DOCUMENT RESUME

ED 459 835 IR 058 371

AUTHOR	Christel, Michael G.; Maher, Bryan; Begun, Andrew
TITLE	XSLT for Tailored Access to a Digital Video Library.
SPONS AGENCY	National Science Foundation, Arlington, VA.
PUB DATE	2001-06-00
NOTE	11p.; In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (1st, Roanoke, Virginia, June 24-28, 2001). For entire proceedings, see IR 058 348. Figures may not reproduce well.
CONTRACT	DUE-0085834; IRI-9817496
AVAILABLE FROM	Association for Computing Machinery, 1515 Broadway, New York NY 10036. Tel: 800-342-6626 (Toll Free); Tel: 212-626-0500; e-mail: acmhelp@acm.org. For full text: http://www1.acm.org/pubs/contents/proceedings/dl/379437/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Libraries; Access to Information; Computer Interfaces; Computer System Design; *Electronic Libraries; Higher Education; Library Collection Development; Multimedia Materials; *Videotape Recordings
IDENTIFIERS	Carnegie Mellon University PA

ABSTRACT

The Informedia Project at Carnegie Mellon University has created a multi-terabyte digital video library consisting of thousands of hours of video, segmented into over 50,000 stories, or documents. Since Informedia's inception in 1994, numerous interfaces have been developed and tested for accessing this library, including work on multimedia abstractions, or surrogates, which represent a video document in an abbreviated manner. The utility and efficiency of these surrogates have been reported in detail elsewhere, validated through a number of usability methods, including transaction log analysis, formal empirical studies, contextual inquiry, heuristic evaluation, and cognitive walkthroughs. This paper begins with an introduction to a few of these interfaces and their implementation history. The promise of Web technologies is then discussed, particularly the recommendations of the World Wide Web Consortium (W3C), leading to a presentation of the Informedia digital video library delivered through a Web browser via XML and XSLT. Emphasis is placed on the tailored accessibility offered by this information architecture, with specific examples given as evidence. The paper concludes with a discussion of next steps planned for the Informedia library work. (Contains 19 references.) (AEF)

XSLT for Tailored Access to a Digital Video Library

Michael G. Christel
CS Dept. and HCI Institute
Carnegie Mellon University
Pittsburgh, PA 15213
1-412-268-7799
christel@cs.cmu.edu

Bryan Maher
Computer Science Dept.
Carnegie Mellon University
Pittsburgh, PA 15213
1-412-268-8970
bsm@cs.cmu.edu

Andrew Begun
Computer Science Dept.
Carnegie Mellon University
Pittsburgh, PA 15213
1-412-268-2029
apb@andrew.cmu.edu

ABSTRACT

Surrogates, summaries, and visualizations have been developed and evaluated for accessing a digital video library containing thousands of documents and terabytes of data. These interfaces, formerly implemented within a monolithic stand-alone application, are being migrated to XML and XSLT for delivery through web browsers. The merits of these interfaces are presented, along with a discussion of the benefits in using W3C recommendations such as XML and XSLT for delivering tailored access to video over the web.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *video*. H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – *standards, dissemination, user issues*.

General Terms

Design, Human Factors, Standardization.

Keywords

Digital video library, XML, XSLT, surrogate.

1. INFORMEDIA INTERFACES

The Informedia Project at Carnegie Mellon University has created a multi-terabyte digital video library consisting of thousands of hours of video, segmented into over 50,000 stories, or documents. Since Informedia's inception in 1994, numerous interfaces have been developed and tested for accessing this library, including work on multimedia abstractions or *surrogates* which represent a video document in an abbreviated manner [4, 5]. The utility and efficiency of these surrogates have been reported in detail elsewhere [1, 2, 3, 14], validated through a number of usability methods, including transaction log analysis, formal empirical studies, contextual inquiry, heuristic evaluation, and cognitive walkthroughs. This paper begins with an introduction to a few of these interfaces and their implementation history. The promise of web technologies is then discussed, particularly the recommendations of the World Wide Web Consortium (W3C), leading to a presentation of the Informedia digital video library delivered through a web browser via XML and XSLT. Emphasis is placed on the tailored accessibility offered by this information architecture, with specific examples given as evidence. The paper concludes with a discussion of next steps planned for the Informedia library work.

JCDL '01, June 24-28, 2001, Roanoke, Virginia, USA.

1.1 Informedia Surrogates

Video is an expensive medium to transfer and view. MPEG-1 video, the compressed video format used in the Informedia library, consumes 1.2 Megabits per second, and looking through a ten minute video for a section of interest could take a viewer ten minutes of time. Surrogates can help users focus on precisely which video documents are worth further investigation, reducing viewing and video data transfer time. Example Informedia surrogates include brief titles and single thumbnail image overviews, as shown in Figure 1 for 12 documents.

