Using Metadata to Enhance Multimedia Query Semantics[1]

P. Merialdo, G. Sindoni - Università degli Studi di Roma Tre

Via della Vasca Navale, 84 - 00146 - Roma - Italy

{merialdo, sindoni}@inf.uniroma3.it

{merialdo, sindoni}@dis.uniroma1.it

Abstract

One of the main problems to face in the design and implementation of advanced multimedia information systems is how to give enough semantics to data and queries. This problem arises in particular from the need to share multimedia data among different applications in different domains in geographically distributed environments, like the ones emerging from the "vertical" growth of the WWW.

We think that only with a massive use of *metadata* it is possible to achieve the goal to realize multimedia information systems capable of constructing, querying (also on content) and browsing collections of documents.

We show how an increasing use of metadata in system design improves data and query semantics. We present a general reference architecture for multimedia information systems with the focus of attention on metadata. A multimedia information system in a hospital enviroment, designed following these principles, is under development.

1. Introduction

Data semantics concerns the meaning and the use of data [WOO85]. One of the best interpretation of data semantics has been sentenced by Navathe [SHE96], who defines data semantics as “anything about data that has a conceivable practical consequence in application". Semantics covers aspects like interpretation, use of data and rules used by people to convert data into information. As semantics is dependent on humans, it is difficult to catch it in the context of machines.

In the multimedia information system context, the importance of human factors in semantics definition is even higher. Multimedia objects are in fact unstructured: there are not well defined and well founded structures and rules describing their nature and their use. Hence, it is not possible to query, for example, a database with images basing on values, because there are no values at all in an image content and it is not clear which are the values that can represent an image. There are two main issues in the multimedia quering problem: *to define meaningful representatives for multimedia objects and to define similarity criteria for objects comparison and retrieval*. Besides, the same multimedia object is usually accessed and manipulated for different purposes and since most of the times it can't be replicated for efficiency reasons, this means that giving unique semantics to multimedia is not only impossible, but even not so useful [CKO92].

To give multimedia objects different semantics we need a way to describe and query them with respect to their media types, contents and different applications and domains. To do this we need *information describing the nature of the data and the set of interpretations that may be placed on the data* [SGG96]. This is probably one of the best definitions given to the concept of *metadata*, which are broadly known as *data about data*.

In this paper we show how metadata can be useful in defining correct semantics to multimedia objects with respect to their accessibility and manageability by different applications with different aims in a distributed environment. Besides, we address the need of a novel approach in developing

[1] This work has been partially supported by MURST, Università di Roma Tre and Consiglio Nazionale delle Ricerche.

architectures for distributed multimedia information systems with the focus of attention on metadata. Moreover, we address the hypermedia paradigm as a mechanism to semi-automatically define metadata over a multimedia object.

To date, the interest of researchers on the use of metadata is motivated by two main factors: on the one hand, metadata use has been investigated in order to manage efficiently complex multimedia documents [KSH94], on the other hand, this approach seems promising in order to solve problems in heterogeneous and autonomous systems [KSH95]. This, also because of the growth of the global information network, is stressing the development of heterogeneous, autonomous, distributed and multimedia. For the former problem, metadata assumes importance as they allow to perform content based search on non-traditional data such as images [GFS94], video [JHA94], speech [GSW94], and so on. For the latter, metadata are used in order to browse, navigate and retrieve information with the focus of attention on the information content, without relying on the structure and the organization of the single databases [SSK95].
A lot of modern information systems present both the needs discussed above. For example, in the medical context, diagnostic exams produce a huge amount of data from different media; besides, health care organizations are composed of distributed structures which manage their own information in an autonomous and heterogeneous environment. Crime fighting organizations need to manage a lot of multimedia data (photos, shadowing movies, tracks from phone interceptions) and to correlate this data with existing databases, often belonging to different agencies. Teaching issues have often the need to explicitly correlate text with other kind of data, as in arts history, whose textbooks contain a lot of explicit references to the single parts in the described pieces of work. For an efficient use of these data, it is important for their owners to access multimedia data by their content; moreover, effective solutions are needed in order to share information about objects whose history involves more than one organization.
In this paper we address the massive use of metafata as an useful methodological approach to the design of multimedia systems. We show how, through the introduction of metadata representing the relationships between multimedia document components, it is possible to query on content a database with still images. Besides, we propose a general pourpose metadata classification that can be used as a reference framework for the design of system architectures.

