A first analysis of the digital corpus of the Vatican Apostolic Library

Giuliano Giuffrida

Vatican Apostolic Library, Vatican City, Vatican State.
`g.giuffrida@vatlib.it`

Abstract. The Vatican Apostolic Library is digitizing its collection of manuscripts. About 6 Million of FITS images have been acquired so far, distributed among 21000 shelfmarks, for a total of 400 TB of data.
In collaboration with the University of Rome Tor Vergata, we started an analysis of the high resolution (400 dpi) images.
We propose a new approach to discriminate pixels containing ink from "empty" pixels - i.e. pixels that represent only the parchment without ink - based on the analysis of millions of Pixel Intensity Histograms. This first work focuses on 4622 shelfmarks and 1.7 Million of images, the corpus object of this analysis being distributed among 16 collections (8 Latins, 5 Greeks and 3 Hebrews) produced between the III and XIX century.
A second analysis concerns the extraction of the filling coefficient using two parallel independent approaches. The first one uses aggregated data to extract a representative value for the whole shelfmark, whereas the second one extract a value for the filling coefficient for each page using morphological operators. This second analysis is focused on the shelfmarks from Vat.lat.1 to Vat.lat.2000, covering a large range of centuries (VI to XIX sec.). Both analysis are performed using original Java and Python codes together with Apache Common Math, OpenCV, and other libraries.
These works are part of an ambitious project that aims to extract codicological parameters from the whole Vatican Apostolic Library digital corpus. We plan to analyze the dataset using classification and clustering algorithms and share it with scholars and researchers for further analysis.

Here I will show some preliminary results of our work, highlighting the different approaches that we are following and comparing our findings with previous works.

Keywords: Digital Humanities · Codicology · Manuscript Studies · Image Processing