The Figure 1 interface shows query-based thumbnail images: the image is selected from the neighborhood of the document where the highest match scores occurred. In this example, the first few documents show weather maps, indicating that most of the matching to the query "cold snow ice avalanche" occurred in portions of the documents where weather maps were shown. By contrast, the ninth document shows a snowplow, indicating footage of snow and a plow where the query terms are discussed most frequently in the story. Past work showed the utility of choosing thumbnails based on context rather than simply choosing the first visual for a document, and for packing the result set with thumbnails rather than solely listing text titles, document durations and broadcast dates [1].

The vertical bar to the left of each thumbnail indicates relevance to the query, with color-coding used to distinguish contributions of each of the query terms. The document surrogate under the mouse cursor, the eighth result, has its title text displayed in a pop-up window, and the query word display is also adjusted to reflect this particular document. The document is part of the results set primarily because it mentions "avalanche" frequently with some mention of "snow." In Figure 1, "cold" and "ice" are grayed out to show they don't apply to the currently focused document, and the vertical relevance bar for the document shows only two colors: a small patch for "snow" and a large extent for "avalanche." Hence, the display of Figure 1 makes use of relevance bars, query word color-coding, context-specific thumbnail selection, and additional pop-up text information to present a page of documents to the user.

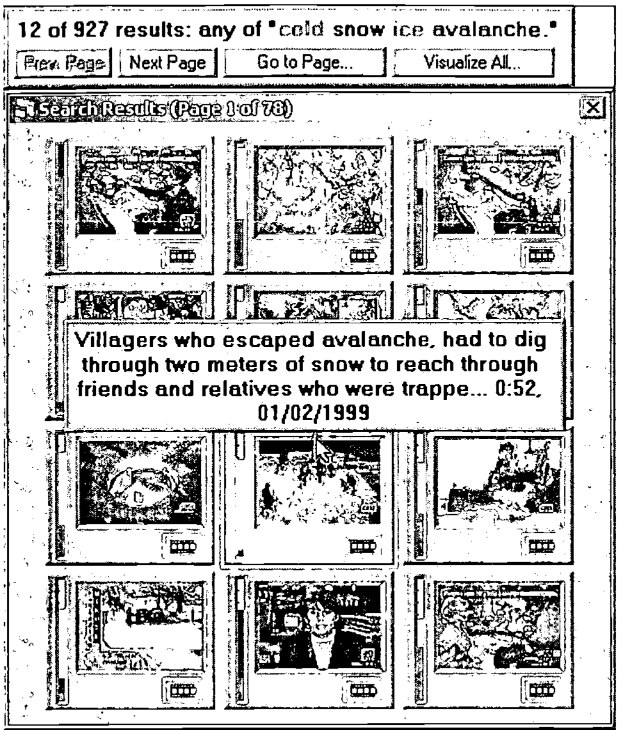

Figure 1. Thumbnail results page for 12 documents, with one pop-up title shown.

From Figure 1's interface, clicking on the filmstrip icon for a document displays a storyboard surrogate with the visual flow of that document, along with locations of matches to a query, as shown in Figure 2.

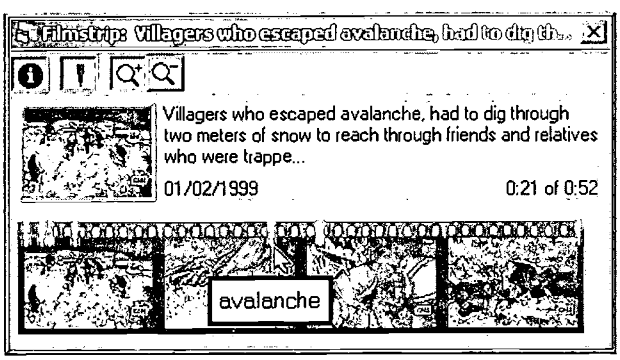

Figure 2. Storyboard, showing that "avalanche" is discussed 21 seconds into the 52-second video document.

Such an interface is equivalent to drilling into a document to expose more of its details before deciding whether it should be viewed. Storyboards are also navigation aids, allowing the user to click on an image to seek to and play the video document from that point forward. For example, Figure 3 shows the video playback window for this document, complete with synchronized transcript, started at this point by clicking on the Figure 2 storyboard's second image. These surrogates are built from metadata automatically extracted by Informedia speech, image, and language processing modules, including transcript text, shot boundaries, key frames for shots, and synchronization information associating the data to points within the video [14].

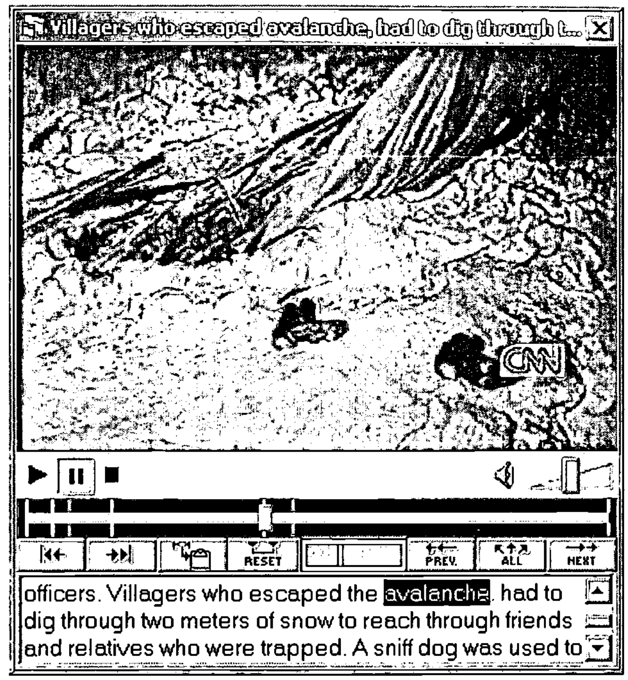

Figure 3. Video playback window, complete with match lines and scrolling transcript.

