FRESNO STATE
Discovery. Diversity. Distinction.

Web Programming (CSci 130)

Department of Computer Science
College of Science and Mathematics
California State University Fresno
H. Cecotti

Learning outcomes

- From printed documents, to structured & web documents
  - Printed documents
  - Markup languages
  - Web
  - XML/HTML

Printed documents (1)

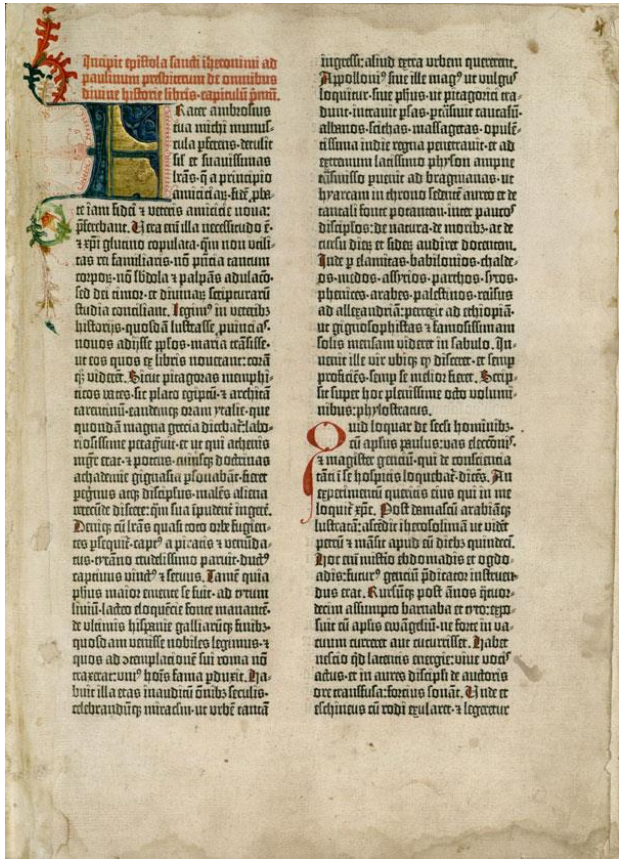
- Before going to Web documents, what is a document?
  - Books
  - Magazines
  - Article
    - Scientific article, newspapers,…
    - No default input from the user but the possibility to add annotations, notes
      - (Meta data)
  - Forms
    - Questionnaire, bank cheque,…
    - Input from the user (pen)

Printed documents (2)

- Structured documents
  - Title, Abstract, Chapters, Sections, Subsections, paragraphs, list of items, equations, tables, figures…
- Access
  - Linear (you start from the first page)
  - Index (table of contents, table of figures, glossary)
    - Page (access the **physical** location)
  - With some jumps: references.
    - In some old books (e.g. role playing game)
- Links between:
  - The way a document is accessed
  - The way a document is created
    - Introduction, Rationale, Arguments, Examples…

Printed documents (3)

- Examples:



