

F.Galli -E. Nieddu, (Universita della Svizzera italiana & Roma 3)

***In Codice Ratio*: using VREs in the study of the Medieval Vatican Registers**

In Codice Ratio is a joint research project with the Departments of Engineering and Humanities at Roma Tre University, and the Vatican Apostolic Archive (formerly Vatican Secret Archive). Our goal is to build tools to enable the study of medieval manuscripts at scale, including an automated transcription system based on cutting-edge OCR technology.

As a case study, we focus on the volumes (9-13) of the Vatican Registers, containing copies of official papal letters and documents written under Pope Honorius III (1216-27). As well-known, this collection is a highly valuable historical source, comprehending important scripts such as a copy of the Rule of St. Francis, as approved in the Papal Bull "Solet Annuere" (29 Nov. 1223, nr. 4582). As of today, Vatican Registers are mostly available as digital images (they have been published only partially) and thus it requires manual search by expert humans to retrieve a specific content or a phrase. In order to speed up search at large scale, we leverage new technology, such as machine learning and crowdsourcing, and modern automated knowledge extraction techniques, such as, full text search, named entity recognition and text mining.

In this context, we will present two virtual research environments:

1. a crowdsourcing platform to collect training data for an automatic transcription system, while disseminating and teaching basic paleographic, historical and linguistic knowledge to high school students;

2. a search application prototype for full-text search and advanced queries on pages automatically transcribed.

Based on our first-hand experience, we will discuss lessons learnt, potentials and limits of the aforementioned systems and tools, both as an aid to scholars who deal with manuscripts (paleographers, philologists, historians) and as dissemination and 'training' means for non-specialist audiences (young students, general public).

In Codice Ratio was presented as a showcase project at the last Google AI Tensorflow Summit.

*

Proponents: Dr. Francesca Galli (Università della Svizzera italiana, postdoc assistant and adjunct lecturer in Medieval Philology) and Elena Nieddu (Università Roma Tre, PhD student in Computer Science).

Contacts: francesca.galli@usi.ch ; elena.nieddu@uniroma3.it

Project website: <http://www.inf.uniroma3.it/db/icr/>