Student Online Teaching Advice Notice

The materials and content presented within this session are intended solely for use in a context of teaching and learning at Trinity.

Any session recorded for subsequent review is made available solely for the purpose of enhancing student learning.

Students should not edit or modify the recording in any way, nor disseminate it for use outside of a context of teaching and learning at Trinity.

Please be mindful of your physical environment and conscious of what may be captured by the device camera and microphone during videoconferencing calls.

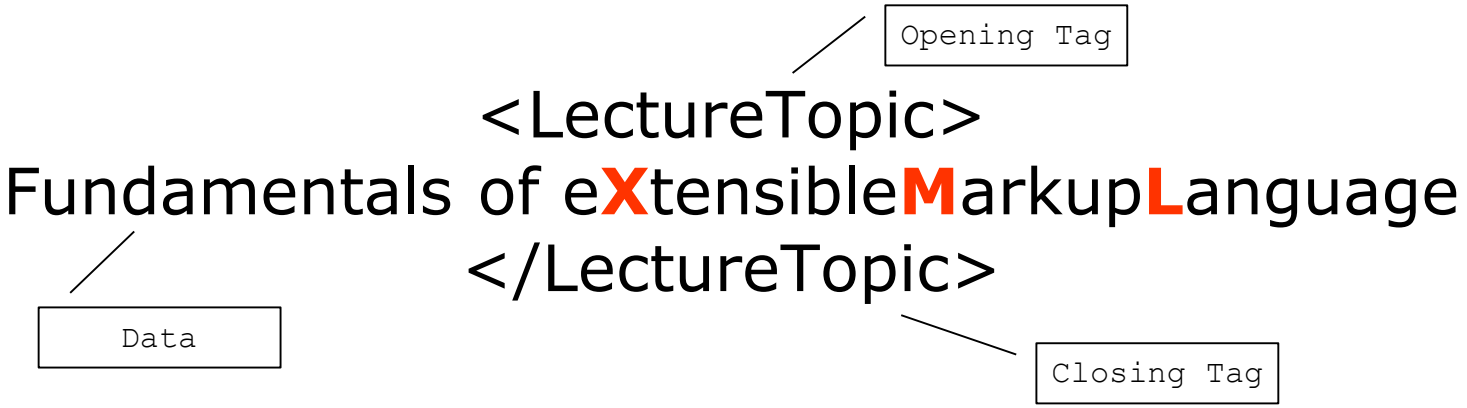
Recorded materials will be handled in compliance with Trinity's statutory duties under the Universities Act, 1997 and in accordance with the University's policies and procedures.

Further information on data protection and best practice when using videoconferencing software is available at https://www.tcd.ie/info_compliance/data-protection/.

CSU22041: Information Management I

Moving on to XML

Gaye Stephens gaye.stephens@tcd.ie



XML is a Markup Language.

In XML, Data is Marked up using opening and closing Tags

Moving on… to XML (The assignment)

- Each group will build on the work they have done in part 1 of the group assignment