REVIEWED ON PC | 24 AUG 2018 6:14 PM PDT

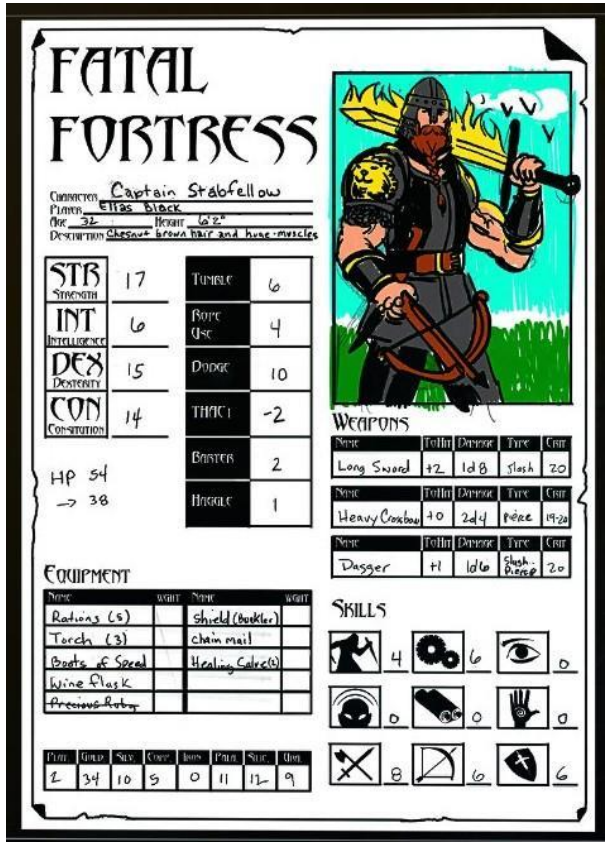
WORLD OF WARCRAFT: BATTLE FOR AZEROTH REVIEW

Share. The battle is joined in Blizzard's latest content-packed expansion.

BY JOSEPH BRADFORD → The Horde vs. Alliance struggle has always sat at the heart of World of Warcraft, although it has at times faded into the background while the world banded together for a common cause. But now that the world-destroying threat from 2016's excellent Legion expansion is diminished, Blizzard has once again rekindled the burning hatred these factions have for one another, bringing a welcome refocus on our character's epic storylines and those of some of Warcraft's greatest characters in the new Battle for Azeroth expansion. Legion left behind some pretty big shoes to fill, and while Battle for Azeroth doesn't bring many of sweeping changes like its predecessor, it does feel right at home in the Warcraft story.

Battle for Azeroth picks up right after the major events of Legion's last big update, which saw the Titan Sargeras imprisoned - but not before he plunged a sword into Azeroth itself. The resulting wound caused the planet to begin bleeding its "life-blood," which solidified into a substance now known as Azerite. This surprisingly plentiful substance enhance the desires and abilities of those who possess it, making it an incredibly valuable resource and raising tensions between the factions as they vie for its control. The stakes rise even higher as you realize that Azeroth herself is dying thanks to the giant blade embedded into the heart of Silithus.

The pre-expansion events set the stage for the conflict in Battle for Azeroth, though the stakes of the new questing pale in comparison to Legion's world-ending scenarios. In one of their more sinister and memorable acts, the burning of Teldrassil perfectly sets the stage for the struggle to come. Like many other Alliance, I couldn't wait to hop into Darkshore with my Pandaren Monk and make them pay. In comparison to previous expansions, the carnage seemed personal.



Documents on your computer (1)

- Notepad: just text
- MS Word
  - Text + Presentation
- Still many people don't know how to use MS Word properly
  - **Key issues**:
    - Going to the next page: press enter until the new page vs. insert a line break
    - Adding a space between a section title and paragraph vs. changing the style of the section title
    - Figure in the text: writing the figure number directly vs. insert a cross reference
  - Results: you can have the correct presentation **but** with no internal logical meaning
    - It **looks** correct but the way it has been created is not correct!
      - The presentation is tricking the structure

Documents on your computer (2)

- What does it mean …
  - For a piece of text to be bold? Important?
  - For some text to have a font of size 14? A title?
- It is a key issue because
  - The document is more difficult to **maintain**
  - Modifications will have a key impact on the document
    - Example:
      - Addition/deletion of a new figure
      - Addition/deletion of a new reference (bibliography)
- In a document, the choice of the font, style, the structure…
  - Nothing is innocent and it must be carefully thought
- Separation between the logical part from the presentation
  - Consistency: it is a title of section, and all the titles will be presented the same way
  - “bla bla” is a title with some properties for its presentation (size, font,…)
  - “bla bla” is presented in a particular way (size, font,…)
    - **You** see it as a title because of its presentation **but** the computer does not know it is a title !!