Figures 1, 2, and 3 show the typical interaction progression of users during the first years of the library. A text search was entered, results were returned as in Figure 1, titles and thumbnails were browsed, with optionally more detailed surrogates as that of Figure 2 examined, leading to some videos being played with the interface of Figure 3. Many fewer videos were actually played compared to the total number returned by text searches.

While the surrogates were put to use, they were not sufficient to deal with the richness of a growing library. As the Informedia collection grew from tens to thousands of hours, the results set from queries grew from tens to hundreds or thousands of documents. Whereas a query on "cold snow ice avalanche" might have produced 30 results that could all be shown on a single screen, later queries against years of CNN news produced too many documents to afford a direct examination of each thumbnail. Figure 1 shows the results of a query against 1998 and 1999 news, producing 927 results. Visualization techniques were added to provide overviews of the full result set and to enable user-directed inquiries into spaces of interest within this result set.

1.2 Informedia Visualization Techniques

The three main visualization techniques employed in the Informedia library interface are:

- Visualization by Example (VIBE), developed to emphasize relationships of result documents to query words [12].
- Timelines, emphasizing document attributes to broadcast date [4].
- Maps, emphasizing geographic distribution of the events covered in video documents [5].

Each technique is supplemented with dynamic query sliders, allowing ranges to be selected for attributes such as document size, date, query relevance, and geographic reference count. The visualizations shown here convey semantics through positioning, but could be enriched to overlay other information dimensions through size, shape, and color, as detailed elsewhere [4, 5].

By combining multiple techniques, users can refine large document sets into smaller ones and better understand the result space. For example, the 927 documents of the query in Figure 1 produce the VIBE plot shown in Figure 4. By dragging a rectangle bounding only the points between words, and excluding the points at just a single query word, the user can reduce the result set to just those documents matching two or more of the terms "cold snow ice avalanche." This operation is shown in Figure 4, reducing the focused result set from 927 documents to 281.

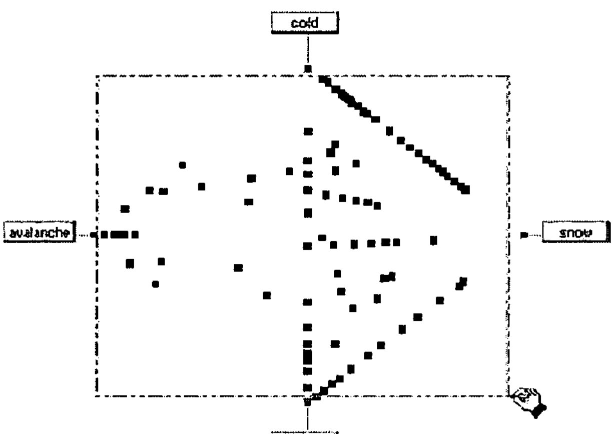

Figure 4. Selecting area of VIBE plot mapping to "two or more of the terms 'cold snow ice avalanche'".

VIBE allows users unfamiliar or uncomfortable with Boolean logic to be able to manipulate results based on their query word associations. For video documents such as a news corpus, there are other attributes of interest besides keywords, such as time and geography. Figure 5 shows a timeline that portrays the obvious (considering that the news corpus originates in the Northern Hemisphere): results from the "cold snow ice avalanche" query cluster in the winter months of November to March.

Figure 6 shows a snapshot of a sequence of interactions that trim down the 281 documents from Figure 4's interaction to a very manageable set of 11. A map view of the results shows a number of highlighted countries, some mentioned only once peripherally in news stories discussing two or more of "cold snow ice avalanche." By highlighting only countries mentioned 4 or more times, tangential references are given less consideration. The user can drag a time window, through the date slider shown below, to set a time period for which to plot results. The user can also manipulate the map, zooming into Europe as a region of focus. In this manner, the user discovers that when looking at February 1999 the documents are concentrated in Austria and Switzerland.

