Project proposals for IT master/bachelor students

1. Swedish Caribbean Colonialism Project

   Development and modelling of our research database, presently developed in Filemaker. The database will probably eventually be ported to an open source SQL version when we have finalised the main part of our metadata inventory of the historical material being digitised in the project, i.e. the archival holdings pertaining to Swedish Saint Barthélemy (1784–1878).

   Design and programming of web-based search interface connected to the research database. Here we need to set up trial versions to be able to do live searches and test GUI for ourselves and eventually end users.

   Image retrieval from the university library hosted Alvin repository and presentation of image data that is connected to the searches in the research database.

   Contact: Ale Pålsson, ale.palsson@hist.uu.se
   Contact: Victor Wilson, victor.wilson@hist.uu.se
   Contact: anna.foka@abm.uu.se

2. Classification and semantic analysis of visual and textual data in a historical dataset of medical publications

   Apply and compare NLP (Natural Language Processing) approaches and possibly AI/ML methods to a dataset of historical medical journals to classify them automatically and analyse the content semantically. Use this to build a searchable database with a user frontend suited to humanities researchers. Categorization and semantic analysis could include: word meanings, geographical informations, entities, people, chronologies, bibliographic data (authors, editions, etc). The data also contains images/visual data so various types computer vision/image analysis approaches could also be examined within the same aims. Solving OCR and document analysis problems would also be relevant for the dataset.

   The results should be accessible to historical researchers through a cloud environment provided collaboratively by The Centre for Digital Humanities/Uppmax/University IT.

   Several master projects could cover subsets of this task if it is too big for one student to solve. The student/students will be collaborating with The Centre for Digital Humanities and The Centre for Medical Humanities and Social Sciences.

   https://www.idehist.uu.se/centre-for-medical-humanities/
   Contact: matts.lindstrom@abm.uu.se
3. Periegesis project – researching and implementing automated mechanisms for text extraction concerning place names.

Project website
http://www.periegesis.org

The project aims at creating an enriched digital edition of an ancient traveling guide, incorporating and combining contemporary archaeological data.

Two datasets and suggested implementation options for text mining and automatic subject classification:

Archaeological surveys are catalogues of information registers of place and sites and monuments. 1990s- 2021s (Volume 30-60 tops)

   a) Text mining Archaeological surveys (English, Swedish) and thus creating a GIS database of all place mentions in archaeological reports, with additional metadata (year of discovery, persons involved, artefact/ monument classification).

Skills: regular expressions; basic scripting (e.g. Perl, Python); Geographic Information System (GIS) or Google Earth. Natural Language Processing, i.e applying named entity recognition methods to recognize place names.

   b) Historical Maps and extracting textual data. Use Machine Learning/ Image Pattern Recognition methods to extract place names and thus creating a database of all place mentions in a number of historical maps in TIFF format with additional metadata (year of print, cartographer(s), names, place/ artefact/ monument classification)


Supervision: Anna Foka/ Anders Hast
Contact: anna.foka@abm.uu.se

4. Quantifying Culture Project

Project description

Supervision: Anna Foka, Fredrik Wahlberg

The aim of this project is to establish mathematical models and algorithms for
AI/ML/Image Pattern Recognition generated descriptions of art and heritage. The objective is to use those to automatically classify a variety of datasets from the 19th and 20th Centuries. Data providers and partners are: the Swedish National Heritage Board, the National Museums of World Culture, the National Archives of Sweden and Uppsala University Museum (Gustavianum).

In supervision, the MA student(s) will explore how state-of-the-art ML algorithms infer and encode nuanced and meaningful descriptions on a mathematical level. This research can have two parallel tracks connected to Digital Humanities:

a) theoretically study how issues of a qualitative nature (bias, diversity, ethics) as connected to the quantitative nature of the internals of AI/ML methods used for heritage institution digitalisation.

b) Implementation of heritage relevant ML algorithms in code in order to test the applicability of our common ideas. We will examine how overlapping taxonomies and statistical uncertainty can be encoded in relation to categorisations and descriptions.

All code produced will be open source. For this part the student can mainly use research on ML bias and ‘explainable AI’ from the fields of computer vision and natural language processing as the basis.

Contact: anna.foka@abm.uu.se

5. Face recognition on historical photographs (Anders Hast).


Face recognition on historical photos is challenging since rather few examples of photos of each person are available. So called one shot learning has proven to be efficient for such cases. Nevertheless, a CNN has to be trained in order to be able to create relevant features. Siamese networks have been used in the past and have proven to make good predictions. The idea is to (1) make a literature overview of face recognition methods and techniques; (2) investigate what already trained CNN network can be used for transfer learning, for the purpose of creating relevant features for historical photos; (3) try two or more deep learning pipelines (siamese etc) for one-shot (or a few examples) recognition and compare their results.

Deep learning for face recognition requires computational resources, especially GPU’s so it is presumed to be necessary to use SNIC super computer resources for this type of work. It is required that the student has taken some course in Machine Learning and is interested in programming, especially in the field of deep learning. Any Image Analysis courses are an advantage, but not a requirement.
Contact: anders.hast@it.uu.se