Metadata for Multimedia Documents

Klemens Böhm and Thomas C. Rakow
GMD Integrated Publication and Information Systems Institute (IPSI)
Dolivostraße 15
64293 Darmstadt
Germany
{kboehm, rakow}@darmstadt.gmd.de

Abstract

In this article metadata for mulimedia documents are classified in conformity with their nature, and the different kinds of metadata are brought into relation with the different purposes intended. We describe how metadata may be organized in accordance with the ISO standards SGML, which facilitates the handling of structured documents, and DFR, which supports the storage of collections of documents. Finally, we outline the impact of our observations on future developments.

1 Introduction

As concepts and technology for single digital media types are maturing, problems concerning the composition of multimedia documents from single-media components are becoming more acute. The reflections making up this article are based on the perspective that the type of multimedia documents itself is again a digital media type.[1] To be able to better distinguish it from those other media types we will refer to those media types as *single media types* when in doubt. Furthermore, with *multimedia items* we refer to both components of multimedia documents as well as multimedia documents in their entirety. In this article the different metadata that are necessary or advantageous for handling multimedia items are identified and classified. From one perspective, metadata for multimedia documents describe single media types, intra-document relationships, i.e. relationships between components, and inter-document relationship, i.e. between different sets of documents.

Clearly, some sorts of metadata are also useful fo other application domains. In this article, however, we do, to some degree, rely on observations that have been made with two applications developed at our institute. The Calendar of Events (CoE, [19, 16]) is an information system to administer multimedia descriptions of, say, cultural events: The user formulates a query according to his information needs. When composing the answer, both technical characteristics of the client-side (e.g. 'Only Quicktime movies can be presented.') as well as preferences of the person posing the query (e.g. 'Only presentations up to one minute should be returned ') can be taken into account. The MultiMedia Forum (MMF, [18]) is another application that processes multimedia documents. It is an interactive online journal. Here, the reader or consumer himself selects news items that seem to be of interest to him from a table of contents. An important difference between the MMF and the World-Wide Web (WWW) are the facts that the individual issues of the MMF are composed by an editorial team, and each piece of news corresponds to exactly one issue.

With this article we intend to make explicit the different kinds of metadata for multimedia items and their usage. Section 2 contains that classification. In Section 3 we go into the facets of the ISO standards SGML and DFR that are important in this context. In Section 4 the usage of the categories of metadata introduced is elaborated. Section 5 deals with some implementation issues and further developments.

2 Classifying Metadata for Multimedia Documents

Clearly, the different ways of using multimedia documents impinge on the metadata and the way they are organized. In the sequel we describe the different kinds of metadata for multimedia documents we

[1] The approach that a type text reflects the intensional characteristics of documents can be encountered in [2]: Relational database technology has been extended so that a type `text` for column entries is available. There, however, 'text' means 'SGML document'. In other words, it seems to us that the type name `SGML_Document` would better reflect the speciality of the approach.

have been able to identify. First of all, the differentiation between metadata themselves on the one hand and on the other hand information how metadata for a concrete collection of multimedia documents are organized, which is again metainformation, is relevant. For instance, a multimedia document may be furnished with an attribute levelOfExpertize that reflects how familiar the reader or consumer should be with the field the document relates to. In order to deal with that information in a reasonable way the range of the attribute must be known. - This differentiation is, to some degree, orthogonal to the following classification of metadata for multimedia documents:

