Universal Metadata Definition

Mohib ur Rehman, Mohammad Haseeb Anwer, Nadeem Iftikhar

***Abstract*—**The need to have standards has always been a priority of all the disciplines in the world. Today, standards such as XML and USB are trying to create a universal interface for their respective areas. The information regarding every family in the discipline addressed, must have a lot in common, known as Metadata. A lot of work has been done in specific domains such as IEEE LOM and MPEG-7 but they do not appeal to the universality of creating Metadata for all entities, where we take an entity (object) as, not restricted to Software Terms. This paper tries to address this problem of universal Metadata Definition which may lead to increase in precision of search.

***Keywords*—**Metadata, Standard, Universal, XML.

I. Introduction

EVERY object must have a definition i.e. some keywords that define its structure and working, thus we can refer to this object through these keywords and a collection of these objects. These objects, even if formed, will also have the same definition. The objects must have some relationships to be of some use i.e. no object is useful in a standalone manner. This brings up an interesting fact that today all forms of objects that exist do not have proper implication of the above definition and by term object we refer to a number of fields such as Hardware, Software, Multimedia and other forms of physical or virtual structures. To define an object one needs to look into the specific details of that object and its relationships. Metadata [1], as defined in the proper sense is a step towards defining the information related to an object.

II. Basic Idea

Today, many forms of objects such as multimedia and software use Metadata to define the structure and information about the media. The problem is with the low usage of a standard that does not really exist in terms of definition and structure. Every field (Hardware, Software, Multimedia and other forms of physical or virtual structures), uses its own way of defining the object in hand. The need is to create a standard metadata definition for each of the fields. Once a user tries to find an object of his need he must look into every possible relationship for that term. For Example,

A user tries to find a term "audio", this particular term has many ways to it. An audio term can relate to a software, hardware, media, files etc. Thus if he does not know about the relationship between all these aspects, he is mighty sure to be stuck and thus his search is lost.

Search engines today such as Google [2] try to bring every possible result to a user's query through minimal possible retries and maximum probability of success. But it does not tell a user what he is searching for. Suppose that the user wanted to search for hardware and wanted to buy it after he came to know about the exact specification of the hardware. The search might not possibly bring him the physical object but can relate links to that hardware.

IEEE [3] defined standards about certain media types such as software and multimedia files that are starting to take roots but are not widely used throughout the world. USB [4] is another example of minimizing the amount of brands in the world created by many of the companies. The idea is to give the user a "Z" after processing "A-Y", where “Z” is the common interface such as the USB and “A-Y” is the company’s own design and specification of the hardware, which however does not possibly restrict them. This idea of having something common throughout the lineup of these objects is the basic concept that can help in the creation of metadata based upon common fields and universally accepted standards throughout the world.

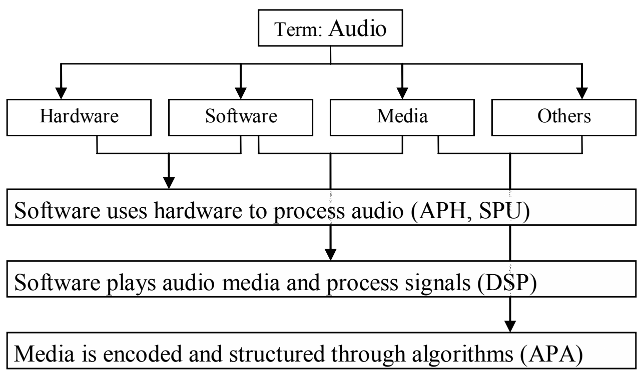

Figure 1: “Audio” Meta term & Relationships

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).
Mohammad Haseeb Anwer is with Muhammad Ali Jinnah University, Islamabad Campus, Jinnah Avenue (Blue Area) Islamabad, Pakistan (phone: 92-596-541520; e-mail: mhaqs@ hotmail.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).

Open Science Index, Computer and Information Engineering Vol:1, No:5, 2007 publications.waset.org/5777.pdf

This is where the idea of a Universal Metadata Definition comes into play. We take out our previous example i.e. "audio" See Figure 1. Any object related to audio definition must have something in common and something that can define relationships between these objects. Co-Relationships existing between objects help us to create a graph of these objects and extract Metadata.