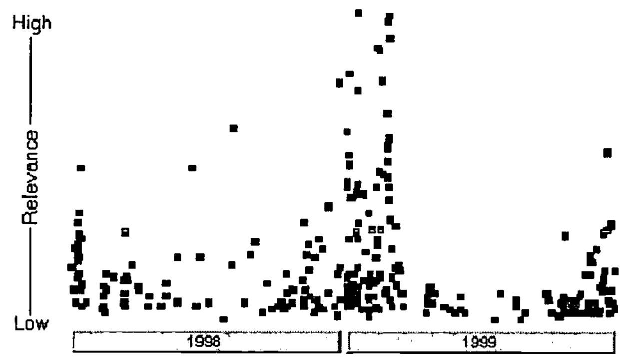

Figure 5. Timeline plot for Figure 4 subset, showing density of results in winter months.

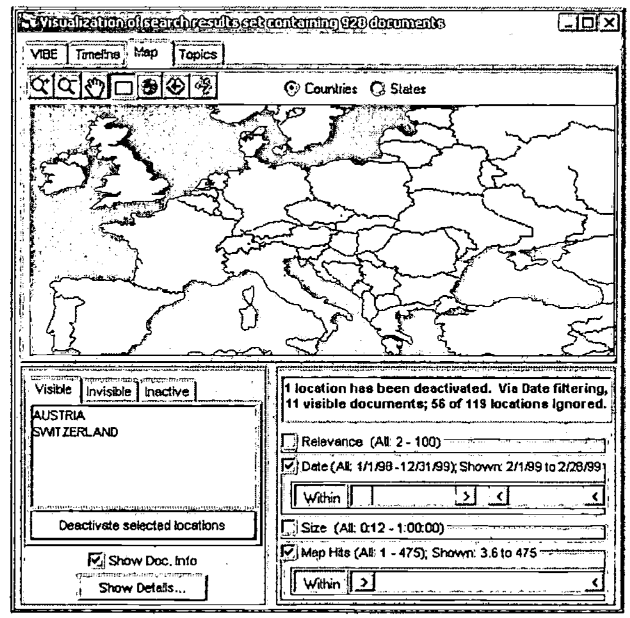

Figure 6. Map plot and dynamic query sliders, showing two European countries within February 1999 time focus.

This section has outlined through example the evolution of Informedia digital video library interface work. This work began with surrogates to enable the exploration of a single video document without the need to download and play the video data itself, and migrated to visualization techniques to allow the interactive exploration of sets of documents. A monolithic, Visual Basic Windows application provides these interfaces, allowing users to query or browse through text, image, and map searches, refine the result space with visualization techniques, and browse through surrogates such as titles, thumbnails, and storyboards.

The developments of the past year, particularly new W3C Recommendations and their implementation in major Web browsers, provided the opportunity to migrate this video library work to the Web. The remainder of this paper discusses this migration, emphasizing the benefits offered and the flexible library interface front-end provided to the user.

2. XML AND XSLT

"The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding" (verbatim from www.w3c.org). A number of key W3C Recommendations were published in 1999, enabling the separation of authoring from presentation in a standardized manner. In the case of the Informedia library, these recommendations allow the separation of video metadata from the library interface. Last year saw gradual implementation and support for these recommendations, documented at the W3C web site. The Informedia work described below makes use of the Microsoft XML Parser 3.0, an Internet Explorer add-on released by Microsoft in November 2000. The W3C Recommendations used in migrating Informedia interfaces to a Web browser include the following:

- XML (Extensible Markup Language): the universal format for structured documents and data on the Web, W3C Recommendation February 1998 [16].
- XML Schema: express shared vocabularies for defining the semantics of XML documents, not yet a full W3C Recommendation as of January 2001 [18].
- XSLT (XSL Transformations): a language for transforming XML documents, W3C Recommendation Nov. 1999 [19].
- XPath (XML Path Language): a language for addressing parts of an XML document, used by XSLT, W3C Recommendation November 1999 [17].

Other emerging standards for synchronized media metadata, such as MPEG-7 [9] and SMIL [15], will be tracked and incorporated as they become adopted by video streaming services and web browsers.

"Metadata" describes an information resource; it is "data about other data" [8]. A metadata record consists of a set of attributes necessary to describe the resource in question. For the Informedia video library, some attributes such as the producer, copyright holder, and broadcast date are given. A number of other attributes, such as start and end times, shot sequences, thumbnails, and transcripts, are automatically derived as input video is processed, segmented into documents, and catalogued.

The Informedia metadata is stored in a relational database and accessed through the application overviewed in Section 1. Such a closed system makes interoperability with other digital libraries difficult. A separate video collection might be described with a different set of metadata, or have that metadata stored in a different fashion.

An idealistic vision is to have a standard video metadata scheme, so that all video collections could be described to the same level of detail, accessed in the same manner, and have identical surrogates and interfaces built from the common scheme. However, video genres like news, sports, situation comedies, travel, lectures, and conference presentations have such diverse features that deriving a detailed, general video library metadata scheme will be a difficult if not impossible task. More likely, a common metadata framework will evolve, probably with input from professional societies in related disciplines like the Association of Moving Image Archivists. Using this common metadata framework as a foundation, more specific metadata could be added to more accurately describe resources in particular video collections.