- **Metadata for the Representation of Media Types.** As opposed to, say, plain ASCII-text certain additional information is mandatory in connection with the presentation of multimedia data. This includes format, coding and the compression techniques that have been applied. Either a current name for format, coding etc. may be given, or it may be explicitly described. For example, a datatype AUDIO is described by the number of samples per second, number of channels, and the coding in which it has been recorded, e.g. ulaw or linear. From our point of view, certain attributes of textual document components not reflecting the content also fall into this category. An example is an attribute language bearing the language a textual component is written in.
- **Content-Descriptive Metadata.** These metadata are determined intellectually or by means of semi-automatic or automatic methods. In the last two cases, these methods are media-type-specific. An example from our institute is the DREAM parser [5] that enriches documents with SGML-like markup. Examples of content-descriptive metadata are a list of persons or institutions having some relation to a particular multimedia document's content. Metadata of this type for CoE data are the location of the event and the price of the tickets, to provide further examples.
- **Metadata for Content Classification.** The distinction between content-descriptive and technical metadata for simple digital media types is well-known (cf. [14]). Furthermore, however, while content-descriptive metadata reflect a document's or a document component's content, metadata for content classification are additional information that can be derived from the document content. For instance, metainformation such as the level of expertise in the field required by the reader falls into this category. To our knowledge, coming up with algorithms how that kind of metainformation can be derived from the document content automatically or at least semi-automatically is an open research issue.
- **Metadata for Document Composition.** Logical components of multimedia documents have, as a rule, certain semantics or, in other words, a role as part of the document. A document component's role does not unmistakably follow from the document. Rather, it is dependent on the kinds of roles that are allowed within documents of a certain type. From a slightly different perspective, a document component's role depends on the author's intentions. In the cooperative authoring environment SEPIA [17], to give an example, there is a differentiation between claims and facts. Frequently, it is subject to interpretation whether a statement is a claim or a fact, to continue the example. Hence, the knowledge about document components' roles is metainformation. The relationships between document components have certain characteristics. Composition-specific metadata are knowledge about these characteristics. As opposed to some other kinds of metadata, this one makes sense only for multimedia documents and not for simple media types. For instance, the direct content elements of a document component may be ordered, as with SGML documents [8, 7], or not, as with diverse hypertext models. Using data modeling terminology, the different semantics of these instances of the partOf-relationship [6] is reflected in the operations that would be provided by the corresponding modeling primitives. In the first case a method getNextComposite returning the composite that follows the target object makes sense. This method, however, would not have a counterpart in case of unordered composites.
- **Metadata for Document History.** With the sorts of metadata that have been described so far multimedia documents are described only as static entities that do not seem to change over time. However, in multimedia publishing environments it is a common practice to record the status of multimedia items. E.g. there may be a property status with possible values approvedByChiefEditor, notApproved etc. Metadata for document history need not necessarily apply to the entire document, but instead may be different for individual document components. Further examples for metadata of that kind are the author's or composer's name for a multimedia document component or the calendar date of such a component's last update. Additionally, the

derivedFrom-relationship in versioning models, i.e. the information from which (older) document components a more recent one has been derived, falls into this category. Sometimes metadata of this kind can be recorded automatically, e.g. the date of an audio component's last update. However, when considering the attribute status described above it can easily be seen that this need not necessarily be the case. Finally, from a consumer perspective the information whether a document component has already been visited also falls into that category.

- **Metadata for Document Location.** Multimedia documents are not inclined to be duplicated and distributed, as compared to conventional ones. Rather they are accessed by the consumer on demand. This facilitates continuous modification of documents' content without that the document actually becomes another one. It is a prerequisite, however, that the multimedia documents can always be localized unambiguously [10].

The kinds of metadata that have been mentioned so far relate to individual multimedia documents. Additionally, there are metadata for collections of multimedia documents. Here, we give an example of such metadata which will be referred to as *statistical metadata.* Consider the case that multimedia documents are stored within a database. Metainformation that may be relevant in this context is the frequency of documents or document components with certain characteristics. Metadata of that kind may be based on other metainformation. E.g. the number of documents written by a certain author in a collection of documents is based on the corresponding piece of metainformation for individual multimedia documents.

3 Organizing Metadata for Multimedia Documents

Due to space limitations, with our reflections about the organization of metadata in this section we limit ourselves to the possibilities according to two international standards for document handling.

3.1 SGML - a Means to Deal with Structured Documents

In this context, a prominent role falls to the share of the SGML standard. SGML documents conform to so-called *document-type definitions (DTDs)*, which are essentially attributed grammars. Arbitrary DTDs can be defined to specify the generic logical structure of documents of a certain type, such as, say, newspaper articles or scientific publications. Document components such as titles, paragraphs, footnotes and the like are referred to as *SGML elements.* A document's elements form an ordered tree, the *parse tree.* The tree's edges stand for the partOf-relationship. In this context, the fact that the parse tree's leaves, i.e. the atomic document components, need not only be of type text, but rather may be of arbitrary types, is important. This may cause problems when processing and presenting documents. A straightforward solution for a distinct set of DTDs is to leave the interpretation and presentation to the application. More sophisticated solutions are beyond the scope of this article [4].

