

Manuscripts Online Project Meeting

Friday 11 May 2012, 11:00
Humanities Research Institute, Sheffield

Present: Mike Pidd, Kathy Rogers, Orietta Da Rold, Wendy Scase, Linne Mooney, Estelle Stubbs, Sharon Howard, Peter Findlay (JISC)
Apologies: Jeremy Smith

Review of NLP samples

KR: the initial 2-page samples have produced reasonable results, better than expected, although notes that they use punctuation/capitalisation so are not the most difficult

Some blips - eg

- 'Advent' marked up as a Location;
- overenthusiastic treating 4-digit numbers as years - this can be tweaked fairly easily
- confused by Roman numerals

NB: While part dates are marked up, there has to be a full numerical year to index a date

Issues

- resolving movable feasts/regnal dates to calendar dates
- disambiguation, eg 'Trinity' is both a date and location

Would be useful to identify MS references/shelfmarks for searching.

KR - where these are tagged it's straightforward to do this as long as we have info on what the tags are.

We need project partners to identify fields/tags in their datasets that they want specifically marked up - this will also need to be done for datasets from other sources

At present everything between tags will be indexed for keyword search - if anything should be excluded we need to know this too.

KR will put marked-up collections online to work with

Responsibilities

Each member of the Editorial group to take responsibility for their own datasets plus:

MED - WS

Blake Canterbury Tales - OR

Auchinleck - JS

Europa - speak to Toby Burrows about this

Cause Papers and Taxatio - MP/KR

British History Online - SH

BLMO - to arrange later (no project members seem to have access to this - check JS and JT)

EEBO-TCP - JT

Parker - ES

TNA - OR

BL catalogue - LM

ME Corpus - JS

METS - MP

Resource Descriptions

Taking the Connected Histories descriptions as basic template +

- 'Keywords'
- Introductory reading (where it's available)

Technical methods - not needed on every page, focus on where documentation on the resource website is lacking

Strengths and limitations -> 'Scope'

Order

Summary
URL
Keywords
Description
Scope
Intro reading
Tech
About
Contact

Summary information

Document types:

- Literary Manuscripts
- Non-literary Manuscripts
- Official Documents (government, civic, legal, religious)
- Literary Printed Books
- Non-Literary Printed Books

Resource format:

- Full-text transcription
- Summary or Calendar
- Manuscript Description

Language(s)

Access

Time period

Publisher

Search Engine / User Interface

Work progressing, slightly behind schedule. We've added Matthew Groves to the project team to work on the interface

By the next meeting basic search should be available for discussion

Prototype and internal testing during May up to the next meeting, followed by wider testing among project partners

Roll out for full user testing in July (NB that this is during the long vac)

Will need to schedule a focus group and compile a survey for online testers

Should flag up in advance to testers - write a post to the mailing list to report on progress at the same time

User Focus Group #2

OR suggestion via JISC usability workshop - rather than putting together a group of people, identify specific users for 1 to 1 feedback (could be via Skype or in person)

Solves problems with one person dominating a group

Only issue is that it could be time-consuming

Internal Testing

Project partners to engage in scenario-based testing. Partners to propose research scenarios to the team.

Review of Stakeholder Panel feedback

Areas for clarification:

Search paths to be a private feature

Public word cloud of popular search terms

Will collect activity data but unlikely to be able to do much with it during funded project

Web 2.0 specification

Crowdsourcing of dialect variants

OR: is it really worth doing this? WS pointed to the difficulty of the task, small numbers of people who could do it meaningfully; and that it's highly disputed among experts.

It was decided not to try to do it after all.

However, mapping still possible using provenance information already marked up in datasets - again, will need to know what to look for in both partners' datasets and external ones.

Material Transfer Agreements

Completed MTAs:

Manuscripts of the West Midlands (Birmingham)

Production and Use (Leicester)

Medieval Scribes (York)

Middle English Dictionary and Corpus (Michigan)

British Literary MSS (Gale Cengage)

Middle English Texts (W. Michigan)

Draft MTA awaiting final comments: TCP-EEBO (Michigan)

Data bundle 1 in early stages/awaiting responses on initial draft

Imagining History and Geographies of Orthodoxy (Belfast) (have data)

MEG (Glasgow) (have data)

Auchinleck (have data)

Europa Inventa (don't have data)

Cause Papers (have data)

Anglo-Norman Dictionary (don't have data)

Dissemination Activities

Upcoming

WS: New Chaucer Society Congress (Portland), July, and Computing and Palaeography workshop (Germany), September. - arrange publicity materials

OR: Paper at Bergen, June.

Quadrivium

Cathy Shrank has made a list of people with medieval interests at Sheffield - OR to email them next week

Project conference

Date has been booked

How to go about organising papers/sessions? Andrew Prescott lined up for keynote

CFP - what should it do/cover (bearing in mind it can be useful publicity in itself).

Possibly a mix of formats

- talks (invited speakers)
- roundtable discussion
- workshop
- short talks
- poster session

OR (? I think) to put something together for the June meeting agenda

Publicity for attendees - try to target future potential resource providers

JISC Reporting

Interim report is due 31 May

MP to draft and circulate.

AOB

SH to organise Doodle poll for June and August meetings