The Dublin Core Metadata Initiative provides a fifteen-element set for describing a wide range of resources. While the Dublin Core "favors document-like objects (because traditional text resources are fairly well understood)" [8], it has been tested against moving-image resources and found to be generally adequate [7]. The Dublin Core is also extensible, and has been used as the basis for other metadata frameworks, such as an ongoing effort to develop interoperable metadata for learning, education and training, which could then describe the resources available in libraries like the Digital Library for Earth System Education (DLESE) [6]. Hence, Dublin Core is an ideal candidate for a high-level metadata scheme for the Informedia video library. An outside library service, with likely support for Dublin Core, would be able to make use of information drawn from the Informedia video library expressed in the Dublin Core element set.

The Dublin Core metadata for Informedia documents can be expressed as XML and validated through the use of a data type definition, or XML schema. More detailed metadata is necessary to produce the interfaces shown in Figures 1 through 6, but this metadata too can be expressed as XML and validated through a more comprehensive XML schema. In fact, a richly detailed XML document can be transformed into a minimal Dublin Core view, or transformed into views like those shown in Figures 1 through 6, with transformations performed via XSLT. Multiple XSLT transformations, e.g., one for low bandwidth users, another for high bandwidth users, optional additional ones for specific languages, age groups, etc., allow the video data to be widely disseminated in different forms based on W3C standards.

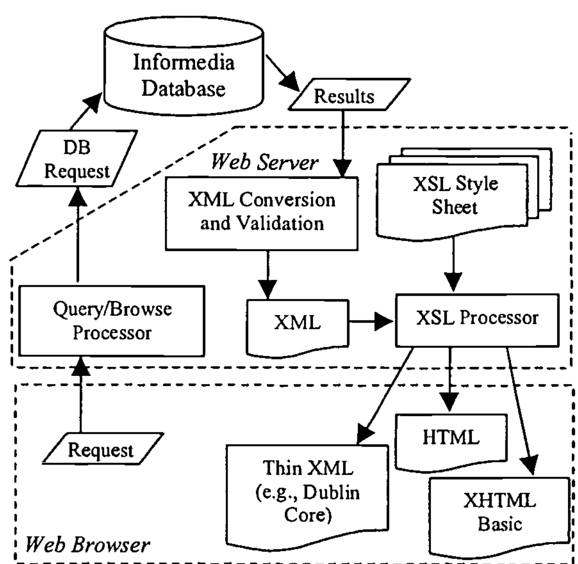

Figure 7. Architecture showing multiple outputs from XSL processing.

Figure 7 shows the process of a query or browse request against the Informedia database, producing XML results that are validated and data-typed via an XML schema. These XML results can be processed with different XSL style sheets to produce different library interfaces, such as an XML view consisting of Dublin Core elements, an HTML view that may look like Figure 1, or an XHTML Basic view suitable for display everywhere, including tiny PDAs. The next section gives specific examples, and discusses how tailored library access can be enhanced with XSL processing done in the client web browser rather than the web server.

3. TAILORED ACCESS TO DIGITAL VIDEO LIBRARY MATERIALS

In a recent editorial on "informationitis", Ramesh Jain notes that today's Web users and digital library patrons are overwhelmed by too much information. The traditional means for retrieving information has been keyword indexing and search, but abstracting the search level to keywords removes a great deal of relevant context for multimedia documents. In addition, presenting a list of documents returned from a keyword query involves perhaps a painstaking linear traversal of the list to find a document, with no gestalt view of the query space nor the results, i.e., no understanding of the relationship between result documents [11]. The editorial reinforces the Informedia interface conclusions drawn in the opening section: as the library contents increase in quantity, information visualization approaches need to be employed to facilitate understanding and navigation through larger document sets.

Speech recognition, image processing, and natural language processing allow automatic derivation of metadata to use as building blocks for subsequent generation of interfaces such as those shown in Section 1 [14]. The same metadata can be stored as XML and converted into numerous views through XSLT, where the views are tailored to a user's needs and bandwidth requirements. This section presents examples of XML and XSLT that implement such views, and discusses an architecture fostering quick presentation of multiple views into the digital video library, based on user selection. Users drive the library exploration and navigation, highlighting different aspects of document context to address their information needs and overcome "informationitis."

3.1 Informedia Access through XML and XSLT

Consider Figure 1 once again, showing a thumbnail view for a set of documents retrieved through an Informedia search service. These documents could be described in XML, as follows (listing shows only first and eighth result for Figure 1, to save space):

```
<IDVSet xmlns:im="x-
    schema:idvSchema.xml">
  <im:doc>
    <im:id>160814</im:id>
    <im:pos>1</im:pos>
    <im:shot>1961294</im:shot>
    <im:d_yr>1999</im:d_yr>
    <im:d_mo>1</im:d_mo>
    <im:d_day>14</im:d_day>
    <im:score>100</im:score>
    <im:dur>151250</im:dur>
    <im:mmss>2:31</im:mmss>
    <im:title>On Monday that cold air in place over
       upper midwest and great lakes with
       showers over midwest and snow in great
       lakes ...</im:title>
  </im:doc>
  <im:doc>
    <im:id>157053</im:id>
    <im:pos>8</im:pos>
    <im:shot>1931480</im:shot>
    <im:d_yr>1999</im:d_yr>
    <im:d_mo>1</im:d_mo>
    <im:d_day>2</im:d_day>
    <im:score>80</im:score>
    <im:dur>52120</im:dur>
    <im:mmss>0:52</im:mmss>
    <im:title>Villagers who escaped avalanche,
       had to dig through two meters of snow to
       reach through friends and relatives who
       were trappe...</im:title>
  </im:doc>
</IDVSet>
```