Documents on your computer (3)

- The separation between the logical and layout structure can be tricky
  - Printed documents: you look at it and you know what is what
    - Based on experience, visual processes…
      - Big = important (title) , bold and rare = important
      - Small = detail (footnote)
  - Web documents: must be understood by the computer
    - A need to think logical/layout
- The challenges
  - Not to write the “content of the document” (e.g. a news)
  - But to generate the logical and layout structure that will allow to accept the content
    - From a user interface
    - From a database

Markup languages (1)

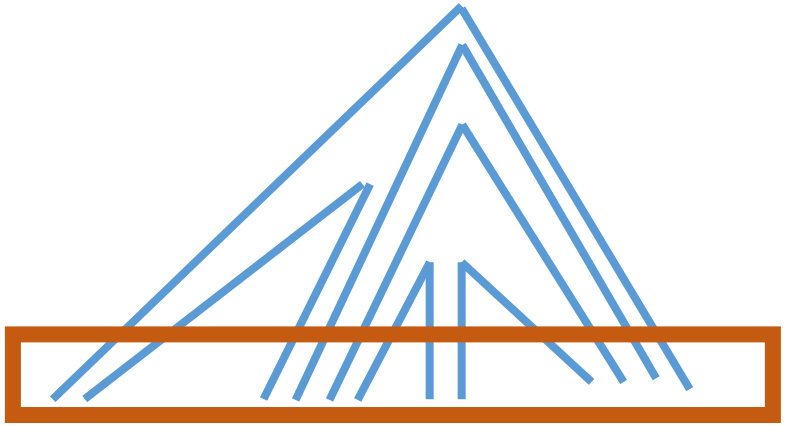
- A system for annotating a document
  - Markup ← marking up
  - Paper/pen: to underline some text, to put a frame around a piece of text
    - To tell what is what (primary school: subject-verb-complement in a sentence)
    - To get information about the structure of the document (logical and/or layout)
  - Beginning/end of a block: tag
    - Block X is composed of block(s) $Y_1, Y_2, \ldots Y_N$
      - Named elements delimited by angle brackets
    - Example:
      - LaTeX: \begin{equation} … \end{equation}
      - HTML: …

Correct (no overlapping lines)

... continually amazed ...

... continually amazed ...

Incorrect (the sets of tags cross over each other)



ABBCBDDEEBCA

Sequence of tags

Markup languages (2)

- Different types:
  - Presentation
    - Type of font, font size,…
  - Procedural (to process the text)
  - Descriptive (logical markup)
    - Description of the text in a logical way
      - Title, sections, subsections, citations, references to figures…
- Different level of complexity
  - TeX → Wikipedia
    - https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Layout

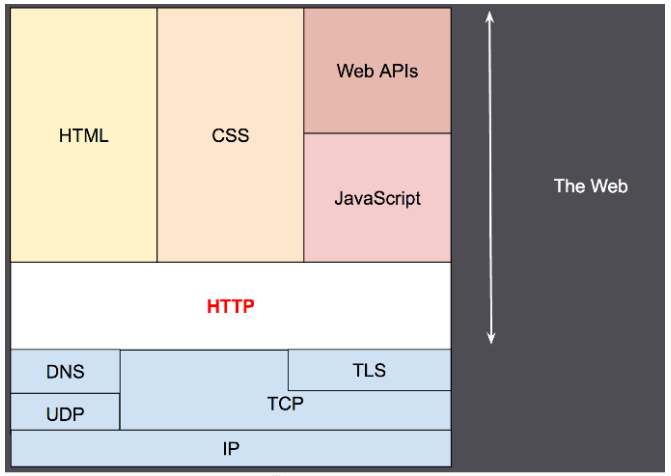
The Internet

- A massive network of networks, a networking infrastructure
- Connection of millions of computers/devices together globally
  - A large network where any computer can communicate with any other computer as long as they are connected to the Internet
- The information travels via protocols:
  - Internet Protocol (IP)
  - Example: Transmission Control Protocol (TCP) (TCP-IP)
- For the Web, email, ftp…

