World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:5, 2007

Generic Multimedia Database Architecture

Mohib ur Rehman, Imran Ihsan, Mobin Uddin Ahmed, Nadeem Iftikhar and Muhammad Abdul Qadir

***Abstract*—**Multimedia, as it stands now is perhaps the most diverse and rich culture around the globe. One of the major needs of Multimedia is to have a single system that enables people to efficiently search through their multimedia catalogues. Many Domain Specific Systems and architectures have been proposed but up till now no generic and complete architecture is proposed. In this paper, we have suggested a generic architecture for Multimedia Database. The main strengths of our architecture besides being generic are Semantic Libraries to reduce semantic gap, levels of feature extraction for more specific and detailed feature extraction according to classes defined by prior level, and merging of two types of queries i.e. text and QBE (Query by Example) for more accurate yet detailed results.

***Keywords*—**Multimedia Database Architecture, Semantics, Feature Extraction, Ontology.

I. INTRODUCTION

AS we all know that Multimedia now-a-days, is perhaps the most diverse and rich culture around the globe. The diversity of this aspect of multimedia leads to one dilemma and that is a variety of media formats. Currently, richly used Media Formats such as, Video (MPEG, MOV, WMV etc.), Audio (MP3, OGG, MIDI etc.), Image (JPEG, GIF, PNG) and Documents (PPT, PDF, TXT etc.) [1] serve as a source of information and archival and thus the fact arises that such a large collection of media should be organized and properly indexed for search and retrieval in a standard way, keeping aside the type and domain of media.

The need to have standards in the current computer age is proving to be a must for all fields that require it and thus Multimedia is no exception. XML [2] stands as a powerful and widely accepted UMF (Universal Media Format) and thus it can be used as transfer and storage medium for media. XML Schema[3] can be utilized as a way to define and store Metadata. A media unit must have a Metadata [11] if it is required to be transferred over a rich medium such as the Internet. This Metadata should follow some standards like; IEEE LOM [4, 5], MPEG-7[6, 7], Dublin Core [8] etc.

The Search of multimedia objects is facing a well known problem that is defined as semantic gap [9, 10]. The process of querying the multimedia Objects is complex depending upon not only what and how the information can be retrieved, but also how this information can be link / merge logically to reduce the semantic gap. As the three level of complexity in retrieving the information defined in [10] are; Level 1, 2 & 3 are on basis of Primitive, Logical and Abstract features respectively. But there exists a semantic gap among these three levels. Some of the work is done on reducing the semantic gap between Level 1 and Level 2 and still a lot more to be done. But as far as Level 2 and Level 3 is concerns very little is done.

Manuscript received March 31, 2005.

Mohib-ur-Rehman is with Muhammad Ali Jinnah University, Islamabad Campus, Jinnah Avenue (Blue Area) Islamabad, Pakistan (phone: 92-300-5115672; e-mail: mohib.ur.rehman@ gmail.com).

Imran Ihsan is with Muhammad Ali Jinnah University, Islamabad Campus, Jinnah Avenue (Blue Area) Islamabad, Pakistan (phone: 92-333-5118897; e-mail: iimranihsan@gmail.com).

Mobin Uddin Ahmed is with Muhammad Ali Jinnah University, Islamabad Campus, Jinnah Avenue (Blue Area) Islamabad, Pakistan (phone: 92-300-5115603; e-mail: mobin.uddin@ gmail.com).

Nadeem Iftikhar is with Muhammad Ali Jinnah University, Islamabad Campus, Jinnah Avenue (Blue Area) Islamabad, Pakistan (phone: 92-300-5164815; e-mail: nadeem @ jinnah.edu.pk).

Muhammad Abdul Qadir is with Muhammad Ali Jinnah University, Islamabad Campus, Jinnah Avenue (Blue Area) Islamabad, Pakistan (phone: 92-300-5011600; e-mail: aqadir @ jinnah.edu.pk).

II. PROPESED FRAMEWORK

Keeping in view the nature and demands of media, we have proposed a generic framework for multimedia database system (see Figure 2). Some of the key modules of the system are explained below:

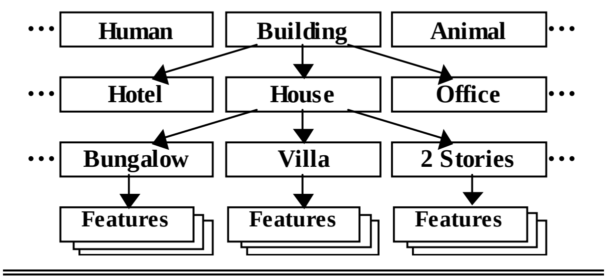

Figure 1: Hierarchy of Library Classes

A. Feature Extractor:

The main job of this module is to extract all possible features from a media file i.e. images, audio, video or text [12, 13, 14]. It consists of number of different components. These components are special kind of filters that are applied on a media and can generate the corresponding features. The module is flexible in nature and can adapt to any new filter that is available. The media undergoes all of the filters one by one and generates a complete set of extracted features.