- From Part one of your assignment

a) UML Classes will be transformed into XML documents and DTD documents

b) UML Use Cases will be the basis for Xquery/Xpath queries

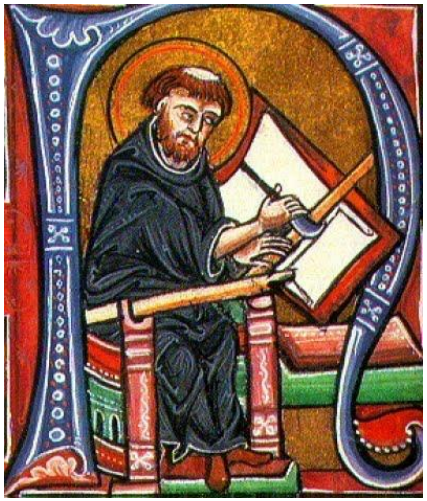
Overview

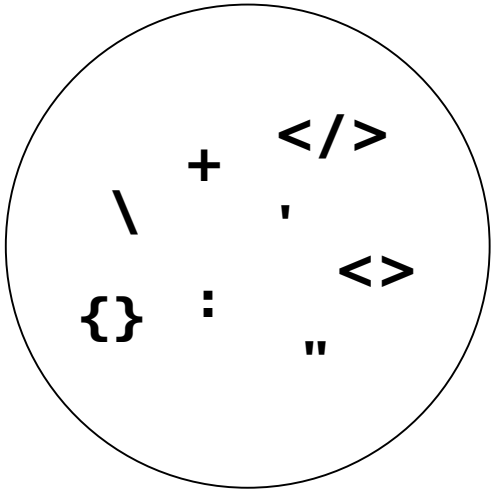
- What is a MarkUp Language
- Where XML came from
- What XML looks like
- XML Processing software

- Three separate short Videos to follow
  - Parts of an XML document
  - How to move from UML to XML
  - Using XML processing Software

What is Markup

- Sequence of characters within a text or word processing file to define
  - Print properties
  - Display properties
  - Document's logical structure
- Markup indicators are often called "tags"
  - Examples
    - RTF
    - EDIFACT
    - XML





Mark Up: RTF

```
\li0\ri0\sb240\sa60\keepn\widctlpar \aspalpha\aspnum\faauto\outlinelevel2 \a
djustright\rin0\lin0\itap0
\b\f1\fs26\lang2057\langfe1033\cgrid\langnp2057\langfenp1033
{\lang6153\langfe1033\langnp6153 Entity Relationship Diagram
\par }\pard\plain \s1\ql
\li0\ri0\sb240\sa60\keepn\widctlpar \aspalpha\aspnum\faauto\outlinelevel0 \a
djustright\rin0\lin0\itap0 \cbpat17
\b\f1\fs24\lang2057\langfe1033\kerning32\cgrid\langnp2057\langfenp1033
{\lang6153\langfe1033\langnp6153 Entity Type
\par }\pard\plain \ql
\li0\ri0\widctlpar\aspalpha\aspnum\faauto\adjustright\rin0\lin0\itap0
\fs24\lang2057\langfe1033\cgrid\langnp2057\langfenp1033
{\b\fs20\ul\lang6153\langfe1033\langnp6153
Def.:}{\b\fs20\lang6153\langfe1033\langnp6153 }{
\fs20\lang6153\langfe1033\langnp6153 An object or co ncept that is
identified by the enterprise as having an independent existence.
\par }\pard\plain \s1\ql
\li0\ri0\sb240\sa60\keepn\widctlpar \aspalpha\aspnum\faauto\outlinelevel0 \a
djustright\rin0\lin0\itap0 \cbpat17
\b\f1\fs24\lang2057\langfe1033\kerning32\cgrid\langnp2057\langfenp1033
{\lang6153\langfe1033\langnp6153 Entity
\par }\pard\plain \ql
```

Mark Up: EDIFACT

```
'''ED2'''OPENET:1111111:OVT':003705655815:OVT'ABC1234567'0'TYP:ORDERS'N
RQ:1'''

UNA:+.?
'UNB+UNOC:2+003705655815:30+1111111:30+980729:2233+4++ORDERS911+++KKK
KATE+1'UNH+

1+ORDERS:001:911:UN:FI0030'BGM+640+1234567'DTM+4:19981201:102'DTM+2:199
90101:102'DTM+2:9901:616'RFF+BC:123'RFF+VN:123456'NAD+BY+003705655815:1
00'

NAD+SE+11111111::92'NAD+PL+53432::92++KAUPPA:KAUPUNKI+KATU
9+KAUPUNKI++00007'NAD+CN+-::ZZ++TERMINAALI+OVI 42+TOINEN
KAUPUNKI++00069'UNS+D'LIN+1++23442423234

:EN'PIA+5+3244:MF'PIA+5+2341234324:ZBU'PIA+5+234243:ZCG'IMD+F+8+-
::91:KUKKAPUR

KKI:SAVI'QTY+21:8:KPL'FTX+AAA+++T.HARMAA:V[RI'FTX+AAA+++10:KOKO'PRI+NTP
:7.23:+

RP:7.32:PE'TAX+7+VAT+++::22.00'LIN+2++543434554345:EN'PIA+5+535:MF'PIA
+5+45:

PCE'UNT+38+2'UNZ+2+4'

'''EOF'''9'
```

