding was initially specific to an application using
    - codes for individual characters of the text
    - & codes for formatting commands
- descriptive markup became seen as the fundamentally correct approach
- descriptive said to identify and describe the parts of a document
- procedural was a command or instruction invoking formatting
- logical v graphical

Text Encoding

Descriptive Markup - Advantages Part 1

- simplified composition
- edit with structure-oriented editors
- generated alternative document views are supported

Text Encoding

Descriptive Markup - Advantages Part 2

- generic formatting can be specified and easily modified
- textual apparatus can be automated
- enhanced output device support
- portable and interoperable

Text Encoding

Descriptive Markup - Advantages Part 3

- support for information retrieval
- support for analytical procedures

Text Encoding Initiative

Consortium

- maintains a technical standard
- set of guidelines
- user wiki
- set of tools for development and processing
- currently predominant in social sciences and humanities, in particular
    - textual studies
    - literary studies
    - linguistic studies

Text Encoding Initiative

A gentle introduction to the TEI Guidelines

- guidelines for electronic text encoding and interchange
- define and document a markup language for representing
    - structural
    - rendition
    - logical & semantic
    - analytic
- rules and recommendations rather than standard
- wide variety of possible solutions for encoding material
- freedom of expression of personal textual theory
- current version = P5
- over 513 elements, 214 attributes...
- customisations such as TEI Lite

Text Encoding Initiative

A gentle introduction to the Guidelines

- structural features
    - organisation of information in text
- TEI defines overall text structure using
    - front
    - body
    - group
    - back
- a consideration of rendition features

Text Encoding Initiative

A gentle introduction to the Guidelines

- rendition features such as
    - distinct fonts
    - colours
    - alignments
    - italics, underline, bold
    - font weight...
- highlighting

Malory Example

<hi>italic words...</hi>

<hi rend="italic">italic words...</hi>

Text Encoding Initiative

A gentle introduction to the Guidelines

- logical and semantic features such as
    - emphasis
    - foreign words
    - linguistically distinct words, phrases
- we can also consider
    - quotation marks
    - quotes
    - cited quotation...

Text Encoding Initiative

A gentle introduction to the Guidelines

- analytical features such as
    - notes and comments
    - marking for indexing
    - regularisation
    - editorial statements

Introduction to XML

- XML = EXtensible Markup Language
- markup language similar to HTML
- designed to carry data but not display data
- XML tags are not pre-defined, you must define your own
- designed to be self-descriptive
- XML is a W3C recommendation

Introduction to XML

```
<calendar>
<date>23rd September 2014</date>
<title>DIGH 401</title>
<location>CTSDH, Loyola Hall...</location>
</calendar>
```

Introduction to XML

- separation of data from display
- simplifies data sharing and transport
- simplifies and eases transition to new platforms and upgrades
- XML becomes more accessible
- XML has also been used to create new languages

XML Structure

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<calendar type="personal">
<date>20th September 2012</date>
<title>DIGH 401</title>
<location>Room 201 Cudahy Library</location>
</calendar>
```

XML Structure

Syntax

- there must be closing tags
  eg: a new paragraph... and not
  a new paragraph...

- tags are case-sensitive
- elements must be properly nested
  eg: <bold><italic>a new phrase...</italic></bold> and not
  <bold><italic>a new phrase...</bold></italic>

- must have a root element
- attribute values must be quoted
- entity references
  eg: using a character such as < instead of <

XML Structure

Entity References

<	<	less than
>	>	greater than
&	&amp;	ampersand
'	&apos;	apostrophe
"	&quot;	quotation mark

XML Structure

Syntax

- comments

```
<!-- this is a comment in XML -->
```

XML Structure

Elements

```
<library>
  <book category="fiction">
    <title>To the Lighthouse</title>
    <author>Virginia Woolf</author>
    <year>1927</year>
  </book>
</library>
```

- everything from the start to end tag of an element, inclusive of the tags
- can contain
    - other elements
    - text
    - attributes
    - or a mix of all of the above...

XML Structure

Elements

- Naming rules
    - elements can contain letters, numbers, and other characters
    - elements cannot start with a number or punctuation character
    - elements cannot start with the letters xml (or XML, or Xml, etc)
    - elements cannot contain spaces
- no words are reserved considering the above rules
- best practice examples for element naming
- XML elements are extensible

XML Structure

Attributes

<book type="print">Hannibal's Footsteps</book>

- additional information about the element
- attribute values must be quoted
- double or single quotes
- no strict rules when to use attribute and when to use element
- some issues with attributes are:
    - cannot contain multiple values (elements can)
    - cannot contain tree structures (elements can)
    - not easily expandable (for future changes)
- attribute values for metadata

<book id="4024">A Good Year</book>

XML Viewing

- view without styling in a web browser, editor... - DEMO

- render with CSS by associating a stylesheet with a given XML file - DEMO

- transform the document using XSLT

- access, query and manipulate XML using Javascript

- parse XML and render using a language such as PHP
    - PHP & Javascript - DEMO

XML Tests

Think about how you might encode the following information:

- a car

- a sports team / orchestra

- 2 Musical CDs including each song per album per CD (each album has 5 songs)