B. Semantics Libraries

This module contains a number of classes that describe different features. Each class can have sub classes within it.

This module is also flexible and new classes/sub classes can be added. Features in each class can be compared with the extracted features in order to classify a media. These classes can be of different type such as Buildings, Humans, and Animals etc. Taking an example of Building Class, there can be a number of sub classes such as Houses, Hotels, Offices, and Universities etc. Further more a House Subclass can have another set of sub classes such as Bungalows, Villas, and Two Stories Houses etc (see Figure 1).

C. *Metadata Schemas*

There are number of different Media Metadata Standards available but any one standard cannot be applied to all types of media. For example IEEE LOM Standards are used for multimedia files that are for Education purposes, where as MPEG 7 are used for Motion Pictures such as videos. Therefore to make the system as generic as possible, there can be more than one metadata standards. This module contains such Metadata Standards that are used to collect the metadata of a media. The process of gathering metadata is semi-automatic, where the user only needs to specify as few as possible i.e. with the help of "User Profiles".

D. *Ontology*

Ontology contains a description of the learning concepts that will add meaning to it. It is a kind of a dictionary that contains knowledge representations that are very similar to "Libraries" in structure. Thus it consists of different classes that are used to expand the query keyword within a specific domain or class. The module is adaptable to new classes as well.

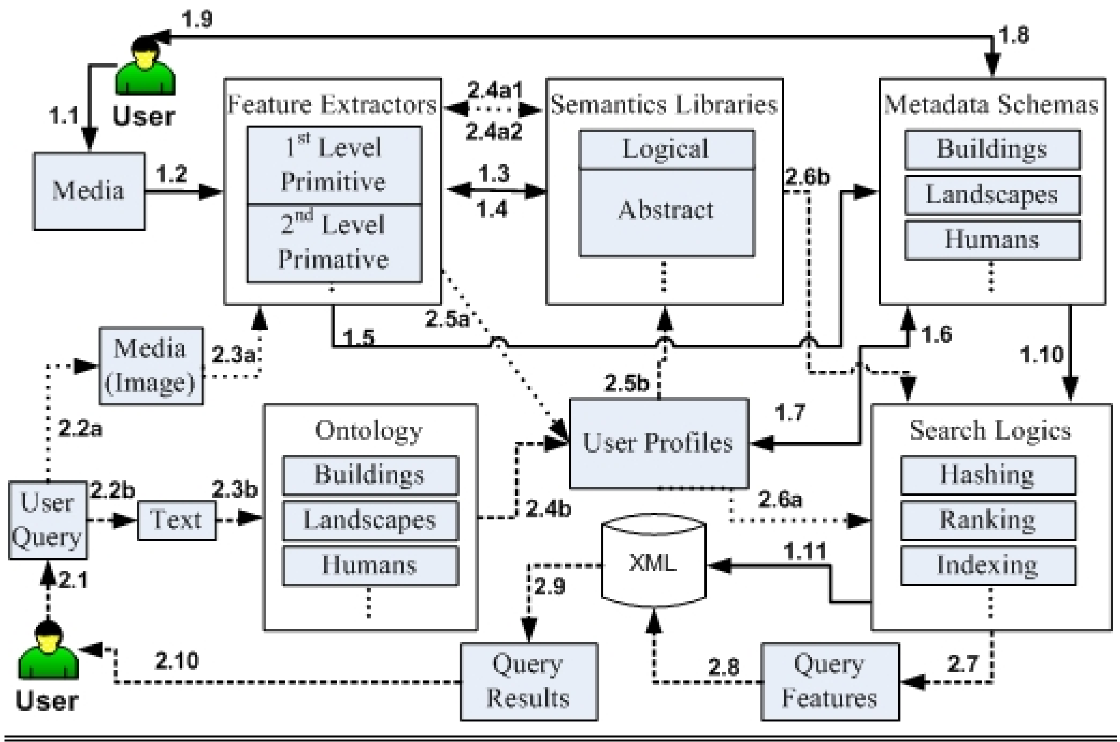

Figure 2: Complete Architecture

E. *User Profiles*

The user needs to define profiles that can interact with the system. The definition of a profile is a context in order to specify a domain and some basic data about the user. A user can specify more than one profile according to his need. These profiles play key roles in the collection of metadata and by making a query more related to what is required.

F. *Search Logics*

Search Logics is a kind of a search engine with all the possible components that are required for efficient search and retrieval of a media file. These components include indexing, hashing and ranking etc.

