Chapter 4.5
Semantically Driven Multimedia Querying and Presentation

Isabel F. Cruz
University of Illinois, Chicago, USA

Olga Sayenko
University of Illinois, Chicago, USA

ABSTRACT

Semantics can play an important role in multimedia content retrieval and presentation. Although a complete semantic description of a multimedia object may be difficult to generate, we show that even a limited description can be explored so as to provide significant added functionality in the retrieval and presentation of multimedia. In this chapter we describe the Delaunay^{View} that supports distributed and heterogeneous multimedia sources and proposes a flexible semantically driven approach to the selection and display of multimedia content.

INTRODUCTION

The goal of a semantically driven multimedia retrieval and presentation system is to explore the semantics of the data so as to provide the user with a rich selection criteria and an expressive set of relationships among the data, which will enable the meaningful extraction and display of the multimedia objects. The major obstacle in developing such a system is the lack of an accurate and simple way of extracting the semantic content that is encapsulated in multimedia objects and in their inter-relationships. However, metadata that reflect multimedia semantics may be associated with multimedia content. While metadata may not be equivalent to an ideal semantic description, we explore and demonstrate its possibilities in our proposed framework. Delaunay^{View} is envisioned as a system that allows users to retrieve multimedia content and interactively specify its presentation using a semantically driven approach.

Delaunay^{View} incorporates several ideas from the earlier systems Delaunay (Cruz & Leveille, 2000) and Delaunay^{MM} (Cruz & James, 1999). In the Delaunay^{View} framework, multimedia content is stored in autonomous and heterogeneous sources

annotated with metadata descriptions in resource description framework (RDF) format (Klyne & Carroll, 2004). One such source could be a database storing scientific aerial photographs and descriptions of where and when the photographs were taken. The framework provides tools for specifying connections between multimedia items that allow users to create an integrated virtual multimedia source that can be queried using RQL (Karvounarakis et al., 2002) and keyword searches. For example, one could specify how a location attribute from the aerial photo database maps to another location attribute of an infrared satellite image database so that a user can retrieve images of the same location from both databases.

In Delaunay^{View}, customizable multimedia presentation is enabled by a set of graphical interfaces that allow users to bind the retrieved content to presentation templates (such as *slide sorters* or *bipartite graphs*), to specify content layout on the screen, and to describe how the dynamic visual interaction among multimedia objects can reflect the semantic relationships among them. For example, a user can specify that aerial photos will be displayed in a slide sorter on the left of the workspace, satellite images in another slide sorter on the bottom of the workspace, and that when a user selects a satellite image, the aerial photos will be reordered so that the photos related to the selected image appear first in the sorter.

In this paper we describe our approach to multimedia querying and presentation and focus on how multimedia semantics can be used in these activities. In "Background" we discuss work in multimedia presentation, retrieval, and description; we also introduce concepts relating to metadata modeling and storage. In "A Pragmatic Approach to Multimedia Presentation", we present a case study that illustrates the use of our system and describe the system architecture. In "Future Work" we describe future research directions and summarize our findings in "Conclusions."

BACKGROUND

A multimedia presentation system relies on a number of technologies for describing, retrieving and presenting multimedia content. XML (Bray et al., 2000) is a widely accepted standard for interoperable information exchange. MPEG-7 (Martinez, 2003; Chang et al., 2001) makes use of XML to create rich and flexible descriptions of multimedia content. Delaunay^{View} relies on multimedia content descriptions for the retrieval and presentation of content, but it uses RDF (Klyne & Carroll, 2004) rather than XML. We chose RDF over XML because of its richer modeling capabilities, whereas in other components of the Delaunay^{View} system we have used XML (Cruz & Huang, 2004).

XML specifies a way to create structured documents that can be easily exchanged over the Web. An XML document contains *elements* that encapsulate data. *Attributes* may be used to describe certain properties of the elements. Elements participate in hierarchical relationships that determine the document structure. XML Schema (Fallside, 2001) provides tools for defining elements, attributes, and document structure. One can define typed elements that act as building blocks for a particular schema. XML Schema also supports inheritance, namespaces, and uniqueness.

MPEG-7 (Martínez, 2003) defines a set of tools for creating rich descriptions of multimedia content. These tools include *Descriptors*, *Description Schemes (DS)* (Salembier & Smith, 2001) and the *Description Definition Language (DDL)* (Hunter, 2001). MPEG-7 descriptions can be expressed in XML or in binary format. Descriptors represent low-level features such as texture and color that can be extracted automatically. Description Schemes are composed of multiple Descriptors and Description Schemes to create more complex descriptions of the content. For example, the MediaLocator DS describes the location of a multimedia item. The MediaLocator