Mark Up: XML

```
<fragment>
  <section>
    <title>Introduction</title>
    <para>Since the emergence of <acronym refid="xml">XML</acronym> in
    early 1998 and it's subsequent adoption across diverse application
    domains, one of the key benefits it enabled was the separation of
    content and presentation <bibref refloc="Bos97"/>. <acronym
    refid="xml">XML</acronym> borrowed this model (along with other
    important concepts) from the <acronym.grp><acronym
    refid="sgml">SGML</acronym><expansion id="sgml">Standard
    Generalised Markup Language</expansion></acronym.grp>. An
    <acronym  refid="sgml">SGML</acronym> document consists of
    logically structured content and uses a separate file (style
    sheet) to specify how the content should be formatted for
    [...]
    <figure id="img1">
      <title>ePublishing Components</title>
      <graphic href="02-04-03-fig01.jpg" width="321" height="214"/>
    </figure>
  </section>
</fragment>
```

- Background on Where XML came from.

- Lots of Technical abbreviations on next few slides e.g.
  - SGML, HTML, XML, Xpath, Xquery, ISO, XHTML, DTD, SVG, XSL, Xlink…….

What is SGML?

- Standard Generalised Mark-Up Language
- ISO standard since 1986
- Meta-language for defining document mark-up vocabularies
- Uses logical mark-up (structure, content) and not Physical Markup (how document looks on printed page)
- Platform-, system-, vendor- and version-independent documents
- Very powerful, but contains a number of complex features
- HTML is an SGML vocabulary

What is HTML?

- HTML, the *de facto* standard for publishing Web content, is an SGML vocabulary
- Supporting full SGML on the Web was too difficult so HTML made some simplifications
  - not extensible
  - limited structure
  - not content oriented
  - cannot be validated
- HTML is a simple language to understand and use
- The success of the World Wide Web is largely due to the simplicity of HTML.

```
<html>
  <head>
    <title>The SICK ROSE</title>
  </head>

  <body>
   <h1>The SICK ROSE</h1>

   <p>
     O Rose thou art sick.<br />
     The invisible worm,<br />
     [...]
   </p>

   <p>
     Has found out thy bed<br />
     Of crimson joy:<br />
     [...]
   </p>
  </body>
</html>
```

What is XML?

- eXtensible Markup Language, using user-defined tags
- XML is a simplified subset of SGML
- Can also be used to define document markup vocabularies (e.g. XHTML)
  - These can have a strictly defined structure (DTD)
- Retains the powerful features of SGML (extensibility, structure, validation)
- XML documents look similar to HTML documents
- Separates structure and presentation (like SGML)

