

The background of the slide features a light blue gradient with a faint, semi-transparent image of classical architectural columns on the left side. The columns are white with detailed capitals and are set against a darker blue background. The entire slide is framed by a thin black border and a thicker brown border.

XML for Legal Information

Terry Butler

Director Research Computing

Arts Resource Centre

University of Alberta

CALL/ACBD 2006

Edmonton Alberta

2006 May 8



Outline

- Introduction to XML
- Structured Documents
- Metadata and Searching
- Indexing
- Legal Text Mining
- Humanities Computing Projects
- Question and answer

The background of the slide features a light blue gradient with a faint, semi-transparent image of classical architectural columns on the left side. The columns are white with detailed capitals and fluted shafts, set against a darker blue background.

Introduction to XML

what is XML?

XML (Extensible Markup Language) is a simple, flexible text format

it allows people to define their own markup schemes, and to share information based on those schemes

many areas have developed standards for markup and information exchange

XML is easy to use with web technologies

The background of the slide features a light blue gradient with a faint, semi-transparent image of classical architectural columns on the left side. The columns are white with detailed capitals and fluted shafts, set against a darker blue background.

Introduction to XML

the history of XML

first came SGML, a complex markup language for publishing

SGML was difficult to use, and had impact only in large organizations which could control document workflow

meanwhile, proprietary file formats came and went ...

and then the www changed everything

Introduction to XML

HTML and the world wide web (1)

HTML was based on the idea of SGML but went farther, faster because

- it was simple
- there was software to support it
- it took advantage of emerging networks

Introduction to XML

HTML and the world wide web (2)

HTML **prescribes** certain tags, and browser software displays them

the web strikes a balance between information creators and consumers

- the tags describe **content** but actually only affect presentation
- creator has less control than with word processing or desktop publishing
- different operating systems, browsers, and fonts means total visual control is impossible

Introduction to XML

HTML and the world wide web (3)

`<p>Question: Why does this work at all?`

`
Answer: Users care more about
information than about rules!`

[minimal](#)

[basic](#)

[better](#)

The background of the slide features a light blue gradient with a faint, semi-transparent image of classical architectural columns on the left side. The columns are white with detailed capitals and fluted shafts.

Structured Documents

discovering structure in documents

information may have inherent structure; the structure is integral to its meaning
documents express this structure visually
document classes share certain properties
so **document analysis** can lead to an understanding of meaning and relationships among the information elements

[Canadian Constitution 1982](#)



Structured Documents

distinguish presentation from meaning

a citation vs. the article itself

[Romanow - Nation Building citation](#)

[Pelletier - Morganatic marriage - citation](#)

[Pelletier - Morganatic marriage - article](#)



Structured Documents

how to create structured documents

- *post-facto* tagging
- authoring in a controlled context
- dynamic documents

[Masterplots - Carol](#)

[Masterplots - Carol - browser view](#)

[DTD for Masterplots](#)

The background of the slide features a light blue gradient with a faint, semi-transparent image of classical architectural columns on the left side. The columns are white with detailed capitals and fluted shafts, set against a darker blue background.

Metadata and Searching

“data about data”

library catalogue records and citations are good examples

can be powerfully combined with full-text searching

familiar examples include:

- [CANLII](#)
- [Lexum](#)



Legal Text Mining

software to create structured text (1)

transcript and document analysis software

Text Map 2

features: word indexes, formatting, linking to
case management software



Legal Text Mining

software to create structured text (2)

qualitative analysis software

nVivo

features: manage, describe, code, report textual data; complex but powerful interface

ATLAS.ti

features: a powerful workbench for the qualitative analysis of large bodies of textual, graphical, audio, and video data

fully compatible with XML



Humanities Computing Projects

An Electronic Index to the Notebooks of
Samuel Taylor Coleridge

overview of the project

[demonstration](#) of the live linking



Humanities Computing Projects

Aboriginal Voice in EIA (1)

“To collect, digitize and make available public statements made during Environmental Impact Assessment hearings on large projects in the Canadian North. A specific focus is the aboriginal voice in this process.”

These are public documents which have very little present public utility: difficult to access
After collection and encoding, a web-based searching system will make both text and themes accessible



Humanities Computing Projects

Aboriginal Voice in EIA (2)

Document structure

Tagging scheme

Aboriginal Voice – [sample of tagging](#)

The background of the slide features a light blue gradient with a faint, semi-transparent image of classical architectural columns on the left side. The columns are white and have ornate capitals. The entire slide is framed by a thin black border and a thicker brown border.

Question and Answer

I am glad to answer any questions you may have.

Contacts

Terry Butler, Arts Resource Centre, University
of Alberta

www.arts.ualberta.ca/~tbutler/

Terry.Butler@UAlberta.ca

Chris Fletcher, Department of Anthropology,
University of Alberta

www.ualberta.ca/~cf6/