

“Heterogeneous image indexing in the BVH Project (Virtual Humanistic Libraries in Tours)”

Marie-Luce Demonet (*Centre d'Études Supérieures de la Renaissance, Université François-Rabelais, Tours-CNRS, UMR 6576*),

The Bibliothèques Virtuelles Humanistes (BVH, or Virtual Humanist Libraries, a digitization project begun in Tours in 2003, <http://www.bvh.univ-tours.fr>) offers two types of digital representations of a selection of 2000 books printed (or manuscripts written) during the Renaissance: the image of a copy (its "facsimile") on the one hand, and its transcription on the other hand, without additions besides corrections or variations that are essential for understanding the text. Two levels of images must be encoded: the image of the page itself, and the images included within the page (illustrations, ornaments, *picturae* of emblems, portraits, ornamented letters, graphs, and so on), properly indexed according to their nature, to feed separate and connected databases. 10 to 20 % of the books (mainly in French) will be totally transcribed, manually and semi-automatically, with specialized OCRs generating TEI encoded files. The TEI recommendations (P5 release) allow the encoding of illustrated matter, with pointers to URLs and/or keywords embedded in the tags: the search engine, adapted to all Internet browsers, will process not only the metadata and the databases of illustrations, but also the full text encoded with TEI tags, to which we can add “types”, i.e. the generic classifier of the illustration and the keyword taken out of the Iconclass thesaurus. To speed up image indexing, similarity software (designed by the Navidomass Project, a consortium with six computing teams) automatically applies previous indexed images to a new corpus, and the scholar validates the proposal. By March 2010, more than 19000 illustrations have been indexed according to Iconclass, and ornamental letters have their adapted sorting system. This paper will discuss the current results of this project.