Section 2 presents an example about how metadata can improve multimedia semantics. In Section 3 we discuss features and classifications of metadata from literature and a specific proposal for our problems. Section 4 presents our reference architecture. Conclusions and future work are illustrated in section 5.

2. How the Use of Metadata Enhances Multimedia Semantics

In this section we present an example that explains how, using metadata, it is possible to increase data and query semantics for a collection of multimedia documents. We consider documents which are composed by text and still images. Of course it is possible to extend the document with components from other media, such as video, graphics, animation, sound, and so on. Moreover, the single components need not to be stored in the same site, but we can have a *distributed multimedia environment* in which multimedia documents can be created taking components from different sites, for instance on the Web. Multimedia document semantics regards how to use each component of the document and how to provide solutions to improve querying and browsing power on the collection.
Fig. 1 shows three schematic solutions for the management of a collection of distributed multimedia documents. In the lower situation, text component and image component are stored in flat files. To allow querying about the document as a whole, metadata about document composition (e. g. *URL* and *media type* of each component) and metadata about the whole document storage (e.

g. *last modification date of the last modified component*) are needed. Moreover, data about the format and compression technique of the images must be provided in order to allow presentation.
This information is given by the document metadata and allow to query the system for both single components of the document and the document as a whole. For instance it is possible to formulate query as

```
retrieve image component from documents dated after 31/12/95
```

Of course this is a very simple system, but it is important because it emphasizes how the expressive power of the query language depends on the metadata it can manage: the raw-data of the image component and the ASCII data of the text component do not contribute to satisfy the query.
If users are interested in richer queries, this solution is not satisfactory. One step beyond can be done adding more metadata to our document representation. In particular, systems should be able to support querying on components content; for instance, a radiologist querying a report collection could be interested in retrieving images where a given *remark* has been diagnosed or images concerning a given *anatomical structure*. This can be achieved only adding information describing image contents.
We think these metadata represent pieces of information useful for the *interpretation* of the data and not only for their retrieval: their main purpose is to increase the semantics of data and not the performances of the system. In the same manner we consider keywords on the text component, if we consider them as text descriptors, i. e. metadata on text. Moreover, as some of the keywords can refer to objects contained in the image component, if also information about these relationships is managed, such keywords can be used to query the whole document and both its single components. For instance, in the radiological example domain, if the multimedia document represents a report composed by a textual component and by the images the text describes, terms can be found that are *remarks* and/or *anatomical structures*. In this way, the keyword terms contained in the textual part of the multimedia can be seen as metadata also on the image component.
This kind of management is more powerful than the previous one but it still has some limitations. In particular the system describes the content of the components but does not permit to locate the objects inside the components. For instance, in the report example, with this kind of management it is not possible to locate the remarks inside the images and, if more than one remark is detected, it is not possible to distinguish them.