The referenced schema "idvSchema.xml" is used to validate and provide data type semantics for this XML text. Consider this subset of contents from idvSchema.xml:

```
<?xml version="1.0" ?>
<Schema name="IDVResultsSchema"
    xmlns="urn:schemas-microsoft-com:xml-data"
    xmlns:dt="urn:schemas-microsoft-
    com:datatypes">
  <ElementType name="score" content="textOnly"
     dt:type="ui1" />
  <ElementType name="doc" content="mixed">
     <element type="score" maxOccurs="1" />
  </ElementType>
</Schema>
```

These schema definitions limit "score" to appearing at most once for each document "doc", with "score" being an unsigned one-byte integer. The schema defines other requirements and types for "IDVSet." The validated XML can be transformed into the view shown in Figure 8 through the following XSL style sheet, which loops through each im:doc document metadata and converts it into appropriate HTML:

```
<xsl:stylesheet xmlns:xsl=
  "http://www.w3.org/1999/XSL/Transform"
  version="1.0" xmlns:im=
  "x-schema:idvSchema.xml">
  <xsl:output method="xml" indent="yes"
    omit-xml-declaration="yes" />
  <xsl:template match="/">
     <xsl:apply-templates />
  </xsl:template>
  <xsl:template match="IDVSet">
   <xsl:for-each select="im:doc">
    <xsl:sort select="im:score" order="descending"
       data-type="number" />
    <span class="resultStamp" id="R{im:pos}"
       rdb_id="{im:id}"
       onclick="stampClick(this);"
       onmouseover="stampChangeOver(this);"
       onmouseout="stampChangeOut(this);">
    <img id="Stamp_{im:pos}"
       src="graphics/Gstamp.gif" alt=""
       orgsrc="graphics/Gstamp.gif"
```

```
    oversrc="graphics/Gltstamp.gif"
    width="112" height="91" />
  <xsl:variable name="ScoreHt"
    select="round(im:score * 0.8)" />
  <!-- map 100 score to 80 px (im:score .le.100) -->
  <img id="Th_{im:pos}" src="graphics/red.gif"
    alt="">
   <xsl:attribute name="style">
   position:absolute; left:9; width:4; top:
    <xsl:value-of select="85-$ScoreHt" />
   ; height:
    <xsl:value-of select="$ScoreHt" />
   ;
   </xsl:attribute>
  </img>
  <img id="I_{im:pos}"
    style="position:absolute; left:23; top:9"
    width="80" height="55">
   <xsl:attribute name="src">
    <xsl:choose><xsl:when test="im:shot[. !=
          0]">GetShot.asp?<xsl:value-of
          select="im:shot" />
    </xsl:when>
    <xsl:otherwise>Graphics/viddeflt.gif
    </xsl:otherwise></xsl:choose>
   </xsl:attribute>
  </img>
  <img id="tip" src="Graphics/1p-trans.gif"
    style="position:absolute; left:0; top:0"
    width="112" height="91">
   <xsl:attribute name="alt">
    <xsl:value-of select="im:title" />,
    <xsl:value-of select="im:mmss" />,
    <xsl:value-of select="im:d_mo" />-
    <xsl:value-of select="im:d_day" />-
    <xsl:value-of select="im:d_yr" />
   </xsl:attribute>
  </img></span></xsl:for-each>
 </xsl:template>
</xsl:stylesheet>
```

Figure 8. Browser display of XSL-transformed XML into HTML (a view similar to Figure 1).

XSLT is itself an XML document, and so the style sheet above reads as a jumble of starting and ending XML tags that essentially do the following: for each Informedia document, create a green stamp area (Gstamp.gif) with the relevance score in red on a vertical bar, a thumbnail image if given a valid nonzero identifier, and pop-up title text, duration, and broadcast date information. The produced html from this XSLT for document 8 is as follows:

```
<span class="resultStamp" id="R8" rdb_id="157053"
    onclick="stampClick(this);"
    onmouseover="stampChangeOver(this);"
    onmouseout="stampChangeOut(this);"
    xmlns:im="x-schema:idvSchema.xml">
  <img id="Stamp_8" src="graphics/Gstamp.gif"
     alt="" orgsrc="graphics/Gstamp.gif"
     oversrc="graphics/Gltstamp.gif" width="112"
     height="91" />
  <img id="Th_8" src="graphics/red.gif"
     style="position:absolute; left:9; width:4; top:31;
     height:64;" />
  <img id="I_8" style="position:absolute; left:23;
     top:9" alt="" width="80" height="55"
     src="GetShot.asp? 1931480" />
  <img id="tip" src="Graphics/1p-trans.gif"
     style="position:absolute; left:0; top:0"
     width="112" height="91" alt="Villagers who
     escaped avalanche, had to dig through two
     meters of snow to reach through friends and
     relatives who were trappe..., 0:52, 1-2-1999" />
</span>
```