The Web

- The Web is not the Internet
  - A part of the Internet
- The World Wide Web (or Web)
  - A way of accessing information over the Internet
  - Information sharing model at an upper layer of the Internet
- HTTP protocol (client-server protocol)
  - To transmit data
    - To fetch resources (e.g., HTML documents)
    - Application layer protocol that is sent over
      - TCP
      - a TLS-encrypted TCP connection
  - Extensible
    - HTTP headers
      - New functionalities can be introduced

      by a simple agreement between a client and a server about a **new header's semantics**



Hypertext Transfer Protocol

- HTTP (**H**yper**t**ext **t**ransfer **p**rotocol)
  - An **application** protocol for:
    - Distributed, Collaborative, Hypermedia information
  - To exchange or transfer hypertext
- All the addresses start with http (or https)
- Client-server computing model
  - Request – Response (synchronous)
  - Client initiates the request!

Hyperlinks

- Hyperlinks for the creation of the World Wide Web
  - **Tim Berners Lee**: to use hyperlinks to link information to any other information over the internet
  - A reference to data that the user can access by:
    - Clicking
    - Tapping
    - Hovering (e.g. mouse coordinate in within a predefined surface)
  - Pointing to: a new document, a specific element in a document (can be in the same document)
  - Anchor text: the element of text that is used to represent the link
    - Does not need to be a string representing the link (URL)
  - Inline linking:
- Hypertext = a text with hyperlinks

Hypertext

- A text with hyperlinks to other texts
- Hypertext documents
  - Text connected with hyperlinks
- → Hypermedia: media connected with hyperlinks
- Concept coins in 1963 (Ted Nelson)
  - Hypertext system (project Xanadu)

Web documents

- Webpage (HTML)
- Linked between each other via hyperlinks
- Client side
  - Accessed by the user by browsers
    - Firefox, Chrome, MS Edge,…
- Server side
  - The browser retrieve the webpage remotely from web server, or from a local storage (local hard drive).

HTML

- Hypertext Markup Language (HTML)
  - ➢Standard markup language
    - Webpages, web applications (web client-server software application)
  - ➢A document-layout and hyperlink-specification language
- At the beginning,
  - ➢HTML describes the logic structure and the layout ☹
- To be combined with Cascading Style Sheets (CSS) and Javascript

Evolution of HTML

- In the SGML family
  - HTML v1: 1989
  - HTML v2: 1995
  - HTML v3: winter 1997
  - HTML v4: fall 1997

- The main change:
  - HTML5: 2014
  - HTML5.1: 2016

- There is a cousin in the XML family:
  - XHTML

XML (Extensible Markup Language)

- Before XHTML, we need XML…
  - Started in 1996
- Extension of SGML (Standard Generalized Markup Language)
  - Too broad: to define all the types of document
  - Too generic → too useless ☹
  - HTML used a **subset** of the definition of SGML
- Meta markup language
  - Text documents/data (no XML for images)
- HTML made its own way for the web but it was still needed to have a generic way to describe documents

XML (1)

- Early 200s, XML revolution :/
- Warning
  - XML does not replace HTML
    - HTML can be generated from XML
  - It is NOT a presentation format
    - It is possible to convert XML data into a format to be presented
  - It is not a network protocol to transfer data
  - It is not a database

XML (2)

- Useful but it depends on how it is used
  - Portable data
  - Readable by human users (non-experts)
  - Semantics embedded in the data
  - Flexible, customizable
- XML transformation languages
  - Conversion
  - Data exchange between different systems
    - System A file → XML file 1
    - XML file 1 → XML file 2
    - XML file 2 → System B file

XML Structure

- How to define the structure of an XML document
  - To define a structured document, you need a document that explains how to do it :D
    - That is also a structured document … an XML document should explain the XML document
- Two possibilities
  - **DTD (Document Type Definition)**
    - From SGML syntax
    - No datatypes, no namespace
    - No order of the child of an element
    - Limited as not extensible
    - A particular syntax (it can be hard to read)
  - **XSD (XML Schema Definition)**
    - Written in XML (to stay in the logic: a document describing a document)
    - Supports datatypes for elements, attributes(e.g. string)
    - Order of the child elements
    - Extensible (it is XML)
    - XML tags: easier to read and interpret
    - More control on the XML structure

XML - DTD

- **Document type definition** (**DTD**)
  - A set of markup declarations that define a document type for an SGML-family markup language
    - SGML, **XML**, HTML
  - It defines:
    - the legal building blocks of an **XML** document.
    - the document structure with a list of legal elements and attributes.

