



## CENTER FOR TEXTUAL STUDIES AND DIGITAL HUMANITIES

# DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2014

Week 6

## Today's Class

- Discussion Analysis (Perseus Digital Library)
- SciFi Authors Update
- Digitisation (Part 2)
- Text Encoding
- TEI Introduction

## Discussion Analysis

[Perseus Digital Library](#)

SciFi Authors Update

[SciFi Authors Sheet](#)

## 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

'information formally distinct from the character sequence of the digital transcription of a text, which serves to identify logical or physical features or to control later processing.'

- 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

## Text Encoding Initiative

### Consortium

- maintains a technical standard
- set of guidelines
- user wiki
- set of tools for development and processing
- currently predominant in social sciences and humanities, in particular
  - textual studies
  - literary studies
  - linguistic studies

## Text Encoding Initiative

- membership
- Special Interest Groups (SIG)
  - currently 10 active groups
  - wiki and mailing list
- Activities
  - various boards and membership groups
  - conferences
  - workgroups
  - project information

## Text Encoding Initiative

A gentle introduction to the [TEI Guidelines](#)

- guidelines for electronic text encoding and interchange
- define and document a markup language for representing
  - structural
  - rendition
  - logical & semantic
  - analytic
- rules and recommendations rather than standard
- wide variety of possible solutions for encoding material
- freedom of expression of personal textual theory
- current version = P5
- over 513 elements, 214 attributes...
- customisations such as [TEI Lite](#)



## Text Encoding Initiative

A gentle introduction to the Guidelines

- structural features
  - organisation of information in text
- TEI defines overall text structure using
  - front
  - body
  - group
  - back
- a consideration of rendition features

## Text Encoding Initiative

A gentle introduction to the Guidelines

- rendition features such as
  - distinct fonts
  - colours
  - alignments
  - italics, underline, bold
  - font weight...
- highlighting

[Malory Example](#)

`<hi>italic words...</hi>`

`<hi rend="italic">italic words...</hi>`

## Text Encoding Initiative

A gentle introduction to the Guidelines

- logical and semantic features such as
  - emphasis
  - foreign words
  - linguistically distinct words, phrases
- we can also consider
  - quotation marks
  - quotes
  - cited quotation...

## Text Encoding Initiative

A gentle introduction to the Guidelines

- analytical features such as
  - notes and comments
  - marking for indexing
  - regularisation
  - editorial statements