Comp 336/436 - Markup Languages

Fall Semester 2017 - Week 1

Dr Nick Hayward

Course details

Lecturer

- Name: Dr Nick Hayward
- Office: Doyle 307 (LSC)
- Office hours
  - *Wednesday afternoon by appointment (WTC)*
- Faculty Page

Course schedule

Important dates for this semester

- DEV week: 9th to 16th October 2017
  - ***n.b.*** *no formal class: 11th October 2017*
  - *presentation & demo: 18th October 2017 @ 4.15pm*
- Thanksgiving break: 22nd to 25th November 2017
  - ***n.b.*** *no formal class: 22nd November 2017*
- Final class: 6th December 2017
  - *presentation & demo: 6th December 2017 @ 4.15pm*
- Exam week: 11th December to 16th December 2017
  - *Final assessment due on 13th December 2017 by 4.15pm*

Initial course plan - part 1

(up to ~ DEV Week)

- intro and consideration of MLs
- their structure, usage, and implementation
- general manipulation...
- metadata
- semantic structure and usage
- ...

Initial course plan - part 2

(up to the end of the semester)

- MLs processing and manipulation
- project usage and examples
- parsing and rendering of MLs
- stronger semantic structure and organisation
- analysis, visualisation &c. of MLs
- ...

Assignments and coursework

Course will include

- weekly bibliography and reading (where applicable)
- weekly notes, examples, extras...

Coursework will include

- quizzes or group exercises at the end of each section (Total = 30%)
  - *based on course notes, reading, and examples*
- development and project assessment (Total = 70%)
  - *mid-semester assessment (Total = 30%)*
    - end of DEV week
    - demo due Wednesday 18th October 2017 @ 4.15pm
  - *final assessment (Total = 40%)*
    - demo due Wednesday 6th December 2017 @ 4.15pm
    - report due Wednesday 13th December 2017 @ 4.15pm

Quizzes, group exercises...

Course total = 30%

- at least one week notice before quiz
  - *average time ~40 minutes (can be extended...)*
  - *taken towards the end of class*
- group exercises
  - *help develop course project*
  - *test course knowledge at each stage*
  - *get feedback on project work*

Development and project assessment

Course total = 70% (Parts 1 and 2 combined total)

Initial overview

- combination project work
  - *part 1 = mid-semester **DEV Week** work (30%)*
  - *part 2 = final demo and report (40%)*
- group project (max 4 persons per group)
  - *groups posted to Trello*
- design and develop a markup language based app
  - *markup language/s has to form part of the core structure*
  - *clear demonstration of processing and usage of MLs*
  - *combination of structural and presentational usage of MLs*
  - *demonstration of semantic structure and usage*
  - *use and integration of metadata*

DEV Week assessment

- project outline and introduction
- developed using a chosen markup language
- consider and apply metadata schemes and semantic organisation for chosen domain
- current working examples - what does and does not work...
- demo and project report
  - *due on Wednesday 18th October 2017 @ 4.15pm*
- anonymous peer review
  - *similar to user comments and feedback*
  - *chance to respond to feedback before final project*

Final assessment

- working final app
- modify and update DEV week project work
  - *include parsing and rendering of the data suitable for broader publication*
  - *develop and integrate stronger semantic organisation and manipulation*
  - *provide an opportunity for analysis, visualisation...*
- presentation and demo - live working app
  - *due on Wednesday 6th December 2017 @ 4.15pm*
  - *show and explain implemented differences from DEV week project*
  - *where and why did you update the app?*
  - *benefits of updates?*
- how did you respond to peer review?
- final report
  - *due on Wednesday 13th December 2017 @ 4.15pm*

Goals of the course

A guide to developing and publishing interactive client-side web applications and publications.

Course will provide

- examples of processing and manipulation of MLs
- examples of structure and rendering of MLs
- clear understanding of underlying technologies used to
  - *develop and publish markup languages based projects*
- provide a deeper understanding of example markup languages
- relation of MLs to metadata and semantic organisation and usage
- opportunity to consider and develop two project-based applications
- ...

Course resources - part 1

Website

Course website is available at
https://csteach436.github.io

- timetable
- course overview
- course blog
- weekly assignments & coursework
- bibliography
- links & resources
- notes & material

NO Sakai

Course resources - part 2

GitHub

Course repositories available at https://github.com/csteach436

- weekly notes
- examples
- source code (where applicable)

Trello group

Group for weekly assignments, DEV week posts, &c.

- Trello group - COMP 436
  - *https://trello.com/comp436*

Slack group

Group for class communication, weekly discussions, questions, &c.

