



The Digital Library Engine

OVERVIEW

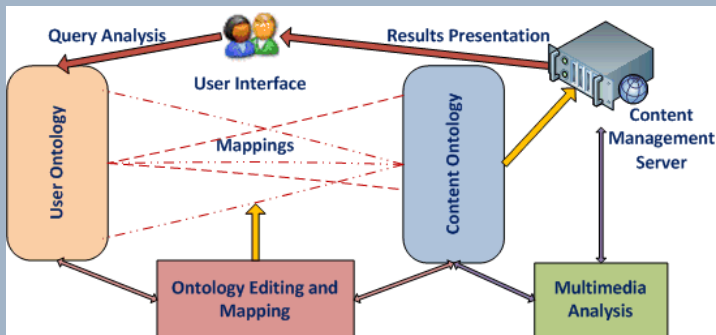
The Papyrus Digital Library Engine is an integrated solution for increasing the value of digital content archives through the 'semantic translation' of content in a way that the archive is rendered useful to more audiences than the originally targeted. In this way we create a dynamic digital library which understands user queries in the context of a specific discipline, looks for content in a domain alien to that discipline and returns the results presented in a way useful and comprehensive to the user.

Papyrus has illustrated this approach with the recovery of history from news digital content. The rationale behind this selection is that vast amounts of digital news content exist in huge archives, which, although being of incredible value, are underused as they are not easily searchable and do not have a significant value if seen as individual news items. The content contained in news archives could very well lead to a dynamic composition of a complete historical library extending to the latest minute in time. However, even though all historical events have been recorded as news, the format, style, and philosophy of a 'news item' differ significantly from their counterparts of a 'historical reference' to the same event. Interpretation of news in the 'language' of history requires suitable modelling of both domains, semantic analysis of the news content and the historians search queries, proper modelling of the semantic correspondences between them and presentation of the results in the context of the 'history' domain.

INNOVATION

Except for the main innovation of providing cross-discipline access to digital content archives, Papyrus also delivers significant advances in the State-of-the-art in the following areas:

- **semantic multimedia analysis:** by introducing knowledge assisted methods which take advantage of existing metadata and content structure models for the understanding of the source content



- context sensitive **query processing** methods: for the understanding of the user demands.
- **knowledge mapping**: for corresponding concepts between the source content and the user queries
- **presentation techniques**: for delivering the results in a manner comprehensive to the targeted users.

BUSINESS IMPACT

The Papyrus Digital Library Engine proposes a new paradigm for repurposing digital content archives to make them useful for new audiences. Although the deployed example showcases only the news-history discipline pair, there are many more cases that such a platform could add significant value to a digital archive providing new avenues for exploiting existing content. Although the initial consideration assumes a local installation of the engine in the premises of the digital archive, the design of the platform allows for a fully distributed deployment both for the software modules and the underlying content. This means that the system can be considered both as an application deployed locally at the archive, as well as a service provided to various content providers through a central content aggregator.

INTEROPERABILITY

A major objective of Papyrus Platform was to decouple functionality and technology. The design of the system is based on Service Oriented Architecture, RESTful loose coupled web services and a modular architecture where all the composite units are separated in layers (tiers). By following these principles, the Papyrus Engine facilitates the rapid, stable and efficient integration of many different technologies, offers high scalability and flexibility and improves the overall operational efficiency.