

2010-11-16

Digital Publishing and Preservation Using XML

Welzenbach, Rebecca; Schaffner, Paul; Hawkins, Kevin

<http://hdl.handle.net/2027.42/109398>



TEI workshop session 3

A *VERY* short introduction to document analysis.

STEP 1. Know everything. Understand everything ...

STEP 2. Give up on that. Acknowledge your limitations.

There is an Infinite and Wonderful Variety in books.

- [1000 cats \(/files/departments/dpp/20101113-pfs/20101113-03b063.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b063.jpg) [1000 cats \(/files/departments/dpp/20101113-pfs/20101113-03b063.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b063.marked.jpg)
- [Graphics plus \(/files/departments/dpp/20101113-pfs/20101113-03b_cartoon1.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b_cartoon1.jpg) [Graphics plus \(/files/departments/dpp/20101113-pfs/20101113-03b_cartoon1.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b_cartoon1.marked.jpg)

Equally infinite choices in what to markup. Markup is always interpretation.

- You can markup what is there.
- You can markup what you choose to find there
- You can markup what is only implicitly there (tagging silences)

One can try to distinguish between categories of features (but they are interrelated)

- Linguistic features
- Presentational features
- Structural/rhetorical features
- Artefacts of print
- Etc.

Example: marking *passages you like* (no visual cue):

- [Kilvert \(/files/departments/dpp/20101113-pfs/20101113-03b088.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b088.jpg) [Kilvert \(/files/departments/dpp/20101113-pfs/20101113-03b088.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b088.marked.jpg)

Mostly you will be responding to obvious physical cues. Asking "what is this thing?" "What is it here for?" "How does it relate to the other things here?" Visual cues aren't everything, they can be misleading, but they are a start. Especially if you are writing instructions for *someone else* to recognize features.

Examples:

- [Spider \(/files/departments/dpp/20101113-pfs/20101113-03b089.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b089.jpg) [Spider \(/files/departments/dpp/20101113-pfs/20101113-03b089.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b089.marked.jpg)
- [Pork \(/files/departments/dpp/20101113-pfs/20101113-03b096.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b096.jpg) [Pork \(/files/departments/dpp/20101113-pfs/20101113-03b096.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b096.marked.jpg)
- [Fixtures \(/files/departments/dpp/20101113-pfs/20101113-03b017.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b017.jpg) [Fixtures \(/files/departments/dpp/20101113-pfs/20101113-03b017.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b017.marked.jpg)
- [Houses \(/files/departments/dpp/20101113-pfs/20101113-03b018.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b018.jpg) [Houses \(/files/departments/dpp/20101113-pfs/20101113-03b018.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b018.marked.jpg)
- [Dante \(/files/departments/dpp/20101113-pfs/20101113-03b003.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b003.jpg) [Dante \(/files/departments/dpp/20101113-pfs/20101113-03b003.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b003.marked.jpg)

Leverage your knowledge. It helps if you know something about...

- genre,
- print conventions,
- language
- subject matter
- author
- period
- the work itself

But expect to be ignorant at least sometimes. Use what you know and allow for incomplete tagging

These can still be marked up, despite being (partly) unfamiliar:

- [Polish \(/files/departments/dpp/20101113-pfs/20101113-03b077.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b077.jpg) [Polish \(/files/departments/dpp/20101113-pfs/20101113-03b077.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b077.marked.jpg)
- [BCP \(/files/departments/dpp/20101113-pfs/20101113-03b002.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b002.jpg) [BCP \(/files/departments/dpp/20101113-pfs/20101113-03b002.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b002.marked.jpg)
- [Cernyweg \(/files/departments/dpp/20101113-pfs/20101113-03b016.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b016.jpg) [Cernyweg \(/files/departments/dpp/20101113-pfs/20101113-03b016.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b016.marked.jpg)
- [Concordance 1 \(/files/departments/dpp/20101113-pfs/20101113-03b025.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b025.jpg) [Concordance 1 \(/files/departments/dpp/20101113-pfs/20101113-03b025.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b025.marked.jpg)
- [Concordance 2 \(/files/departments/dpp/20101113-pfs/20101113-03b040.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b040.jpg) [Concordance 2 \(/files/departments/dpp/20101113-pfs/20101113-03b040.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b040.marked.jpg)
- [Cornish \(/files/departments/dpp/20101113-pfs/20101113-03b043.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b043.jpg) [Cornish \(/files/departments/dpp/20101113-pfs/20101113-03b043.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b043.marked.jpg)

On the other hand, sometimes one is left genuinely at a loss. In my case, here:

- [no way! \(/files/departments/dpp/20101113-pfs/20101113-03b075.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b075.jpg) [no way! \(/files/departments/dpp/20101113-pfs/20101113-03b075.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b075.marked.jpg)

An easy sample

- [Selden \(/files/departments/dpp/20101113-pfs/20101113-03b001.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b001.jpg) [Selden \(/files/departments/dpp/20101113-pfs/20101113-03b001.marked.jpg\)](/files/departments/dpp/20101113-pfs/20101113-03b001.marked.jpg)

SO: pick a couple of samples and look at them:

- **What are the salient features?**
- **How would you instruct someone to recognize them?**
- **How do they relate to each other?**
- **What would you gain by marking up one feature set over another?**

- Are there advantages to *adding* information? normalizing or making explicit?
- Is there anything anomalous or inexplicable? or simply difficult?
- Are there any concurrent (overlapping, conflicting) organizational hierarchies?

Page maintained by [Paul Schaffner \(/users/pfs\)](#)

Last modified: 11/22/2010

Contact

University of Michigan Library

818 Hatcher Graduate Library South
913 S. University Avenue
Ann Arbor, MI 48109-1190
(734) 764-0400
contact-mlibrary@umich.edu



[Give Feedback](#)

Copyright



Except where otherwise noted, this work is subject to a [Creative Commons Attribution 4.0 license](#). For details and exceptions, see the [Library Copyright Statement](#).

©2014, [Regents of the University of Michigan](#)

[Go To Mobile Site](#)