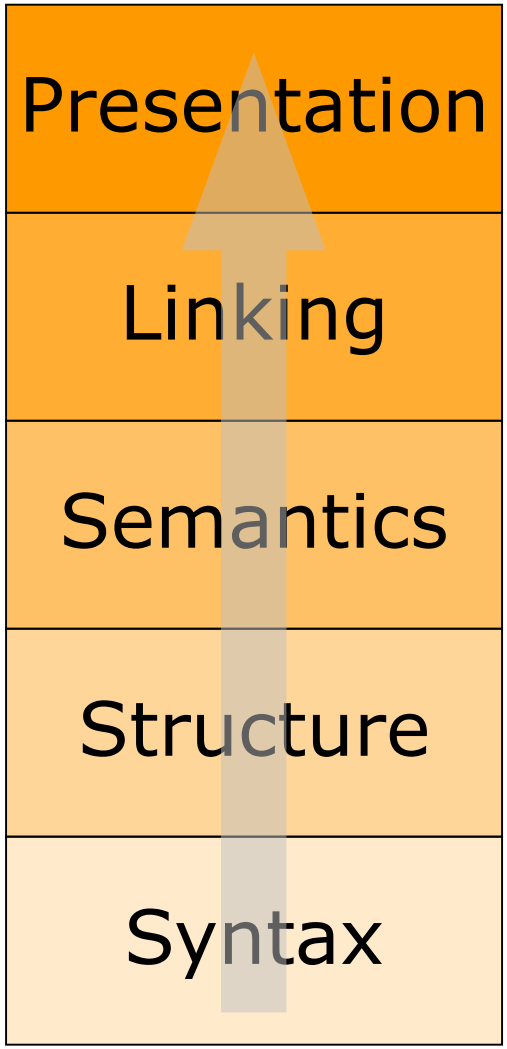
Why is the emergence of XML an important development?

- XML is a tool for defining vocabularies
  - XML vocabularies are easy to read
  - XML is self describing
    - Parse tree embedded in document
- XML vocabularies are easy for computers to process, exchange and display
  - XML tools are ubiquitous, free and conform to established standards
  - Natural affinity with Object serialization
  - Data source neutral

Design goals of XML 1.0 specification

1. XML shall be straightforwardly usable over the Internet.
2. XML shall support a wide variety of applications.
3. XML shall be compatible with SGML.
4. It shall be easy to write programs which process XML documents.
5. The number of optional features in XML is to be kept to the absolute minimum, ideally zero.
6. XML documents should be human-legible and reasonably clear.
7. The XML design should be prepared quickly.
8. The design of XML shall be formal and concise.
9. XML documents shall be easy to create.
10. Terseness in XML markup is of minimal importance.

XML technologies



CSS, Cascading Style Sheets
XSL, Extensible Stylesheet Language
XPath, XQuery

XLink, XBase
XPointer

Topics Maps, Ontology Web Language
Resource Description Framework (RDF)

XML Schema, RelaxNG, RDF Schema,
Document Type Definition (DTD)

XML Namespaces
XML 1.0

XML Example

```
<?xml version='1.0' encoding='ISO-8859-1' standalone='yes' ?>
<doc type="book" isbn="1-56592-796-9" xml:lang="en">
  <title>A Guide to XML</title>
  <author>Norman Walsh</author>
  <chapter>
    <title>What Do XML Documents Look Like?</title>
    <paragraph>If you are ...</paragraph>
      <item>
        <paragraph>The document begins ...</paragraph>
      </item>
      <item>
        <paragraph>Empty elements have ...</paragraph>
        <paragraph>In a very ..</paragraph>
      </item>
    <section>...</section>
    ...
  </chapter>
  <chapter>...</chapter>
</doc>
```

References

- XML
  - Home Page: http://www.w3.org/XML/
  - Tutorial: http://www.w3schools.com/xml/default.asp

- XML Processing
  - Tutorial: http://www.w3schools.com/XML/dom_intro.asp
  - https://docs.oracle.com/javase/tutorial/jaxp/sax/parsing.html
  - Home Pages: http://sax.sourceforge.net/ http://www.w3.org/DOM/

BaseX Software

- A light-weight, high-performance and scalable **XML Database** engine and **XPath/XQuery Processor.**

- Interactive and user-friendly **GUI frontend**

- Different programming APIs to connect to BaseX XML database
  - REST-Style Web API
  - Variety of Client APIs for different programming languages See http://docs.basex.org/wiki/Developing

- YOUR ACTION: Download Core Package Java BaseX to your laptop or your U: drive or to D: drive on PC (http://basex.org/products/download/all-downloads)