**FORTH ICS**

DIATHESIS

An Information System for documentation, management and promotion of historical digital documents

DIATHESIS is an information system for documentation, management and promotion of historical documents that supports both digital library functionality and archival management of the original documents. It includes OCR-based page analysis and subject clipping, subject-level metadata generation, semantic indexing and multifaceted classification of subjects using built-in thesauri. The data produced by the OCR processing of the scanned material are used for the creation of a highly flexible annotation interface which allows users to perform hybrid annotations upon the digitized material assigning semantic properties to specific regions of text that represent a subject. The goal of the documentation process is the creation of a coherent semantic backbone that can be easily enriched with semantic relations. It is not meant to be a complete semantic structure that includes all the semantic relationships and entities (Actors, Places) described in the text.

The query interface enables users to conduct searches on a document as well as on a subject level basis combining both full text and metadata search capabilities. Queries on the document level are based on conventional metadata assigned automatically to the whole document during the import phase while queries on the subject level exploit the semantic relationships that have emerged from the documentation phase. The combination of the different query modes provides a semantic filter that greatly improves the precision of the conducted searches. The subject's metadata are based on a robust top level domain ontology (CIDOC-CRM, ISO 21127) in order to ensure that the produced knowledge can be inter-exchanged between different institutions.

The query result presentation mechanism allows the partial download of the digitized material in order to improve the overall user experience and reduce the download time.
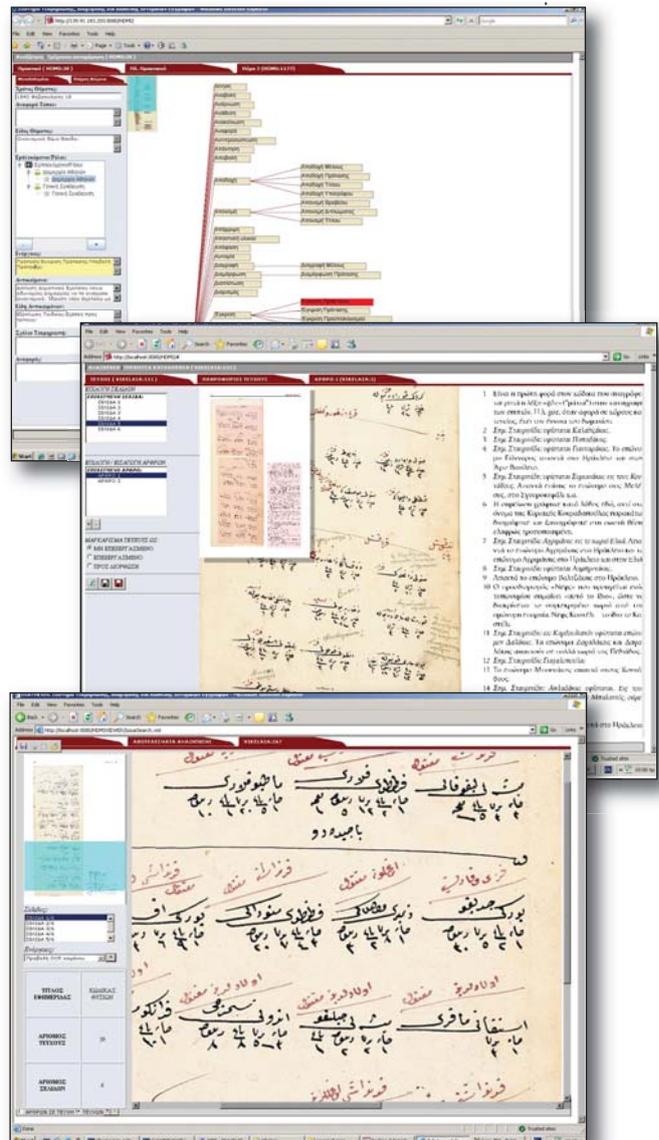
DIATHESIS: *Web interface for the documentation, search and presentation of subjects/articles and documents*

CENTER FOR CULTURAL INFORMATICS

FORTH ICS

DIATHESIS consists of three lightweight, easily deployable and highly configurable Web applications, namely the administration, the documentation, and the querying applications, which allow data import and monitoring, classification and indexing, and search and presentation respectively.

DIATHESIS is currently being used for the archival, documentation and promotion of three historical archives of the Vikelaia Municipal Library of Heraklion, namely the archive of Newspapers and Magazines, the Turkish Archive of Heraklion, and the Municipal Archive ("Archio Dimogerontias"). In a first phase 500.000 pages have been digitized and 20% have already been classified and indexed in the system.

DIATHESIS is also being successfully used for the archival of handwritten manuscripts of "Filekpedeftiki Etaireia", an educational non-profit organization founded in 1836 in Greece. The archival material concerns 10 volumes of minutes of the Board of Directors and General Assembly meetings of the "Filekpedeftiki Etaireia" since 1840.

Currently a new application for the archival, documentation and promotion of the archives of a Greek newspaper is under way.
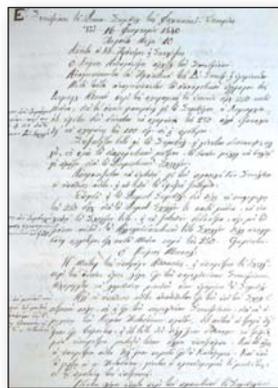
Contact Person:
Maria Theodoridou
Research and Development Engineer
maria@ics.forth.gr

INFORMATION SYSTEMS LABORATORY (ISL)

The Information Systems Laboratory combines expertise in knowledge representation and reasoning, database systems, net-centric information systems, and conceptual modelling. Its principal research challenge is to succeed in the transition from traditional information systems, such as information retrieval systems, database and workflow management systems, to semantically rich, large-scale, adaptive information systems. Such systems are characterised by large-scale semantic interoperation, massive distribution, and high level autonomy and self-adaptation.

Besides conducting theoretical work on the above issues, ISL carries out applied research in a number of application domains, including cultural informatics, biomedical informatics, e-learning, e-commerce and IT security. Work in these domains has a strong interdisciplinary character, since it includes the aspect of understanding and modelling the respective domain through an appropriate conceptualisation. Overall, ISL aims to provide enabling technologies for the emergence of the information society and the knowledge economy.

CENTER FOR CULTURAL INFORMATICS (CCI)

The Centre for Cultural Informatics (CCI) brings together skills in knowledge representation, ontology engineering, knowledge organisation systems, database technology and web technology with expertise in archaeology, museum documentation and management, site and monument management, art conservation, archives and libraries, thesaurus and dictionary management and other cultural disciplines.

Contact Persons:
Grigoris Antoniou
Professor of Computer Science
Head of ISL
antoniou@ics.forth.gr

Dr. Martin Doerr
Principal Researcher
Head of CCI
martin@ics.forth.gr