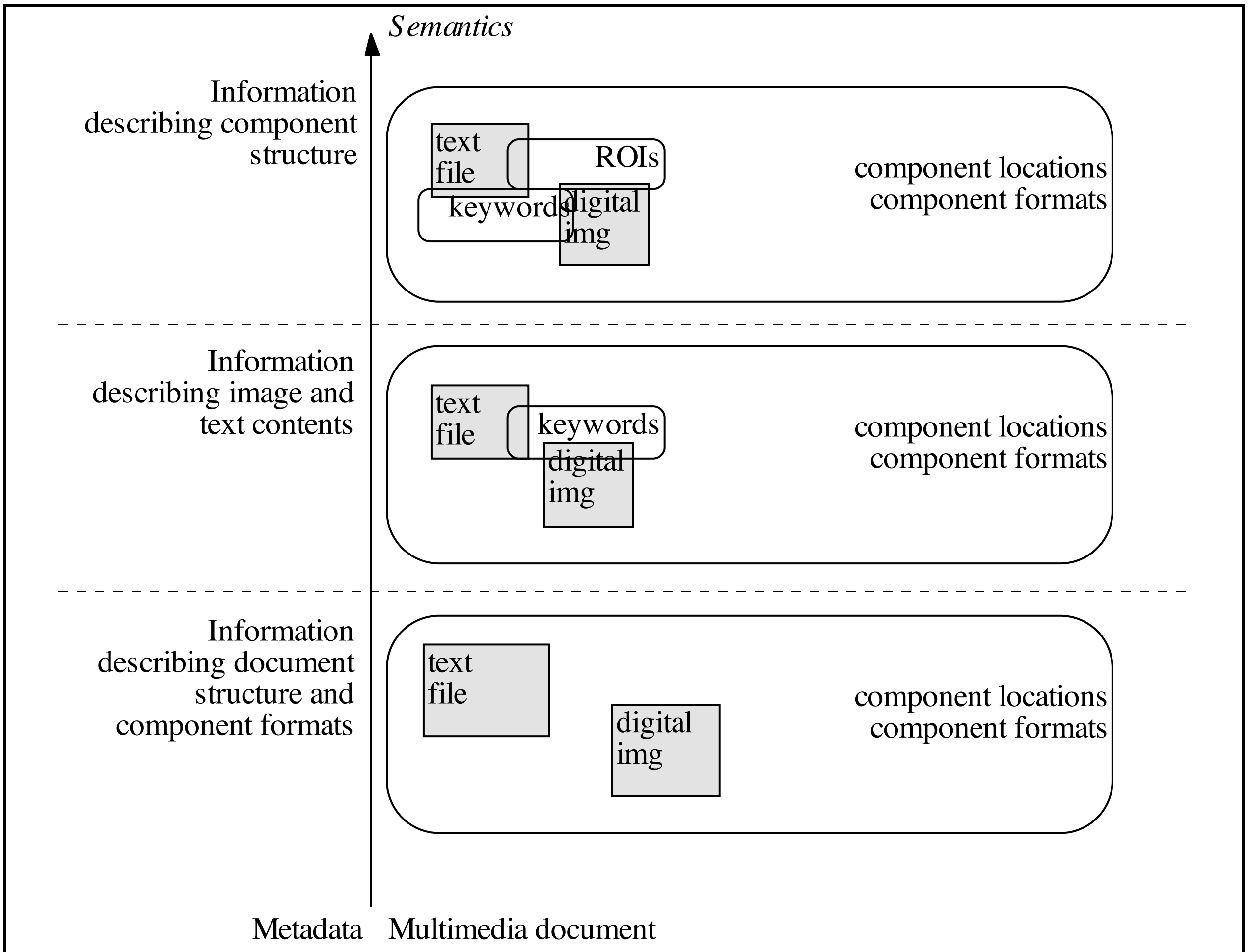

Fig. 1 The increase of multimedia document semantics through metadata.

In the upper part of Fig. 1, a structure is depicted where a new type of metadatum allows to solve the probems existing with the previous ones. We reach in this way the top of our semantics axis.
In particular metadata are used to describe those part of an image which terms refer to and to maintain information about relationships between such parts of image and terms. Queries like:

```
retrieve parts of images where an ulcer has been remarked
```

can so be satisfied.
With the concept of Region Of Interest (ROI), we can *go inside* multimedia. A ROI is a zone of an image which delimit the edges of some interesting feature of the image. In the hypermedia context, it is possible to define anchors on ROIs into images, allowing to attach links to ROIs. The concept of ROI can be easily extended to texts (a ROI is a sequence of characters) and to other type of multimedia objects: ROI descriptors depend on the media-type.

3. Classifications of Metadata

In order to better understand which metadata have to be used for what, some metadata classification methods has been provided.
The proposal by Shet and others [KSS95] considers three kinds of basic metadata:

- **Content-dependent metadata**, which depend only on the content of the original data and can not be extracted automatically: they can be determined exclusively by cognitive or intellectual processes.
- **Content-descriptive metadata**: whose aim is the description of the data; they are sub-classified as **domain-dependent**, which use domain-specific concepts (for example, metadata describing the interpretations of image objects for a specific domain), and **domain-independent**., which rely on domain-independent concepts (for example metadata describing the structure of a multimedia document).
- **Content-independent metadata**: which do not depend on the content (for example metadata about a document files: location, dimensions, ecc.).

With respect to systems managing multimedia document, this classification lacks in describing explicitly the media type dependence and the specific use of metadata for the description and the management of the structure of the single multimedia component (for example, the concept of ROI is Media-type dependent and is used in order to describe a component structure). Besides, considering Domain-dependent metadata as a subclass of Content-descriptive, it is impossible to take into account all those metadata which are Content-independent, but Domain-dependent. For example, in the medical domain, information about standard coding-systems are metadata on coded medical terms, but they are independent from the content of any document in which a coding-system is used.

