

HIMANIS

Historical MANuscript Indexing for user-controlled Search

Partnership

- **Institut de Recherche et d'Histoire des Textes (Project Leader) - FRANCE**
- Universitat Politècnica de València - Sis. Informaticos y Computacion – SPAIN
- A2iA - FRANCE
- Rijksuniversiteit Groningen - Artificial Intelligence Institute - NETHERLANDS

Project Summary

Manuscripts are among the most important witnesses to our European shared cultural heritage. In recent years, large quantities of historical handwritten documents are being scanned and made available through web portals. Yet, the wealth of information conveyed by the text captured in these images remains largely inaccessible. General users and researchers more and more expect to query handwritten resources in plain text like printed books, but current handwritten text recognition technology is still far from offering sufficiently accurate transcripts on historical documents. This situation raises the need of searching approaches specifically designed for text images.

HIMANIS aims at developing cost-effective solutions for querying large sets of handwritten document images. To this end, innovative keyword spotting, indexing and search methods will be developed, tested, adapted and/or scaled up to meet the realworld conditions required.

HIMANIS will associate Computer Science (A2iA, UPVLC and RUG), Humanities (IRHT) and Cultural Heritage (The European Library, the Archives Nationales of France and Bibliotheque Nationale de France) institutions and produce technology to generate new, research-based knowledge from historical manuscripts. As a challenging and particularly interesting case study, the large collection of registers produced by the French royal chancery (14th-15th c.) will be used as a test bed. This will allow sustainable use and management of this important cultural resource. Implementing user feedback will also increase our understanding of the meanings that it holds for people and how they perceive, use and interpret it, thereby meeting societal challenges in promoting its use and re-use.

The expected outcomes of this project are: a) a new indexing/searching technology for historical manuscripts, brought to the market by A2iA; b) a new paradigm to study our historical heritage, as conveyed by manuscripts, by using full text search technology. c) a new vision of the raise of nation states in Europe via a new study of the corpus under this paradigm.