Organising and Maintaining Dynamic Digital Collections

M. Nikolaidou[1], D. Anastasiou[2], D. Anagnostopoulos[2], M. Hatzopoulos[1]

[1] Department of Informatics,
University of Athens
Panepistimiopolis, 15771 Athens, Greece
{mara}, {mike}@di.uoa.gr

[2] Department of Geography
Harokopion University of Athens
70 El. Venizelou Str, 17671, Athens, Greece
{danasta}, {dimosthe}@hua.gr

1. Introduction

To provide access to research material to researchers and physicians of Athens Medical School (AMS), a Digital Library system (AMS DL) was developed. It maintains medical images produced by the laboratories of the Medical School. End-users access the system through the Web. Images are catalogued and processed partially by laboratory scientific staff and partially by cataloguers in the Central Library of Heath Sciences. AMS DL system is based on a multi-tiered client-server model and is implemented using Java and *IBM Content Manager* platform [1]. Since the system should be open, it supports existing standards regarding the metadata scheme and communication interfaces. Thus, the metadata scheme adopted is an extension of Dublin Core, while results are obtained as XML pages [2]. Laboratory requirements significantly differ, thus it was decided to develop a different collection for each laboratory. Since the number of collections needed is not static and predefined, we have identified two requirements: a. the need to easily create new collections and b. the need to extend or modify *collection description* (e.g. the metadata information used to characterize the images included in each collection). We introduced the term *dynamic collection management* to denote the support of automated collection definition and management within an integrated digital library environment.

2. Dynamic Collection Management

This concerns the creation of new collections and the modification of collection description. For each collection added in digital library environment, the corresponding object structure and metadata model must be defined. Collection description can be derived from existing collections by extending the object structure and metadata model, e.g. a collection description can be defined as the descendant of an existing collection description, while additional object parts and metadata fields can also be defined. This feature allows flexibility during collection definition and facilitates collection description in a simplified manner. In the case of AMS DL, each collection consists of the medical images produced by a specific laboratory. Thus, they are characterized by common general metadata and domain specific metadata, useful for researchers in the specific domain [3]. The general metadata model and a basic object structure corresponding to *medical image objects* are used to describe the *Medical Image Collection*. The Medical Image Collection Description is used to easily define collections corresponding to each laboratory as its descendants, while the collection is practically empty. In order to efficiently support dynamic collection management, AMS DL facilitates dynamic interface creation both for processing and cataloguing and collection search. The same interface is used for all collection, while screens are dynamically formed based on collection description.

3. Data and Metadata Representation

The following parts are included in the *medical image objects* belonging in the collections: *Original Image*, a high quality image strongly protected regarding copyright issues, *Derivative Image* produced from the original image to be accessed through the Web, *Watermarked Image*, *Screen Size and Thumbnail Image* and *Image Description* in Greek and English. The original image and the description are produced by the researcher, while all other formats are produced by the cataloguer.

The Dublin Core metadata scheme is used to describe general metadata information [2]. Additional fields were used to represent domain-specific metadata. Since these fields are collection specific, we adopted a DC-like structure for their representation in XML, where the collection is depicted similar to a DC qualifier:

```
<AMS:local_field_name>
  <AMSq:collection>collection_name/< AMSq:collection>
    <rdf:value>local_field_value/<rdf:value>
/< AMS:local_field_name>
```

Medical Image data and metadata internal representation using Content Manager constructs are presented in figure 1. As indicated in this figure, the digital object used to represent Medical Image Objects consists of all derived images and image descriptions. Since the system must support both exact and approximate search in combined multi-valued metadata fields, the searching capabilities of a relational database were too poor to ensure Collection Search application performance. Thus, database search is applied for exact numerical and date metadata field search, while string exact and approximate search is performed using free text search capabilities Metadata information is stored both in the underlying database and within a tag-structured text part in the corresponding Medical Image digital object *(metadata part)*. Different tags are used to support Greek and English languages, while all properties of a specific metadata field, are included within one tag.

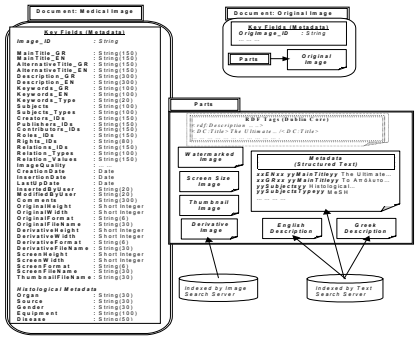

Figure 1. Medical Image Data and Metadata Representation

References

1. Darmoni SJ et.al: CISMeF: A Structured Health Resource Guide. Methods Information Medicine 39 (2001)
2. Weibel S., Hakala J.:DC-5: The Helsinki Metadata Workshop. D-Lib Magazine 4 (1998).
3. Content Manager Documentation. IBM Corporation (2000)