K. Bohm and T. C. Rakow [BRA94] identify six categories:

1. **Metadata for the Representation of Media Types**: they provide information in connection with the presentation of multimedia data, that is, format, coding and (if applied) compression techniques.
2. **Content-descriptive Metadata**: these metadata are used to describe the content of a multimedia document and they can be media-type dependent. They are determined intellectually or by automatic or semi-automatic tools.
3. **Content Classification Metadata**: they are additional information that can be derived from the document content. In the medical context these information can be useful as they can describe, for instance, the categories of people allowed to read the document.
4. **Document Composition Metadata**: these metadata describe knowledge about relationships between document components and the "role" of each component in the document.
5. **Metadata for Document History**: they provide information which are related to the history and the status of the document or its components.
6. **Metadata for Document Location**: the information provided by these category of metadata permit the access to data. They are important with multimedia documents which, because of the huge amount of storage needed, cannot be duplicated.

From the analysis of these proposals and from the experience in developing our system, we summarized a metadata classification method through two features: **description** and **dependence,** which are essentially orthogonal.
In particular metadata are used for: **Content-description**, **Structure-description** or **Location-description**.

- **Content-description Metadata**: they are pieces of information that give an abstract description of component contents, their values are relevant for the application but they are not directly accessible. Example of metadata which have this feature are keyword and index terms for textual documents or for large multimedia objects like those in [BPJ93].
- **Structure-description Metadata**: they are used in order to describe the data structures that model a multimedia document. Examples of metadata which have this feature are the well known concepts of links and anchors in hypermedia systems [GFS94][HSC94].
- **Location-description Metadata**: they describe the logical and physical location features of the components. Well known examples of this kind of metadata are the URL and/or logical path of a file.

The description feature for metadata allows to identify their use, as it has been shown in section 2, and it is useful in order to examine their proper dissemination across the architectural modules of a multimedia system, as it will be shown in the following.

Dependence metadata can be classified as:

- **Media-Type Dependent Metadata** are related to some specific media type. Examples of this kind of metadata are information about an image format [BRA94][AST94], or about audio interval [CHK94]. Also features extracted from images and used as indexes [BPJ93][CHH92] present this kind of property.
- **Domain-Dependent Metadata** refer to concepts which exist, and which have a specific meaning, in the domain of the application. An example of domain-dependent metadata can be the information about coding standards used in the health care information systems. While metadata about the acquisition technique for a NMR-image are both Media-type dependent and Domain-dependent.
- **Content Dependent Metadata** depend only on the content of the original data, that is on the extension of the document collection. Statistics on the updates of a database are examples of this kind of metadata.

Each piece of metadata falls in exactely one of the description cathegories, and at least one of the dependence cathegory. In fact descriptions are mutually exclusive, whereas dependences can refer to various aspects at the same time

4. A reference Architecture with Metadata

The previous classifications allow to reason about metadata dissemination across architectural modules. In this section we show a general reference architecture for distributed multimedia systems. It is a two-level architecture in which at the lower level there are raw-data managed by multimedia databases and at the higher level there are the applications managing and accessing multimedia documents.