XML - DTD

- **PCDATA**
  - Text that will be parsed by a parser.
  - Tags inside the text will be treated as markup and entities will be expanded.
- **CDATA**
  - Text that will **not** be parsed by a parser.
  - Tags inside the text will not be treated as markup and entities will not be expanded.

DTD - Examples (1)

- TV schedule

```
<!DOCTYPE TVSCHEDULE [

<!ELEMENT TVSCHEDULE (CHANNEL+)>
<!ELEMENT CHANNEL (BANNER,DAY+)>
<!ELEMENT BANNER (#PCDATA)>
<!ELEMENT DAY (DATE,(HOLIDAY|PROGRAMSLOT+)+)>
<!ELEMENT HOLIDAY (#PCDATA)>
<!ELEMENT DATE (#PCDATA)>
<!ELEMENT PROGRAMSLOT (TIME,TITLE,DESCRIPTION?)>
<!ELEMENT TIME (#PCDATA)>
<!ELEMENT TITLE (#PCDATA)>
<!ELEMENT DESCRIPTION (#PCDATA)>

<!ATTLIST TVSCHEDULE NAME CDATA #REQUIRED>
<!ATTLIST CHANNEL CHAN CDATA #REQUIRED>
<!ATTLIST PROGRAMSLOT VTR CDATA #IMPLIED>
<!ATTLIST TITLE RATING CDATA #IMPLIED>
<!ATTLIST TITLE LANGUAGE CDATA #IMPLIED>
]>
```

DTD - Example (2)

- Newspaper

```
<!DOCTYPE NEWSPAPER [

<!ELEMENT NEWSPAPER (ARTICLE+)>
<!ELEMENT ARTICLE (HEADLINE,BYLINE,LEAD,BODY,NOTES)>
<!ELEMENT HEADLINE (#PCDATA)>
<!ELEMENT BYLINE (#PCDATA)>
<!ELEMENT LEAD (#PCDATA)>
<!ELEMENT BODY (#PCDATA)>
<!ELEMENT NOTES (#PCDATA)>

<!ATTLIST ARTICLE AUTHOR CDATA #REQUIRED>
<!ATTLIST ARTICLE EDITOR CDATA #IMPLIED>
<!ATTLIST ARTICLE DATE CDATA #IMPLIED>
<!ATTLIST ARTICLE EDITION CDATA #IMPLIED>

<!ENTITY NEWSPAPER "Vervet Logic Times">
<!ENTITY PUBLISHER "Vervet Logic Press">
<!ENTITY COPYRIGHT "Copyright 1998 Vervet Logic Press">

]>
```

XSD – Example (1)

- XML document

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

<shiporder orderid="889923"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="shiporder.xsd">
  <orderperson>John Smith</orderperson>
  <shipto>
    <name>Ola Nordmann</name>
    <address>Langgt 23</address>
    <city>4000 Stavanger</city>
    <country>Norway</country>
  </shipto>
  <item>
    <title>Empire Burlesque</title>
    <note>Special Edition</note>
    <quantity>1</quantity>
    <price>10.90</price>
  </item>
  <item>
    <title>Hide your heart</title>
    <quantity>1</quantity>
    <price>9.90</price>
  </item>
</shiporder>
```

XSD – Example (2)

- With the corresponding XSD
  - Which is an XML document
- Comparison with OOP
  - XSD: name of classes, property names
  - XML: name of objects and their properties
- Comparison with databases
  - XSD: name of the tables, and columns in tables
  - XML: name of the cells in the table

