

DE Specs Working Group Meeting

Klaus Thoden

08 September 2008

1 Legacy Specs

Two data entry specifications were distributed, one for texts in the Latin alphabet and one for Chinese. For both writing systems, ECHO pages served as examples.

After a few notes on a Chinese text regarding the data entry specifications, a text in Latin alphabet was discussed concerning its idiosyncrasies, e. g. ligatures. A diplomatic transcription of the characters is planned, however, most of the ligatures are to be resolved except for the most common ones (e. g. *æ* and *œ*).

2 Things to deal with

2.1 Character encoding

Characters are to be typed preferably in Unicode, not in XML tags, i. e. they should be directly typable, not as entities.

Unknown characters are to be numbered by the digitizers. Same instances of one unknown character get the same number so that these can be resolved easily.

2.2 Conversion to XML

The basis of the conversion to XML will be XML schemata, rather than DTDs. RELAX NG is the proposed language for writing these schemata.

3 Organisation

The *trac*-pages were introduced¹. They contain timeline, roadmap, version control, source viewer and a wiki. There will soon be an introduction to the use of *trac*.

4 Next steps

- Wolfgang and Klaus are going to look at a few documents in ECHO². They should
 - scan the documents and note interesting things
 - make a list of the points what occurs as a structure that should be marked up
 - collect pages that are good to write specs about
- A list of recommended books and articles is also on the first page of the wiki.
- The library is asked to buy a few relevant books ...

¹<https://itgroup.mpiwg-berlin.mpg.de:8080/tracs/mpdl-project-content>

²Links to these documents are on the first page of the wiki. Firefox is the recommended browser here.