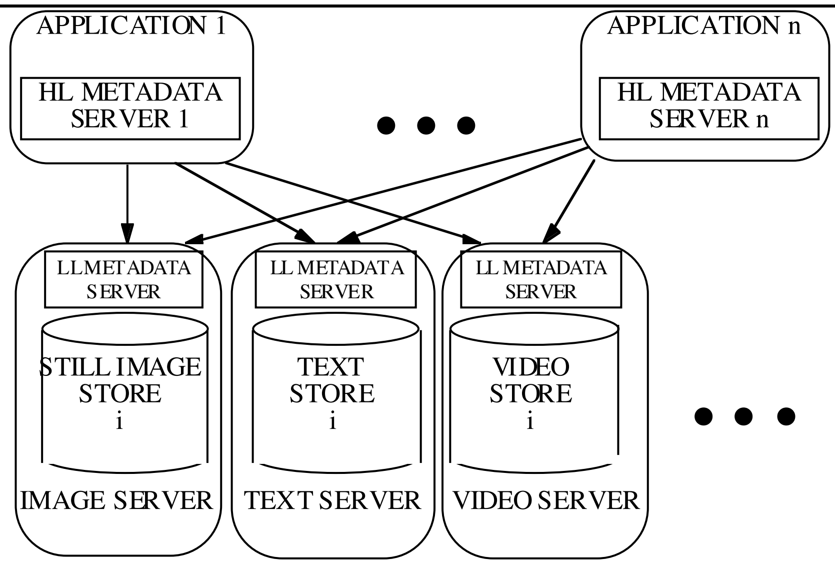

Fig. 2 A reference architecture for multimedia management

Lower-level metadata are those pieces of information that describe raw-data independently from the applications that use them: for instance, considering still images, the format and location of the image or the used compression technique. **Higher-level metadata** describe the different semantics of data from the application point of view. At this level, we find information describing application domains and multimedia document structures. For example, considering the development of a health care information system, radiological images and bio-signals are data that will be subjected to different uses: they will be part of patient folders, they will be inserted in hypermedia educational documents, they will be available for consultation in hospital historical archives. Due to both their huge dimension and legal value, they can't be replicated and they have to be stored in a centralized system. Hence, from the application point of view, there are metadata common to all the single databases which the application interface to, and metadata proper to each single database, which are used by the databases to give the applications a common interface.
These considerations lead to analyze the dissemination of the various kind of metadata on the two levels, in order to define the classes of metadata that are to be taken into account in the design of each module of a system. For example, with reference to the Bohm-Rakow classification, we think that:

- at the higher level, metadata belonging to the 2-3-4-5-6 classes are needed, while
- at the lower level, metadata belonging to the 1-5-6 classes are needed.

The Metadata for Document History class (5) is pertinent to both levels because at the higher level we have information on the history of the whole multimedia complex document managed by each application, while at the lower level we have information about the single components.

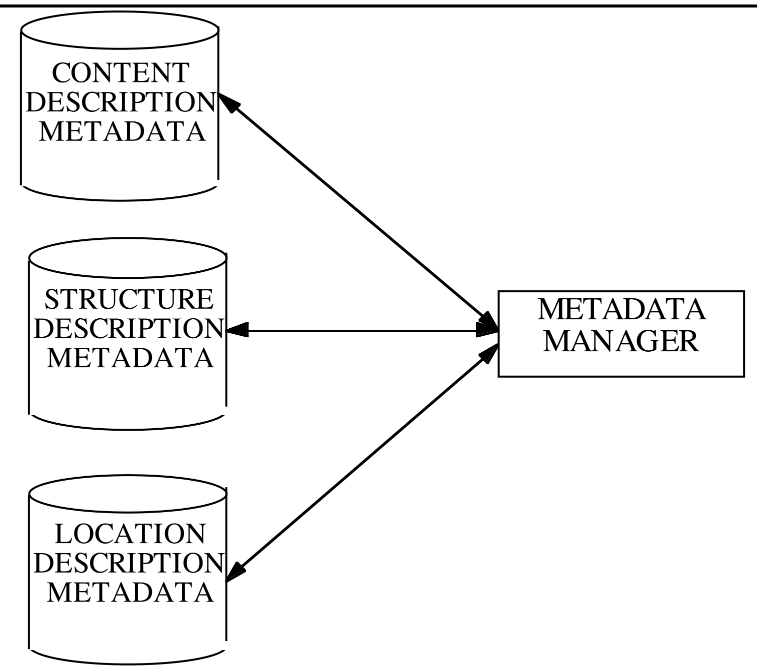

Fig. 3 Metadata Server Architecture

Also in the Metadata for Document Location class (6) are interested both the higher level, for addressing purposes, and the lower, for storing purposes.

We think that our classification introduces something new, with respect to the ones described above, for the following reasons.
The Bohm-Rakow classification is strongly *document-oriented*: it is aimed mainly to structure a minimal set of metadata for multimedia documents. This make it not very useful to analize and define which metadata are to be taken into account in the design of each architectural module. Our classification instead, was conceived as a more general framework to reason about metadata use and design: with the three description features, we can easily figure out which are the required metadata with respect to their uses and aims; with the three dependence features we are facilitated in defining metadata constraints.
The Kashap-Shet classification conversely, even being as general as ours and considering both the description and dependence feature, doesn't catch their essential orthogonality: this may lead to confuse metadata use and constraints during the design activity.

Our metadata classification hence, gives us immediately a simple reference architecture for *metadata servers*, that is those system modules responsible to give to the application all the needed information about data and their interpretations. The reference feature for a metadata server is the description, hence, in fig. 3, we propose an architecture based on the three kinds of description aims a metadatum can concern: Content, Structure and Location.
With respect to the general sytem architecture depicted in fig. 2, the only difference between the metadata servers of the two levels, lies in the fact that at the application level we need information about application domains, hence the Content-description and Structure-description metadata present the Domain-dependence feature, while at the data level we don't need this information.

5. Conclusions and discussion

The use of metadata for the management of a collection of multimedia documents has been discussed in this paper. Metadata are used in order to increase the semantics of multimedia in applications where data access can be performed according to both the hypermedia and querying paradigm. A *metadata-centered* methodology has been presented that has been used in the medical context.