G. *XML Based Database*

This database contains XML files that have three different sections. The first section contains the media file reference, second has the metadata extracted using profiles and given by the user according to standards applied and the last section has the content metadata extracted automatically using Feature Extractor. A sample XML file is shown in Figure 3.

International Science Index, Computer and Information Engineering Vol:1, No:5, 2007 waset.org/Publication/8121

```
- <mediafile id="MMDF000120050311">
  - <media>
      <url>.../filename.jpg</url>
  </media>
  - <contentmetadata>
      - <shapes>
          <rectangle>10,20,56,98</rectangle>
          <rectangle>21,10,100,12</rectangle>
          <rectangle>152,200,220,250</rectangle>
          <circle>40,50,20</circle>
          <circle>17,23,6</circle>
          ...
      </shapes>
      - <colors>
          ...
      - </colors>
      ...
  </contentmetadata>
  - <metadata>

          <filename>filename</filena
          me>
      <filetype>jpg</filetype>
      <filesize>500kb</filesize>
      <filewidth>600px</filewidth>
      <fileheight>600px</fileheight>
      <fileowner>imran ihsan</fileowner>
      ...
  </metadata>
</mediafile>
```

Figure 3: A Sample XML File

III. Flow of The System

A. Media Submission

1.1. **User Input:** User logon and submits any kind of supported media file to the system

1.2. **Feature Extraction:** The 1^{st} Level Feature Extractor will extract generic features from the input, so that it can help Libraries in identifying the possible related classes of the input media.

1.3. **Semantics Libraries:** The 1^{st} level extracted features of the media file are sent to the libraries for related objects classes' identification. After processing the input, the Library module sends a list of related objects & their classes back to the Feature Extractor Module for 2^{nd} Level of feature Extraction.

1.4. **Feature Extraction:** The 2^{nd} Level feature extractor will now extract more detailed classes specific feature extraction.

1.5. **Metadata Schemas:** The extracted features, related object Ids and other information are sent to the Metadata Schemas Module for Metadata collection. This Module fills in the basic information (e.g. filename, Media type, etc), according to Schema.

1.6. **Input to User Profiles:** The Metadata Schema Module asks User Profile Module for possible input to metadata and sends the related Metadata Schema to it.

1.7. **Output of User Profile:** The User Profile Module fills in the information available in the user's profile and sends back to the Metadata Schemas Module.

1.8. **Metadata to User:** The Metadata Schemas Module sends all collected information according to the Schemas to the user.

1.9. **User Input to Metadata:** The user fills in missing fields and modify already filled by system (if required) and sends back to the Metadata Schemas Modules.

1.10. **Input to Search Logics:** The extracted features (Contents Metadata), Metadata and related object Ids are send to the Search Logics Module for further processing, required for search and retrieval.

1.11. **XML Storage:** After applying all available techniques (e.g. Hashing, Ranking, Indexing) on the input, this module stores the output to the XML based database (an XML file).

B. Media Search

The system provides two types of media search. One is through the text query which is a normal search where as second type is Query by Example, where the user provides a media as an example in order to find the appropriate results. Next we have explained how the system incorporates both queries.

Text Query

2.1. **User Query:** User posts the query according to certain requirements.

2.2. **Query Type:** There can be of two types and shown below.

- 2.2.a. **Media:** Audio, Video or Image file
- 2.2.b. **Text:** Key words

2.3.

- 2.3.a. **Feature Extractors:** 1^{st} level of features will extract generic features from a media example for its classification.
- 2.3.b. **Ontology:** Using the knowledge structure, Ontology expands user query.

2.4.

- 2.4.a.
  - 2.4.a.1. **Classes Identification: O**n the basis of 1^{st} level extracted features, the Semantics Libraries identifies classes of the features/Objects.
  - 2.4.a.2. **Feature Extractor:** after getting the possible classes of the features, the feature Extractor performs 2^{nd} level extraction which is specific to the classes identified by the Semantics Libraries.
- 2.4.b. **User Profiles:** Expanded query is quantified on the basis of user's search profiles.

2.5.

- 2.5.a. **User Profiles:** The Expanded query is quantified on the basis of user's Search Profiles.
- 2.5.b. **Classes Identification:** Domains are specified according to the expanded query's knowledge structure.

2.6. **Search Logics:** The refined query is qualified by search logics.

2.7. **Query Features:** Now the query is complete for the searching the database.
2.8. **XML Based Database:** Query is posted to database for execution.
2.9. **Query Results:** Results are calculated and stored here.
2.10. **Results Display:** Results are displayed to the user in a proper format.

IV. Main Features

A. Semantics Libraries

What we have introduced, is the concept of Semantics Libraries. It is Knowledge base Library. It has two main parts, Logical Libraries and Abstract Libraries. The Logical Libraries are to reduce the semantic gap between Level 1 and Level 2 where as Abstract Libraries are to reduce semantic gap between Level 2 and Level 3.

