

# Issues and Challenges in Accessing Metadata Contents of Cultural Heritage Collections

Wafaa Maaitah, Abdullah Zawawi

## Extended abstract

Digital resources use metadata as data schemas or elements to describe their contents. Different metadata standards exist for describing various types of digital objects. Several researches have been reported on how to address issues related to access of digital metadata resources. Most studies involve cultural heritage domain, and this is an indication of the importance of this domain in metadata research and development. Research on metadata in cultural heritage mainly revolves around three fundamental issues: (i) Lack of quality in digital metadata contents, (ii) Difficulty in accessing digital metadata contents due largely to limited user's knowledge on the content of the digital metadata, and (iii) Heterogeneity of the data at the level of schemas which makes the access even more difficult. The lack of quality in the content (structure) of metadata makes it difficult for the users to retrieve and explore information that satisfies their needs. So, in order to make its contents more accessible, rich metadata content and metadata interoperability are required, especially for cultural heritage collections which consist of digital objects (structured documents) described by a variety of metadata schemas.

This paper presents the issues and challenges of accessing metadata contents by providing an analysis of the existing digital metadata approaches in digital metadata text environment with a particular emphasis on cultural heritage collection. This paper firstly looks at metadata approaches which are classified into two categories, namely data retrieval and information retrieval. The classification is based on how digital metadata objects are handled in enriching and retrieving their contents. In data retrieval, many efforts have been made in the development and enrichment of metadata schemas which handle metadata as a database. Data retrieval deals with well-defined structured data and expressive query languages. A data retrieval system aims to determine which objects of a collection satisfy clearly defined conditions as those in a relational algebra expression. In a data retrieval system, a single erroneous object among a thousand retrieved means a total failure. In data retrieval, integration of metadata is also crucial in which mapping of schemas is often employed. In information retrieval (IR), the IR community has traditionally been focusing on searching unstructured content where retrieval result returned by the IR systems is the whole document. In contrast, a number of researchers have shown that metadata retrieval could break away from the traditional IR by allowing users to retrieve the most relevant components of document instead of the whole document as a single large block of relevance information. IR systems also interpret the content of the information items in order to decide their relevance. Digital Metadata resources have also been widely studied by the IR research community. Their research on digital metadata retrieval normally involves two expansion methods which are document expansion method and query expansion method, and many retrieval models which include vector space model, Boolean model, probabilistic model and language model.

In the final part, this paper presents the findings from the analysis and suggestions on how to overcome the issues and challenges in accessing digital metadata contents especially in cultural heritage collection. A framework that aims to improve the effectiveness of retrieval when searching metadata contents is then proposed. The proposed framework consists of approaches and methods that would enrich and enhance the quality of metadata contents. The framework is expected to improve access to digital metadata contents especially in digital heritage collection and particularly for those with limited knowledge on cultural heritage.

Keywords: Metadata, Information Retrieval, Data Retrieval, Digital Cultural Heritage.