Metadata classifications in literature have been discussed and summarized in a proposal which puts the emphasis on their dependence from media, content and domain, in order to capture their nature, and on their descriptive aims, in order to catch their semantics.

We defined a general architecture for multimedia systems with metadata and we used this architecture as reference framework in the design of a system in the medical context. Here, metadata are being used to allow sharing of medical images among different applications. In particular an application for medical reporting has already been developed and its integration with an existing medical record system is under development. On-line graphical interfaces to interact with the system are also currently under development. A demo will be hopefully available at the URL

`http://www.inf.uniroma3.it/medinfo/braque.html`

before December 1996.

REFERENCES

[AMC95] M. Agosti, M. Melucci, F. Crestani. Automatic Authoring and Construction of Hypermedia for Information Retieval. Multimedia Systems (1995) 3: 15-24.

[AST94] J. T. Anderson, M. Stonebraker SEQUOIA 2000 Metadata Schema for Satellite Images ACM SIGMOD Record 23,4 dec. 1994

[BPJ93] J. Bach, S. Paul, R. Jain. A Visual Information Management System for the Interactive Retrieval of Faces. IEEE Trans. on Knwledge and Data Engineering 5(4), 1993.

[BRA94] K Böhm, T. C. Rakow. Metadata for Multimedia Documents. ACM SIGMOD Record 23,4 dec. 1994

[CHH92] S. K. Chang, T. Y. Hou, A. Hsu. Smart Image Design for Large Image Databases. Journal of Visual Languages and Computing (1992) 3, 323-342.

[CHK94] F. Chen, M. Hearst, J. Kupiec, J. Pedersen, L. Wilcox Metadata for Mixed-media access ACM SIGMOD Record 23,4 dec. 1994

[CKO92] S. Christodoulakis, L. Koveos. Multimedia Information Systems: Issue and Approaches in W. Kim, Modern Database System Addison-Wedsley, NY 1992.

[CMS96] F. Consorti, P. Merialdo, G. Sindoni, J. Di Prospero. A Method to Define and Design Tools for Hypermedia Medical Reports Management. Proceeding of the MIE96 Conference. J. Brender et al. Eds, 1996.

[GFS94] W. I. Grosky, F. Fotouhi, I. K. Sethi, B. Capatina. Using Metadata for Intelligent Browsing of Structured Media Objects. ACM SIGMOD Record 23,4 dec. 1994

[GSW94] U. Glavitsch, P. Schauble, M. Welchsler. Metadata for Integrating Speech Documents in a Text Retrieval System. ACM SIGMOD Record 23,4 dec. 1994

[HSC94] F. Halasz, M. Schwartz. The Dexter Hypertext Reference Model. Communications of the ACM, 37(2): 30-39, 1994.

[JHA94] R. Jain, A. Hampapur. Metadata for Video Databases. ACM SIGMOD Record 23,4 dec.1994

[KSH94] W. Klaus, A. Sheth. Metadata for Digital Media. ACM SIGMOD Record 23,4 dec. 1994. http://www.cs.uga.edu/LSDIS/pub.html

[KSH95] V. Kashyap, A. Sheth. Controlled Vocabulary Sharing for Query Processing in Global Information Systems. Technical Report, LSDIS Lab, University of Georgia, Feb. 1995.

[KSS95] V. Kashyap, K. Shah, A. Sheth. Metadata for Building The Multimedia Patch Quilt. (to appear in) “Multimedia Database System: Issue and V. S. Subrahamaniam Eds., Springer-Verlag, 1995.

[SGG96] T. R. Smith, S. Geffner, J. Gottsegen. A General Framework for the Meta-Information and Catalogs in Digital Libraries Proc. Of 1st IEEE Conf. on Metadata 1996. Available at http://www.nml.org/resources/misc/metadata/proceedings

[SHE96] A. Sheth Data Semantics: What, Where and How?. Database Application Semantics Proc. of 6th IFIP W.C. on Data Semantics, R. Meersman and L. Marks Eds. 1996

[SSK95] L. Shklar, A. Sheth, V. Kashyap, K. Shah. InfoHarness: Use of Automatically Generated Metadata For Search and Retreival of Heterogeneous Information. Proceedings of CAiSE 95, june 1995.

[WOO85] J. Wood. What's a Link?. Readings in Knowledge Representation, Morgan Kaufman 1985.