

# The History of Europe App - A pipeline for Humanist-Machine Interaction in the Digital Humanities

## **CUBRIK and the History of Europe App**

The integration of human expertise and machine computation enables a new class of applications with significant potential for the digital humanities. So far this potential remains largely untapped due to the severe requirements of such projects: The implementation and integration of advanced algorithms requires specialized know-how and the final users from the humanities are challenged with defining unprecedented tasks for methods which haven't emerged yet. The FP-7-funded research project CUBRIK ([www.cubricproject.eu](http://www.cubricproject.eu)) implements and integrates research in computer science, the design of human-computation tasks, data visualization, social engineering and the humanities.

In the proposed presentation we would like to showcase one of CUBRIK's case studies, the demo of the History of Europe application. The application introduces an effective interface to access collections of historical sources and to discover links among and entities within them. Upon completion CUBRIK will offer an innovative approach to human-enhanced time-aware multimedia search by synthesizing research in computer science, crowdsourcing and gamification. We will conclude the presentation with an outlook on the future development of the application.

## **Humanist-machine interaction**

The History of Europe (HoE) application is based on a curated collection of more than 3000 images, representing the main events and actors in the history of the European integration. The collection is curated and hosted by the Centre Virtuel de la Connaissance sur l'Europe (CVCE). In a first step, an image indexation pipeline identifies the location of individual faces in the photographs. The location of these faces is verified by a crowd of "click-workers" with no specific training who evaluate for each recognized face if the depicted image shows a human face or not. Following the face verification process, an automatic face recognition process is triggered that associates each of the now verified faces with a list of ten possible identities. This list of candidates is then disseminated for example through Twitter to a crowd of experts that vote and comment for their preferred identity.

Besides the identities of the different persons, all information that is associated to an image, such as the time or the place where the image was taken as well as contextual information about associated historical events can be reviewed by expert users and delegated to a crowd of domain experts for review.

## **Data aggregation, visualisation and analysis**

Building on the computed co-occurrence of persons in images a social graph is constructed that connects them with each other. Connections gain in strength the more often persons appear together in an image. Finally the result of this process is depicted in a visualization of the social graph with a set of analytical tools.

The social graph in the History of Europe App aims at representing and visualizing dependencies between historically relevant persons in the context of European integration. Thereby the weight of the (social) links between person entities relies on their co-occurrence in historic photographs as identified by the aforementioned image indexation process. The more frequently two persons appear in different photographs, the stronger the link between the corresponding entities in the graph.

Users can interact with the History of Europe social graph in different ways, e.g. a click on a node results on an ego-graph of the selected person and clicking on an edge displays documents that relate to both selected relationship. As the documents stored in the collection very often come with a date of creation, the graph can be filtered by date with the timeline, displaying only the connections of documents created within this timespan. This timeline also shows the amount of photos per date that are contained in the collection. Another filtering option is the number of connecting documents, which allows the visualization of those relationships that are only included in an interval of a minimum and maximum number of documents. This feature is useful to highlight highest co-occurrences. Finally, the number of appearances of a person in the processed collection lets us identify people who appear particularly often in any given time frame.

### **Crowd discussion and a new approach to the representation of truth in digital research tools**

Another challenge for the HoE app and the domain of the Digital Humanities in general is the conception of truth, which differs significantly e.g. to the conceptions of truth in Computer Science. Computer Scientists can rely on a stable foundation of what is true: Any experiment can be replicated and measured precisely. In the humanities the concept of truth is far more complex: It is based on the insight, that there is no neutral or objective way to study human environments. The way, in which questions are asked, how data is selected to answer them, by what means this data is analyzed and finally the way in which the results of such analyses are communicated and received all challenge the idea of "one truth".

In order to represent the discursive nature of truth in the humanities within HoE we make use of a community-driven tool for question answering, similar to stackoverflow.com. User have the opportunity to answer questions and thus benefit from the knowledge within the expert crowd. However, the system allows for more than one answer and offers its users the possibility to vote and answer up or down, thereby allowing more than one answer to enter in competition with each other whilst also maintaining the full spectre of the discussion.

### **Summary and outlook**

The History of Europe application takes on the challenge to combine cutting edge research in the domains of computer science, the design of human-computation tasks, data visualization, social engineering and the humanities by identifying

synergies between the disciplines' strengths and by compensating for their weaknesses. We do this by building a pipeline which connects face recognition tools, data visualization and input from humans and creates an ongoing cycle of iteratively improved user input and machine output. The History of Europe application stands in line with a range of other online tools for historical research but introduces new social features as well as crowd sourcing from both click-workers and expert users which continuously improves the system. In the future we will expand the selection of sources to include digitized text documents as well as audio and video interviews from different archives.