Metadata in SGML documents may be organized as follows:

- The DTD may contain certain element types for the metadata. An obvious example is an element type keywordList in a DTD scientificPublication. If the metadata are organized in that way, in essence they cannot be distinguished from the original media data. In that case, an application processing multimedia documents must know the semantics of the individual element types in case metadata are processed differently from original data. From our point of view, however, the distinction between metadata and original data is not clean-cut: An abstract in a document of type scientificPublication is, in a way, metainformation as it is a summary of the content of the (remainder of the) document. On the other hand, it is an original part of the document because the abstract's content may be better suited to illustrate the authors' intentions than the document in its entirety.
- With SGML it is possible to freely define attributes for elements of arbitrary types in the DTD. Possible attribute types are enumeration types, which, in turn, are freely definable, numbers and strings. As attributes can be defined for arbitrary element types, metainformation need not apply to the whole document. Instead there may be a differentiation between individual document components.
- The mapping from document components to element types is metainformation on the logical document structure.
- The document-type definition itself is part of the metadata. It may either be contained in the document's prologue, or the prologue contains a reference to the corresponding DTD.

3.2 DFR - a Framework to Arrange Document Collections

While in this context metadata are largely contained in the document, in other cases it is advantageous to factor out the metadata from the individual multimedia documents. As opposed to SGML, a standard for the logical document structure, the *DFR-standard* ('Document Filing and Retrieval', [9]) deals with the organization of a collection of documents. With DFR there is a set of containers, the *DFR groups*. A DFR group may itself be contained in another one. Hence, the primary structure is hierarchical. The documents themselves are part of a DFR group. The components of a DFR modelling are referred to as *DFR objects*. Names and attributes, i.e. both attribute name and range, of DFR objects can be freely defined. To give an example, consider the CoE which is a DFR application. The location of the event described in the document is the metainformation on which the DFR modeling is based. The DFR groups correspond to different geographic or territorial units such as states or counties. The hierarchical structure is the containedIn-relationship between them. The leaves of the hierarchy are the documents being event descriptions.

4 Exploiting Metadata for Multimedia Documents

While certain ways to exploit metadata are already known from data of conventional types, the nature of multimedia documents leads to a variety of new aspects.

- **Closer Matching Presentation Requirements.** A special problem with multimedia documents is that due to the different technical settings of the client-sides document-processing mechanisms should be more flexible than for conventional document types. On the one hand experience with the MMF shows that in presentation environments it is advantageous to hide away document components from the consumer that cannot be consumed because of the settings of his environment. From another perspective, in a distributed environment it is inefficient to transport data that cannot be presented to the client-side. Metadata is used to deal with both aspects of the problem. The consequence is that metadata are actually used to control the dynamic composition of multimedia documents or, in other words, control viewing of multimedia documents.

- **Browsing.** As it seems that there are no satisfactory means to quickly overview video or audio data, metadata may be used to ease browsing through multimedia documents. A typical example are icons in WWW-pages standing for a big image or a video sequence. Usually the icon is a reduction of the image or of a relevant frame. Besides that, the example shows that metadata need not necessarily be of a conventional type.

- **Posing Descriptive Queries.** Querying is an issue that seems to be fairly independent of the type of metadata. In other words, it is only natural to pose queries to a pool of multimedia documents on the metainformation on the document's history, on content-based document categorization and the like. Query processing on multimedia documents can be optimized using statistical metadata and metadata for the logical document structure. A relevant reference is [12]. Likewise, representation-specific characteristics can be considered in queries on multimedia documents. This aspect of querying is of minor importance with regard to purely textual documents. For example, as it is reasonable to base billing in telecommunication environments on the amount of data that has been transferred size of document components will become a relevant criterium for selection. From a slightly different perspective, the duration of the presentation of multimedia documents shall be another criterium.