```
<?xml version="1.0" encoding="UTF-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="shiporder">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="orderperson" type="xs:string"/>
      <xs:element name="shipto">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="name" type="xs:string"/>
            <xs:element name="address" type="xs:string"/>
            <xs:element name="city" type="xs:string"/>
            <xs:element name="country" type="xs:string"/>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
      <xs:element name="item" maxOccurs="unbounded">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="title" type="xs:string"/>
            <xs:element name="note" type="xs:string" minOccurs="0"/>
            <xs:element name="quantity" type="xs:positiveInteger"/>
            <xs:element name="price" type="xs:decimal"/>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
    <xs:attribute name="orderid" type="xs:string" use="required"/>
  </xs:complexType>
</xs:element>
```

HTML vs. XML (1)

- XML does **not** replace HTML
- XML provides an alternative to allow you to define your own set of markup elements
- What is it for:
  - **HTML**
    - To create web-pages to be rendered on the client side (in the browser)
    - To present and display data (static)
  - **XML**
    - To transport data **between** applications and databases
    - To carry information (dynamic)

HTML vs. XML (2)

- Both are **not** programming languages
- HTML: designed to display data with focus on data presentation, web documents
  - A markup language **itself** (with predefined tags)
  - A presentation language
  - Case insensitive
- XML: designed to be software/hardware independent tool
  - To represent and store data
  - A framework for defining markup languages
    - Custom tags that can be defined by the user of the XML document
  - Case sensitive

XHTML

- XHTML: fusion of HTML 4.0 + XML 1.0
  - ➢HTML is an XML application
  - ➢→ we have document type definitions (DTD) corresponding to HTML 4
- Standard HTML/XHTML
  - ➢Static:
    - Once displayed on the browser, the document doesn't change until the user does something (click on a link).

Static webpage (1)

- Delivered to the user on the browser the same way it is stored
  - Same information for all the users
- HTML doesn't imply static webpages!

Static webpage (2)

- **Pros**:
  - Improved security over dynamic websites
  - Better performance for users
  - Limited dependencies on other systems (performance issues)
- **Cons**:
  - No dynamic aspect, should be implemented on the side

Dynamic webpage (1)

- **Client-side**:
  - HTML + Javascript (JS) that determines how the HTML document is parsed
  - Client scripts
- **Server-side**
  - The generation of the webpage is controlled by an application server
  - Server scripts

Dynamic webpage (2)

- **Plug-in**
  - ➢ It can depend on the system → constraints (security flaws, issues with resource management)
  - ➢ Plug-in capabilities included as standards

- Example: Adobe Flash (maintained until 2020)
  - ➢ It used to be a dominant platform for multimedia content
  - ➢ Replaced by HTML5, WebGL, Web Assembly
    - **WebGL** (Web Graphics Library): a JavaScript API for rendering interactive 3D and 2D graphics within any compatible web browser **without** the use of plug-ins.
    - **Web Assembly**: low-level bytecode format for in-browser client-side scripting, evolved from JavaScript.
  - ➢ No need of Flash anymore (RIP July 25, 2017)

Dynamic webpage (3)

- **Pros**:
  - Interactive website
  - Personalized experience
    - Based on what the user has seen before
- **Cons**:
  - Slower page load times
  - Content delivery network content caching ineffective
  - More expensive to produce and manage
  - Monitoring of effectiveness

Conclusion

- Web documents, in their concepts, are not that far from printed documents
  - →what is written, how it is presented, how it is stored,…
- When you are online, try to develop some awareness about the different entities that are present on the page
- **Key aspects**
  - Static vs. Dynamic (what can be changed and how)
  - Logical structure / Presentation
    - Not always easy to separate both !
    - Needs to think about what is a “document”
- Documents to represent documents ☺
- Documents to represent data
- Documents to represent how to represent documents …
  - https://www.w3.org/2001/XMLSchema.xsd
  - https://www.w3.org/2012/04/XMLSchema.xsd