Logical libraries use primitive features to identify and classify the objects. Logical libraries contain a hierarchy of object in the form of classes. The root class contains a naïve idea about an object that becomes concrete as it goes deeper in hierarchy. The leaf nodes have all possible primitive features that are required to identify an object. Extracted primitive features are compared with leaf node features and after detecting the related nodes, the reverse path to the root node defines the concept of the object detected. Thus we can say that by using primitive features level 2 of complexity can be obtained through logical libraries. Taking an example, the primitive features that Logical Library get circle shape, white and gray of An example of Logical Library is shown by a tree type data structure in figure 1.

As Logical Libraries help reduce the semantic gap between level 1 and 2, similarly Abstract Libraries help to reduce the semantic gap between Level 2 and 3. How this gap is reduced is based on the structure of Abstract Libraries that take the objects and their concepts and try to correlate them in order to describe the abstract concept of the media. These libraries provide the possible relations between different objects and concepts.

B. Levels of Feature Extraction

We have introduced two levels of feature extraction in the architecture. In first level the basic primitive features are extracted so more basic/generic feature extractors are required. In the second level after getting the possible classes of basic primitive features from Semantic Libraries, more detailed and specific features are extracted according to the classes. So in this level more specific to the domain / classes and state of the art feature extractors are required. This is very demanding area and a lot of work has been done in this regard, but still a lot more is required. As the architecture support plug in functionally, new and advanced extractors can be added any time. Even the levels of feature extraction can be increased according to the requirement of the domain of the system, one is implementing on the generic architecture.

C. Merging of Text & Media Query

As one can see in the Section III Flow of the System that user can pose both kinds of queries at the same time i.e. Text and Query by Example. These queries are merged after 2.6a and 2.6b step (see figure2) in the search logics module. This may help in retrieving more accurate yet detailed results. Exactly how these queries are merged is still an open area for research, but the basic logic may be based upon semantic libraries.

V. Conclusion

In this paper we have suggested a Generic Multimedia Database Architecture with three main features; Semantics Libraries, level of feature extraction and merging of text and media query as explained in Section IV. This is an effort in defining a standard architecture for all multimedia applications/systems. As the Architecture is flexible, so new features can be plugged-in and unnecessary parts can be unplugged as required. We strongly believe that using semantics libraries will be helpful to reduce the semantic gaps in different levels of complexity (as explained in Section IV (A)).

References

[1] J.R. Smith and S. F. Chang, “Visually Searching the Web for Content” IEEE Multimedia Volume 4, Issue 3, July 1997, pp 12-20.
[2] Extensible Markup Language http://www.xml.org
[3] W3C: XML Schema Part2: Datatypes, http://www.w3.org/TR/xmlschema-2/#dateTime, W3C Recommendation 02 May 2001, Editors: Biron P V, Malhotra A
[4] IEEE Learning Technology Standards Committee ©2005, http://ltsc.ieee.org/wg12, IEEE Standards for Learning Object Metadata (1484.12.1)
[5] IEEE http://www.ieee.org
[6] International Organization For Standardization, Coding of Moving Picture & Audio, ISO/IEC JTC1/SC29/WG11 http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7/mpeg-7.htm
[7] MPEG-7 Overview: ISO/IEC JTC1/SC29/WG11 N4980, Kla-genfurt, July 2002, Editor: Martínez J M (UPM-GTI, ES)
[8] Dublin Core Metadata Initiative, http://dublincore.org
[9] Chitra Dorai, Andreas Mauthe, Frank Nack, Lloyd Rutledge, Thomas Sikora, Herbert Zettl, “Media Semantics: Who Needs It and Why?” Multimedia’02, December 1-6, 2002, Juan-les-Pins, France. Copyright 2002 ACM 1-58113-620-X/02/0012
[10] Lynne Dunckley, “Multimedia Databases – An Object-Relational Approach” ISBN # 0 201 78899 3, pub 2003, Chapter 5, page 117-122.
[11] W3C Metadata Acitivity,W3C organization http://www.w3.org/Metadata/
[12] Lynne Dunckley, “Multimedia Databases – An Object-Relational Approach” ISBN # 0 201 78899 3, pub 2003, Chapter 5, page 123-135.
[13] S. Kulkarni and B. Verma, "An Autoassociator for Automatic Texture Feature Extraction",Proceedings of the 4th International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'01) 0-7695-1312-3/01 © 2001 IEEE
[14] Job Oostveen, Ton Kalker, Jaap Haitsma, "Feature Extraction and a Database Strategy for Video Fingerprinting", 5th International Conference, VISUAL 2002 Hsin Chu, Taiwan, March 11-13, 2002. Proceedings Lecture Notes in Computer Science , ISSN: 0302-9743.