- **Posing Queries on Document Content and Repositories.** The possibilities of content-based retrieval from multimedia documents are relatively limited, as opposed to retrieval from merely textual documents. This can be inferred from the fact that, quite frequently, adjacent text is exploited to identify the content of images in multimedia documents. If a restriction is made to a set of queries that are allowed the selection of relevant multimedia documents can in principle be facilitated by content-dependent metadata. In the MMF there is the possibility to select documents dealing with certain persons, because the necessary metainformation is part of the documents' header. In the CoE's DFR archive the multimedia documents being event descriptions are classified by means of the event's location. It is part of the DFR archive's functionality to be able to select documents being directly or indirectly contained in a DFR group. Hence, in both cases metadata are used to enhance querying on documents' content. However, there also is a difference: A classification of metadata on multimedia documents based on their usage is the discrimination between *document-specific* and *repository-specific metadata*. This classification is orthogonal to the one made in Section 2, and it does not primarily depend on the nature of the metadata.

- **Cooperative Work on Multimedia Documents; Security.** Metadata for documents' history as well as metadata on the logical document structure ease the authoring and publishing process. This kind of additional information is particularly helpful if multimedia documents are composed and edited by more than one person. In, say, a complex reference work there may well be two dozens of reasons why a word may be italicized. Now assume that a reviewer is reading a section written by an author, and one of his assignment is to ensure the document's consistency. The task is easier if the logical document structure is at hand: The reviewer need not ponder about the reason why exactly a word has been italicized. On the other hand, without the discrimination between logical and layout structure information of that kind would not be available. Summing up, a document in which the components' roles are identified bears more information that might be relevant for authoring and composition. Also in case there is only one person involved in the publishing process it may be advantageous to chronicle the document design process for analogous reasons, namely if it contains various iterations. Besides that, metadata on document history are mandatory to ensure authorized access to documents and document components. In a sample scenario, consumers that are not part of the editorial team shall only be allowed to access documents with a certain status.

- **Easier Modification of Document Layout and Rendition.** By introducing metadata on the logical document structure document components are explicitly identified. Thus, it can be discerned between a document's logical structure and its layout structure. Another impact of this discrimination is that the document itself need not be modified when altering the document's layout. If, say, the layout of facts that used to be displayed using bold letters shall be changed to italics the rule 'Display facts using bold letters.' must be replaced by the one 'Italicize facts.'. Hence, modifications of that kind only impinge on the meta level.

5 Implementation Issues and Conclusions

The objective of our work is the storage of multimedia documents in databases. This is motivated by the innovative publication environments envisioned at our institute where the editing staff works cooperatively and in concurrence with the comsumers accessing the document pools. To this end, the object-oriented DBMS VODAK [11] that has been developed at our institute is extended to become a multimedia DBMS. One facet of the approach is to provide built-in datatypes for continuous media types, e.g. audio [13]. In the context of these built-in datatypes it is important to mention that some metadata are encapsulated and need not be administered by the application developer. The advantage is that document components may be edited, and their internal representation may be altered without affecting the layers above. The category of metadata that need be considered in this context are media-type specific metadata. At a later stage, there shall be a type being supertype of those continuous media types. Namely, some of the processing of data of continuous media types is canonical. It is obvious that in this case updating the canonical metadata (e.g. component size in bytes) must be part of this generic type's operationality. On top of VODAK we have realized an application framework for SGML document storage [3]. The main objective of this work has been to handle SGML documents of arbitrary types. Using metadata terminology, as little metainformation as possible has been hardcoded into the database schema. Instead our goal has been to store these data as database content (cf. [1]). Furthermore, a DFR archive based on VODAK is available [16]. After having established the connection between this DFR archive and the SGML database application the question which metadata shall be used to structure the DFR repository will become relevant.

References

[1] K. Aberer, K. Böhm, and C. Hüser, "The Prospects of Publishing Using Advanced Database Concepts", in *Proceedings of Conference on Electronic Publishing*, April 1994, eds., C. Hüser, W. Möhr, and V. Quint, 469-480, John Wiley & Sons, Ltd., 1994.