Figure 1 shows a clear relationship between the areas, and at some place a "Many to Many" relationship and which is why the use of a graph is necessary to draw it. The next step is to extract Universal Metadata out of these sources.

Hardware - Name, Specs, Features...
Software - Name, Specs, Features...
Media - Name, Specs, Features...
Other - Name, Specs, Features...

Here, Specs tell us about the capability of the object such as the hardware and Features tell us about the add-on features and the way they work.

The above example maps the problem quite a bit. We can take another example to clear out the concept of creating and finding relationships.

A user types in "64 bit". This term is very ambiguous in a lot of aspects and to name a few we have 64 bit processors, software, algorithms etc. We suppose that the user was most probably searching for a 64 bit family of objects such as the 64 bit processor.

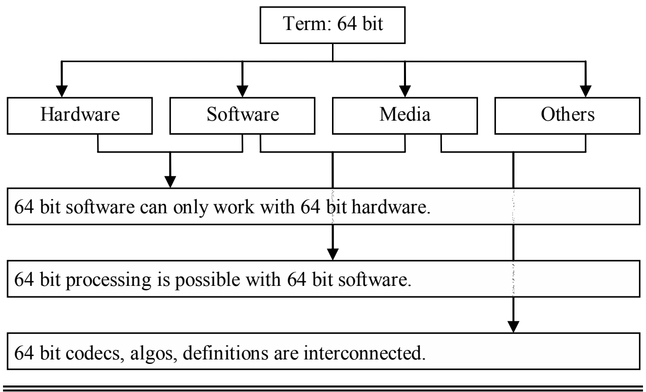

Figure 2: “64 bit” Meta term & Relationships

The above fields in Figure 2 have one very clear distinguishing aspect and that is the calculation of 64 bit structures, which are highly accurate as well as complex at the same time. Thus they contain themselves in a family of 64 bit objects and do not fall under the category of lower bit families but this does not mean that they do not have relationships with other lower or higher bit structures.

Hardware - Name, Type, Specs, Manufacturer...
Software - Name, Type, Specs, Manufacturer...
Media - Name, Type, Specs, Manufacturer...
Other - Name, Type, Specs...

Notice that all of these sources have a lot in common but why is the keyword of "Manufacturer" there is a big question? The answer is that many companies create several architectures based on different schemes but similar specs and types. There remains only one way to relate these objects with the members of their family and that is through relationships between other nodes and parent objects. For example, Intel[5] manufactures 64 bit processors and so does Athlon[6] but the architecture is different and where the architecture differs, the platform is the same i.e. IBM[7] compatible structure of PCs. Microsoft[8] creates 64 bit versions of Windows for IBM based PCs and thus Windows is a common relation between both the companies.

Apple produces PowerPC [7] structured home computers and they differ completely in terms of hardware and software with Windows. But there is one thing that remains exclusive in both the cases; third party developers create products for one or more types of manufacturers with the same name, versions, specs etc. but with different structure. E.g. Macromedia[9] produces Flash MX for both PowerPC and IBM compatible PCs.

This leads us to one conclusion that Metadata can be universalized through extraction of "Common Keywords", found out of relationships between objects of different areas.

III. UNIVERSAL METADATA DEFINITION

Every XML [11] file can have a “Definition” more often referred to as “DTD”, also used as “Schema” [10] in many fields. We will refer to the “Definition” as an XML file itself. As in this case the XML file is representing a Metadata or Definition of another object. As the XML tags are based on relationship rules, we can relate two XML files through some common term. A term in a hierarchy can have more than one parent. Thus, these terms can have “Many to Many” relationships. This relationship between the terms will help to search for two or more objects of the same family. Relationships are also responsible for creating a hierarchy of objects. Whenever a search is initiated, the user will get the most relative definition of the object (search result) and may not see the hierarchy following it but this is left as a choice for the user to choose whether he wants to see the hierarchy or not.

When the Search Engine finds the relative XML definitions according to the User’s query, it must present them in a manner that supports relationships. Graphs are structures that support “Many to Many” relationships among objects. When displaying the hierarchy of the search result, the actual mapping of the objects could be done through a graph.