3.2 Enhancing Views with Match Data

By extending this simple opening example, match data information can be viewed by users in the same way as shown in Figure 1: through color coding of query terms and the vertical relevance score bar. The XML and schema definitions are extended to include information on which entities (in this case, words, but could be geographic regions, image features, etc.) match a video document, by how much and where:

```
<IDVSet xmlns:im="x-
      schema:idvResSchema.xml">
  <im:ScoreInfo>
    <im:ScoreEntity><im:mID>1</im:mID>
      <im:mLabel>cold</im:mLabel>
    </im:ScoreEntity>
    <im:ScoreEntity><im:mID>2</im:mID>
      <im:mLabel>snow</im:mLabel>
    </im:ScoreEntity>
    <im:ScoreEntity><im:mID>3</im:mID>
      <im:mLabel>ice</im:mLabel>
    </im:ScoreEntity>
    <im:ScoreEntity><im:mID>4</im:mID>
      <im:mLabel>avalanche</im:mLabel>
    </im:ScoreEntity>
  </im:ScoreInfo>
  <im:doc>
   {"doc" contents, e.g., im:id, im:pos as before}
    <im:m>
      <im:msrc>3</im:msrc>
      <im:mScore>386</im:mScore>
      <im:mOffset>528</im:mOffset>
    </im:m>
    <im:m>
      <im:msrc>3</im:msrc>
      <im:mScore>484</im:mScore>
```

```
    <im:mOffset>528</im:mOffset>
  </im:m>
  <im:m>
    <im:msrc>1</im:msrc>
    <im:mScore>333</im:mScore>
    <im:mOffset>249</im:mOffset>
  </im:m> {additional im:m continue here...}
</im:doc>
  {other "doc" contents likewise extended with
    match information via the im:m element}
</IDVSet>
```

The XSL style sheet is extended to make use of im:m match information, producing the view shown in Figure 9, which interactively changes the query word colors to indicate which words match in that document, and shows itemized scoring entity contributions in the vertical relevance bar (as done in Figure 1).