- Slack group - COMP 436
  - *https://csteach436.slack.com/*

Group projects

- add project details to course's Trello group, *COMP 436 - 2017 @ LUC*
  - *Week 1 - Project Details*
- create channels on Slack for group communication
- start working on an idea for your project
- plan weekly development up to and including DEV Week
  - *9th to 16th October 2017*
  - *DEV week demo on 18th October 2017*

Intro to markup languages - part I

- a simple consideration of a markup language
  - *highlight, draw attention to, content & parts of a given document*
- markup might provide
  - *instructions, comments, notes*
  - *or simply background information for content or data*
- encounter markup languages & rendered output on a daily basis
  - *access a web page, news feed, or other online source...*
- use of HTML - HyperText Markup Language
  - *simplicity, ease of development & usage*
  - *helped to drive ongoing development & popularity of web*
- HTML has faced shortcomings and limitations

Intro to markup languages - part 2

- markup may also be construed in slightly different terms
  - *as we consider a given target audience*
- e.g. markup of a HTML document for a web browser
- XML, SVG, and many other examples
  - *may have different intended recipients and interpretations*
- HTML5 markup needs to be syntactically precise intended web browser
- XML - eXtensible Markup Language
  - *popular option for storing data*
  - *long time option for the transmission and sharing of data*
  - *now challenged by other options such as JSON*
- HTML was designed to display data - ie: how it looks
- XML was designed to carry data - ie: how it is stored
- presentational vs structural

Markup languages - literary studies

- many uses of markup languages include structured data and metadata
  - *digitised artifacts and objects*
- consider *literary studies*
- assumption of automatic means for analysis and testing
- application of strict, detailed textual preparation if no automated process
- markup as an external intervention
- markup allows a critically interpretative process
- open criteria for encoding textual material

Markup languages - literary studies

- Electronic Scholarly Editions
- TuStep
  - *example*
  - *project homepage*
- personalised markup
- modification of encodings
  - *to match software analysis requirements*

Markup languages - TEI

- The Wonderful World of TEI...
- read a potted introduction and history at the TEI website
  - *http://www.tei-c.org/About/history.xml*
- Malory Project - TEI XML based project
  - *Caxton example - http://www.maloryproject.com/xml/caxton/Caxton.xml*

Digitisation - intro

- crossroads of research and development
- consider options for digitisation
  - *e.g. current analogue record*
  - *historical, literary*
  - *structural*
  - *...*
- inherent benefits of digitisation
  - *accessibility*
  - *flexibility*
  - *ease of manipulation*
  - *aggregation*
  - *storage....*

Digitisation - examples (good and bad)

- many fine examples now exist of digital archives, projects, editions...
  - *Papers Past*
  - *https://paperspast.natlib.govt.nz/*
  - *Library of Congress*
  - *https://www.loc.gov/*
  - *British Library*
  - *https://www.bl.uk/*
- Google Books project
  - *interesting initial 'History' of the project*
- EEBO, ECCO, NCCO...
- others such as Internet Archive
  - *https://archive.org/*
- EEBO-TCP
  - *intro*
  - *http://www.textcreationpartnership.org/tcp-eebo/*

Digitisation - why do we bother?

benefits and costs

- loss of analogue information to precise digitisation

> *We should be cautious about letting the radiance of the bright future blind us to the limitations of this new technology" (Smith, Abby. 1999. "Why Digitize?")*

Digitisation - why do we bother?

> *"analogue information can range from the subtle tone and gradations of the chiaroscuro in a Berenice (bernice) Abbott photograph of Manhattan in the early morning light, to the changes in the volume, tone, and pitch recorded on a tape that might, when played back on equipment, turn out to be the basement tapes of Bob Dylan." (Smith, Abby. 1999. "Why Digitize?")*

Image - Berenice Abbott, New York

Berenice Abbott, Pike and Henry Street, New York, 1936

- Source - Lumiere
- http://lumieregallery.net/wp/167/berenice-abbott/

Digitisation - potential for loss

- dependent upon various factors
  - *e.g. how much information you actually gather*
- whilst digitising we may consider
  - *density of data*
  - *frequency of sampling*
- breadth or depth of information gathered is also a consideration

Digitisation - analogue vs digital

- ongoing saga of 'Analogue vs Digital'
- consideration of cost, time, transmission and storage
- loss of information from analogue to digital
- benefits of digital
- perceived benefits of analogue

> *"...reducing it to 0s and 1s is not unlike a root canal: by extracting the nerves, the tooth is killed in order to save it." (Unknown. 2004. "The Great Analog Versus Digital Debate". VoicePrint Online)*

References

- British Library
  - *https://www.bl.uk/*
- EEBO-TCP
  - *http://www.textcreationpartnership.org/tcp-eebo/*
- Internet Archive
  - *https://archive.org/*
- Library of Congress
  - *https://www.loc.gov/*
- Lumiere - Berenice Abbott
  - *http://lumieregallery.net/wp/167/berenice-abbott/*
- Malory Project - Caxton XML
  - *http://www.maloryproject.com/xml/caxton/Caxton.xml*
- MDN web docs - HTML
  - *https://developer.mozilla.org/en-US/docs/Web/HTML*
- Papers Past
  - *https://paperspast.natlib.govt.nz/*
- Smith, Abby. "Why Digitize?". 1999.
- TEI Project - intro
  - *http://www.tei-c.org/About/history.xml*
- TuStep Project
  - *http://www.tustep.uni-tuebingen.de/tustep_eng.html*
- Unknown. "The Great Analog Versus Digital Debate". VoicePrint Online. 2004.

- Woolf Online
  - *http://www.woolfonline.com*