DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2013

Week 6

Today's Class

- Digitisation (continued)

- Text Encoding

Discussion Analysis

Perseus Digital Library

The act of digitisation

Digitising Text - Options for making your own digital text

- import, automate, or type your own
- scan or photograph analogue to digital
- minimum 300dpi scan or better 600dpi
- book scanners and cradles are also available
- OCR texts
- OCR limitations include poor performance for
  - non-Latin characters
  - small print
  - certain fonts
  - complex page layouts or tables
  - mathematical or chemical symbols
  - most texts pre-19th century

The act of digitisation

Digital Images

- again we can use scanners and cameras
- more specialised material with cradle and book scanners
- consider fidelity to the original and long term preservation
- quality of digital result dependent upon
        - quality of original
        - digitising method employed
        - sample rate of digital to analogue
- minimum 300 DPI, preferably 600 DPI
- uncompressed format, such as TIFF
- optimise to JPEG for online publication

The act of digitisation

Digital Images - metadata

- metadata added to images
- in particular important when the files are not text searchable
- standards such as Dublin Core or METS
- technical information about the file such as
  - resolution
  - compression (if applicable)
  - scanning process
  - copyright and access rights
  - authorship of the image
  - and much more...

good knyght so sir Gryfflet was the firste that sette oute and to hym com
a knyght hys name was sir Ladynas and they com so egirly to gydir that
all men had wondir and they so sore foughte that hir sheldis felle on
pecis and both horse and man felle to the erthe And both the french knyght
and the englysh knyght lay so longe that all men wente they had bene dede
Whan lucas the butler saw sir Gryfflet ly so longe he horsed hym a
gayne a none and they too ded many mervelous dedis of armys wt many
bachelers Also sir kay com oute of a busshemente wt S knyghtis wt hym and
they vj smote othir vj downe But sir kay ded that day many mervay
lous dedis of armys that there was none that ded so welle as he that day
Than there com Ladynas and Grastian ij knyghtis of fraunse and dyd paf
synge welle that all men prayfed them Than com in sir Placidas a good
knyght that mette with sir kay and smote hym downe horse and man wher
fore sir Gryfflet was wroth and mette with sir Placidas so harde that horse
and man felle to the erthe But whan thes knyghtis wyst that sir kay
had a falle they were wroth oute of mesure and there wt all ech of them v
bare downe a knyght Whan kynge Arthur and the ij kynge saw them be
gynne wexe wroth on bothe ptyes they leped on smale hakeneyes and sette
cry that all men sholde departe unto there lodgynge And so they wente home
and unharmed them and so to evynfonge and soupe And aftir soupe the iij
kyngs wente in to a gardyne and gaffe the pryse unto sir kay And unto sir
Lucas the butler and unto sir Gryfflet And than they wente unto coun
ceyle and with them Gwenbaus brothir unto kynge Ban & Bors a wyse
clerke and thidir wente Ulphinus & Brastias & Merlion And aftir they had
in her counceyle they wente unto bedde And on the morne they harde masse
and to dyner and so to theire counceyle and made many argumentis what
were befte to do So att the laste they were concluded that it sholde go w a
tokyn of kynge Ban thatt was a rynge unto hys men And kynge Bors
Gracian & Placidas sholde go agayne and kepe p castels and theire contreyes
And as for kynge Ban of Benwyke and kynge Bors of Gaule had or
dayned them all thynge and so passed the see and com to Benwyke And
whan the people sawe kynge Banys rynge and Gracian & Placidas they
were glad and asked how theire kynge fared and made grete Joy of p welfare

DIGH 400 - Introduction to Digital Humanities Research

The act of digitisation

Audio and Film

- different set of issues and solutions
          - often involves more specialised technical development
- formats, hardware and playback changes over time
- storage of video is another concern
- preparation for online use
- detailed metadata required for searching and cataloguing

The act of digitisation

Audio

- three elements in the creation of an audio file
        - audio stream, container, and codec
- audio can be uncompressed or compressed as with images
- quality loss from CD to MP3
- common archival standards for audio
        - 24 bit audio at 48kHz

The act of digitisation

Video

- consists of the raw video signal (video stream) and the file format
- video file format also contains encoded audio stream
- audio stream uses its own codec
- developing and changing standards
- interlaced or progressive scan

Good summary available at

Make It Digital

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

- 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