Open Science Index, Computer and Information Engineering Vol:1, No:5, 2007 publications.waset.org/5777.pdf

IV. Applications

The amount of work currently done in the field of searching thorough web enabled interfaces such as Google, allow the user to utilize a very fluid interface and thus obtain the most relative results. The use of relationships in this context can lead this search to a better state. The concept of Universal Metadata Definition (UMD) is applicable to any system that can search for objects and this holds not only for web semantics but for physical objects such as the hardware. Once developed, the UMD specification can apply to about any interface; let it be Search Engines, Spiders, and Registers etc.

V. Related Work

A lot of efforts have bean made to standardize the metadata but all these efforts belongs to some specific group or class. Here we are mentioning few of those.

The **Dublin Core Metadata Initiative** (DCMI) [12] is perhaps the largest candidate in defining the Metadata. It is simple yet effective element set for describing a wide range of networked resources and comprises 15 elements, the semantics of which have been established through consensus by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship. According to (Baker, 2000), "Dublin Core is a Small Language for making a particular class of statements about resources". Dublin Core is more suitable for document-like objects (because traditional text resources are fairly well understood). Its use with particular non-document resources depends on how much their metadata bear resemblance to typical document resources and also the rationale of metadata.

IEEE LOM [13], is a metadata standard for Learning Objects. It has approximately 100 fields to define any learning object. The scope of this metadata standard is very limited or you can say specific to Learning Objects.

Medical Core Metadata (MCM) [14] is a Standard Metadata Scheme for Health Resources. It was developed to enhance Internet health document retrieval. So the scope of this standard is focused on Health domain and also limited documents.

MPEG-7 [15] multimedia description schemes (DS) provide metadata structures for describing and annotating multimedia content. Most of the elements focus on low-level attributes of A/V content like shape, color or motion, are intended to be extracted automatically. Some DSs are defined for content management: creation/production, media information and usage (rights).

VI. Conclusion

What we proposed, is an application of Metadata in an abstract manner, and no specific domain. This ultimately achieves the goal of creating a universal standard for metadata definition. Current technologies and systems can exponentially improve their Search Mechanisms through the implication of this Definition. The proposition of graphs can help to map complex relationships in an organized manner and XML stands as a powerful and widely accepted UMF (Universal Media Format) and it can be used as a transfer and storage medium.

References

[1] W3C Metadata Acitivity,W3C organization http://www.w3.org/Metadata/
[2] Google Corporation, http://www.google.com/intl/en/about.html
[3] IEEE http://www.ieee.org
[4] Universal Serial Bus (USB), http://www.usb.org/home
[5] Intel Corporation , Processors, http://www.intel.com/products/processor/
[6] AMD Corporation, Athlon 64 bit Processor, http://www.amd.com/us-en/Processors/ProductInformation/.
[7] IBM Corporation, PowerPC, http://www-03.ibm.com/chips/products/powerpc/
[8] Microsoft Windows http://www.microsoft.com/windows/.
[9] Macromedia Flash MX, Macromedia Corporation http://www.macromedia.com/software/flash/
[10] W3C: XML Schema Part2: Datatypes, http://www.w3.org/TR/xmlschema-2/#dateTime, W3C Recommendation 02 May 2001, Editors: Biron P V, Malhotra A
[11] Extensible Markup Language http://www.xml.org
[12] Dublin Core Metadata Initiative. < http://dublincore.org/documents/>, 2002.
[13] IEEE Learning Technology Standards Committee, http://ltsc.ieee.org/wg12, IEEE Standards for Learning Object Metadata (1484.12.1)
[14] Darmoni, Thirion, "Metadata Scheme for Health Resources" *American Medical Informatics Association, 2000 Jan–Feb; 7(1): 108–109.*
[15] MPEG-7 Overview: ISO/IEC JTC1/SC29/WG11 N4980, Kla-genfurt, July 2002, Editor: Martínez J M (UPM-GTI, ES)

Open Science Index, Computer and Information Engineering Vol:1, No:5, 2007 publications.waset.org/5777.pdf