

# Enhance Metadata To Searching Appropriate Image Data

*I. Wayan S.W., Setia W., Arrummaisha A., Dian K.N., Harya I.*

Gunadarma University

(iwayan, setia)@staff.gunadarma.ac.id,

(arroem, harya\_aja, tyaz\_tyaz\_tyaz)@student.gunadarma.ac.id,

www.gunadarma.ac.id

*Abstract:*

*Traditional text database is moving to multimedia database. One of multimedia data type is image data. Currently, retrieving image data is a difficult task. In this paper, embedded enhance metadata to image file can contribute for image data retrieval.*

*Keyword :* content-based searching, image retrieval, metadata.

## 1 Introduction

Technology is rapidly changing time by time. Currently, data in digital format is being used more than data in paper format. The format can be more efficient because digital data are easier to save and retrieve. The consequent is technology needs more efficient way to retrieve the appropriate data.

Retrieving data in traditional and modern can be based on metadata, such as catalog of library. In previous time, image metadata are based on one concept of namespace, because most traditional model is a closed and single system. However, in Internet era, the system is open and multiple system. Used of Internet as based also motivate us to enhance image metadata. Because of this changing, we have new problem, how to make same perception of concept in metadata and how to put the enhanced metadata in image.

Metadata can be classified into two different types [1], Content Independent Metadata and Content Dependent Metadata. Content Independent Metadata captures information that does not depend on the content of the document with which it is associated. Examples of these type of independent metadata are location, modification-date of a document and type-of-sensor used to record a photographic image. This metadata is typically useful in retrieval of documents from their actual physical locations and checking whether the information is

up to date or not. Content Dependent Metadata depends on the content of the document it is associated with. Examples of basic content dependent metadata are size of a document, max-colors, number-of-rows, number-of-columns of an image.

The purpose of this paper is to enhance image metadata which contains namespace and combines it with the digital image itself, in order to gain a better result for retrieval images in SVG format.

It is the job of an image retrieval system to produce images that a user wants. In response to a user's query, the system must offer images that are similar in some user-defined sense. This goal is met by selecting visual features thought to be significant in human visual perception and using them to measure relevance to the query.

Many image retrieval systems in operation today rely upon annotations that can be searched using key words. These approaches have limitations not least of which are the problems of providing adequate textual descriptions and the associated natural language processing necessary to service search requests.

Standard image retrieval technique with metadata is to create a metadata file that refers to the image link. The metadata file is not embedded to the image. In other word, the metadata file and the image are separated [2]. The approach that we want to do is to integrate the metadata file into the image.

## **2 Approach**

The general concept of the process is to integrate digital image and metadata, while today both of them are separated.

By doing that, we hope the retrieval process can be more accurate and efficient, which is shown in figure 2

To facilitate in storing, searching and changing data, we can develop technique to embed metadata in XML format into image with SVG format. There are many advantages when we use XML in embed metadata such as flexibility in data changing, and can be improved to avoid misunderstanding information by adding namespace.

In general, the approaches are:

1. Convert image from any other format into SVG format.
2. Prepare the XMLS template as XML text data, as the experiment with concatenating the namespace which has known from METS generally.



1. Dependent metadata, such as title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, organization, relation, coverage, rights, record type, measurements, material, technique, location, ID number, style, culture, target audience.
2. Independent metadata, such as subject and description.

### **3 Running Example**

In this part, we will do a comparison between two identical images. The two identical images for example are: (1) First image, which is a digital image with metadata integrated in it, but without namespace. (2) Second image, which is a digital image with metadata integrated in it and also with namespace. And then, we will load both images to a different web and search them via web.

The language that we will use for the elements of namespace is English. It does not matter if user enters a keyword in other language, because system will translate that language into English by using online dictionary.

The different process between first image and second image is the process after keyword has been translated into English. For the first image (with metadata but without namespace), system will directly search the image. Searching result for the first image will not always correct, because system will probably find different perception of the keyword meaning and give a wrong image even nothing. For the second image, system will refer the keyword to the namespace (METS) to find the meaning of the keyword. Because of this process, system will show a better result, because system will not find different perception of the keyword meaning. It means that images with namespace in metadata will easier to retrieve, because system will not misunderstanding about the keyword meaning.

### **4 Conclusion**

Standard namespace of metadata in METS is appropriate as first step to retrieve image data. The result of this research is digital images which have namespace in metadata, are easier to be retrieved than the one which does not have namespace in their metadata. Namespace also can avoid wrong perception about the metadata element. One of information in metadata is about object, we insert the information manually to the metadata. In the future we hope the information about the object can be generated automatically.

## References

- [1] Amit Sheth and Vipul Kashyap. Steps towards media-independent search of web-accessible heterogeneous digital data. March 1997.
- [2] Yunmook Nah Sungyong Hong. *An Intelligent Image Retrieval System Using XML*. <http://csdl2.computer.org/comp/proceedings/mmm/2004/2084/00/20840363.pdf>, 2004.
- [3] TASI. *Metadata Standards and Interoperability*. [www.tasi.ac.uk](http://www.tasi.ac.uk), November 2006.
- [4] Setia Wirawan. Penyisipan data medis pada file citra rotgen format svg sebagai alternatif cara penyimpanan data rekam medis. Master's thesis, Universitas Gunadarma, September 2007.