Lancaster University

JOB DESCRIPTION
Research Associate in Artificial Intelligence and Digital Humanities
Vacancy Ref: A3759

Job Title: Senior Research Associate in Artificial Intelligence and Digital Humanities	Present Grade: 7

Department/College: History

Directly responsible to: Prof Patricia Murrieta-Flores

Supervisory responsibility for: None

The project:

The main objective of Unlocking the Colonial Archive project is to engage artificial intelligence approaches to advance scholarship in the humanities, particularly in the use of historical textual and image collections held at cultural institutions.

The project brings a strong interdisciplinary team to work with some of the most important collections for the study of colonial Latin America at LLILAS Benson Latin American Studies and Collections at the University of Texas at Austin and the General Archive of the Nation in Mexico. Through the use of machine learning methods in three distinct research areas, the project will:

1. Expedite the transcription, query, and retrieval of information from historical documents using and experimenting with state-of-the-art Handwritten Text Recognition (HTR).
2. Advance research in the combined use of Natural Language Processing (NLP), Linked Open Data (LOD), and Corpus Linguistics for the automated identification, cross-reference, and mining of historical information from large multilingual historical text collections.
3. Develop Computer Vision techniques in combination with Linked Open Data to facilitate the automated identification of iconographic elements, as well as the search and analysis of pictorial features in Indigenous maps and printed books.

In doing so, the project will solve the questions:

-How can open computational approaches benefit a broad range of humanities scholars and promote the digitization of cultural and historical archives?

-How can cultural institutions expedite the transcription of, and in doing so, the access to their invaluable historical text collections?

-How can automated methods for the identification of information help researchers in the discovery of data from collections that would take them a lifetime to explore?

-How can cultural institutions create frameworks for the mining of information from their materials that can benefit both archival work and research with their collections?

-How can AI techniques help specialists answer important historical questions related to Mesoamerican and colonial pictorial documents?

-How can the creation of machine learning models with HTR, NLP, LOD, and Computer Vision using 16th- and 17th-century Indigenous and Spanish historical material and datasets advance research in AI?

Focusing on four historical collections crucial to the study of colonial Latin America, the work proposed will greatly facilitate the searching, retrieval, and cross-linkage of large volumes of information, and it will give scholars the capacity to identify patterns and pose new questions that are difficult to answer due to the sheer scale of the colonial archive. For instance, the transcription of thousands of documents through HRT, will enable the first ever longue durée exploration of the Fondo Real de Cholula, the only known extant archive of a Spanish Crown-designated "Indigenous City" in New Spain. The use of LOD and NLP will allow the identification and cross-referencing of information between archival collections facilitating the study of population movement in New Spain, shedding new light into changing settlement patterns throughout the sixteenth century. Furthermore, the unique pictorial and mapping tradition that emerged from the encounter between the Indigenous Mesoamerican and Spanish cultures is extraordinarily rich. Although there is a long tradition of study of Mesoamerican codices, in

the case of colonial maps, understanding of the processes through which this combination of spatial knowledge and conceptions took place, is still far from complete. A framework that facilitates the search, comparison, and cross-reference of iconographic and other elements will help identify changes and continuities in these maps. Their study will allow us to use new technologies critically and with a decolonial perspective, contributing to the study of Indigenous and subaltern cultures.

Major Duties:

- Plan and carry out research on automatisation techniques with computer vision approaches to identify, extract, and analise information from historical documents, maps and images.

- Implement machine learning approaches as necessary to solve the research questions of the project.

- Participate with the team on designing the Linked Open Data framework of the project.

- Work with the team on general automatization tasks for investigating historical documents and images.

- Identify and understand work requirements, prioritising tasks and responsibilities within a timeframe agreed with the Principal Investigator.

- Develop and contribute to peer reviewed publications based on the research of the project.

- Collaborate proactively with the research team.

- Contribute to project meetings at Lancaster and internationally as appropriate.

- Communicate their research effectively across a range of audiences, academic and non-academic.