[2] G.E. Blake et al., "Text/Relational Database Management Systems: Harmonizing SQL and SGML", in *Proceedings of First International Conference on Applications of Databases*, June 1994, eds., W. Litwin, and T. Risch, 267-280, Lecture Notes in Computer Science 819, Springer Verlag Berlin Heidelberg, 1994.

[3] K. Böhm, K. Aberer, and E. Neuhold, "Administering Structured Documents in Digital Libraries", submitted for publication.

[4] K. Böhm, and K. Aberer, "An Object-Oriented Database Application for HyTime Document Storage", accepted for publication in *Proceedings of Conference on Information and Knowledge Management 1994 (CIKM94)*, Gaithersburg, MD, December 1994.

[5] T. Göttke, and P. Fankhauser, "Dream 2.0 User Manual", *Arbeitspapiere der GMD 660*, Sankt Augustin, 1992.

[6] M. Halper et al., "Integrating a Part Relationship into an Open OODB System Using Metaclasses", accepted for publication in *Proceedings of Conference on Information and Knowledge Management 1994 (CIKM94)*, Gaithersburg, MD, December 1994.

[7] E. van Herwijnen, *Practical SGML* (second edition), *Kluwer Academic Publishers*, 1994.

[8] *Information Processing - Text and Office Systems - Standardized Generalized Markup Language (SGML)*, ISO 8879-1986 (E), International Organization for Standardization, 1986.

[9] ISO/IEC 10166-1, *Information Technology - Text and Office Systems - Document Filing and Retrieval (DFR)*, part 1 and part 2, first edition, December 1991.

[10] ISO/IEC, 10031-2 *Information Technology - Text and Office Systems - Distributed Office Applications Model (DOAM)*, part 2: Distinguished-object-reference and associated procedures (DOR), International Standard 10031, 1991.

[11] W. Klas, K. Aberer, and E. Neuhold, "Object-Oriented Modeling for Hypermedia Systems Using the VODAK Modeling Language (VML)", in: A. Dogac, T.Ozsu, A.Biliris, T.Sellis (Eds.), *Advances in Object-Oriented Database Systems*, NATO ASI Series, Springer Verlag Berlin Heidelberg, 1994.

[12] P. Kilpeläinen, and H. Mannila, "Grammatical Tree Matching", in *Proceedings of the Third Annual Symposium on Combinatorial Pattern Matching, 1992*, eds., A. Apostolico et al., 162-174, Springer Verlag Berlin Heidelberg, 1992.

[13] M. Löhr, and T.C. Rakow, "Supporting Audio in a Distributed Object-Oriented Database System", submitted for publication.

[14] K. Meyer-Wegener, *Multimedia-Databases* (in German), *B.G. Teubner Stuttgart*, 1991.

[15] P. Missier, M. Rusinkiewicz, and A. Silberschatz, "Providing Multidatabase Access - an Association Approach", *Technical Report from University of Texas at Austin 93-26*, 1993, available via anonymous ftp, ftp.cs.utexas.edu/pub/techreports/tr93-26.

[16] T.C. Rakow et al., "Development of a Multimedia Archiving Teleservice using the DFR Standard", in *Proceedings of 2nd International Workshop on Advanced Teleservices and High Speed Communication Architectures*, LNCS, Springer, Berlin 1994.

[17] N. Streitz et al., "SEPIA: a Cooperative Hypermedia Authoring Environment", in *Proceedings of the ACM Conference on Hypertext*, Milano, Italy, 1992, ACM Press, 11-22.

[18] K. Süllow et al., "MultiMedia Forum: an Interactive Online Journal", in *Proceedings of Conference on Electronic Publishing*, April 1994, eds., C. Hüser, W. Möhr, and V. Quint, 413-422, John Wiley & Sons, Ltd., 1994.

[19] H. Thimm, and T.C. Rakow, "A DBMS-Based Multimedia Archiving Teleservice Incorporating Mail", in *Proceedings of First International Conference on Applications of Databases, June 1994*, eds., W.Litwin and T. Risch, 281-298, Lecture Notes in Computer Science 819, Springer Verlag Berlin Heidelberg, 1994.