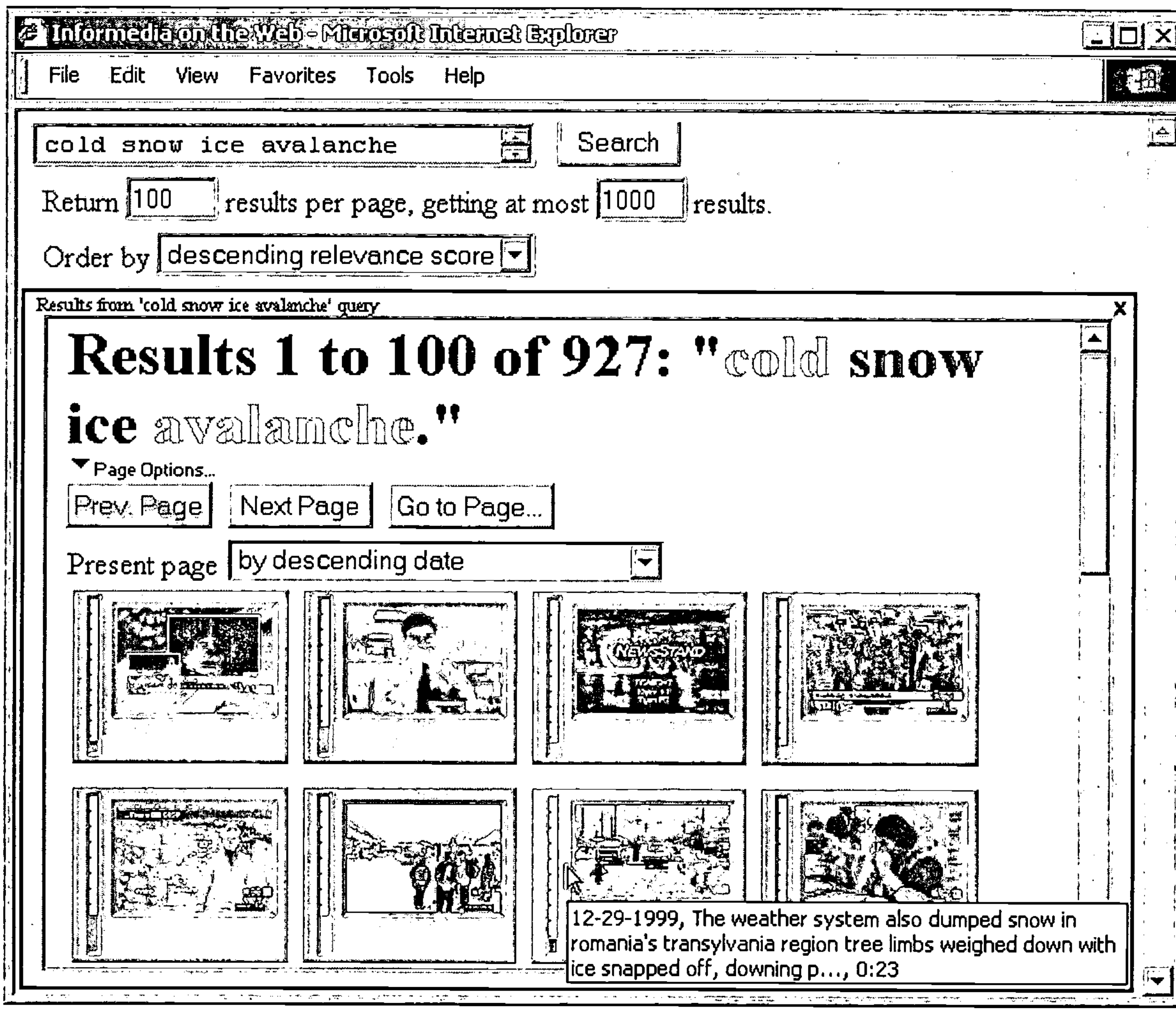

Figure 9. Display of HTML produced via XSL transformation of XML with match data.

3.3 Client-Side XSLT

The addition of XML data provides new interface functionality possibilities. By continuing with this strategy, the Informedia document XML description and its validating schema can be extended to that data necessary to generate all the interfaces described in Section 1, interfaces proven useful through prior investigations. The problem with such an approach is that perhaps the XML or XSLT-produced HTML would grow to huge sizes that take time to download in a Web browser, but never get viewed. Through XSLT in the client browser, however, users have the freedom to choose which views to use, with little or no need for communication back with the Web server.

Figure 9 shows a "Present page" option where the user can select to order the page by relevance, date, or document size in ascending or descending order. The change in sort is accomplished through an XSL style sheet, e.g., the descending date is accomplished via the following:

```
<xsl:sort select="im:d_yr" order="descending"
  data-type="number" />
<xsl:sort select="im:d_mo" order="descending"
  data-type="number" />
```

```
<xsl:sort select="im:d_day" order="descending"
 data-type="number" />
```

The style sheet also reorders the pop-up information to give precedence to the date and lists that first, capitalizing on past experience that when sorting Informedia documents by date the user is more interested in that attribute and prefers such a reordering. Of course, the XSL style sheet could be altered to make the date information even more explicit. Client-side XSL transformations allow the user to sort and present the data to meet his or her specific browsing and information-seeking needs.

Other options available in "Present within page" include a text-centric view, shown in Figure 10, and a VIBE view, identical to Figure 4 and making use of the same XML with match information described in Section 3.2.

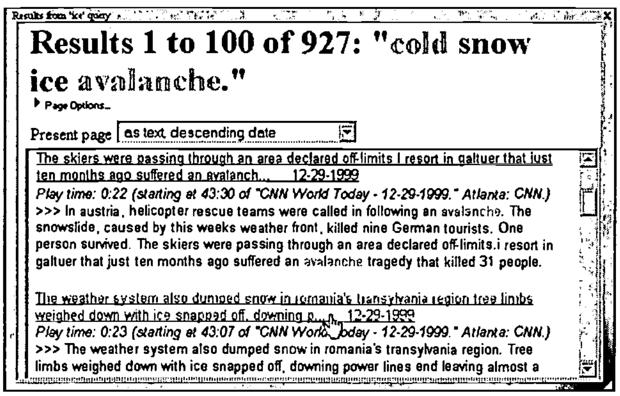

Figure 10. Text view of same XML data presented with thumbnail grid in Figure 9.

The web architecture for the Informedia library is pushing dynamic interface selection to the user by sending XML data and XSL style sheets to the client-side browser. The first time a style sheet is referenced, e.g., to sort a page by date and emphasize that attribute, the style sheet is inserted into the browser's cache. Subsequent use of that transformation can then be applied without the need to contact the Web server. The user is free to explore multiple features of the document space through different views, from text-centric to image-centric, from linear lists to visualization strategies like VIBE.

The interface shown in Figure 9 also lets the user specify the size of the document set to be considered via multiple views, i.e., the "page size" indicating the number of documents described in XML for subsequent translation into HTML via XSLT. The user also sets the maximum number of documents cached at the server for potential future consideration. In this manner, users can control the flow of information to meet their bandwidth restrictions and patience thresholds. For example, a user on a T1 line may set a page size of 1000 and look through image-rich presentations such as multiple storyboards (Fig. 2), while a user with a PDA and 56 Kbps access may set the page size to 20 and make use of text-centric views.

While some transformations may require contacting the Web server to get additional data such as imagery, others are done completely at the client, making use of the original XML or previously cached information, as overviewed in Figure 11. For example, suppose the user initially defaulted to sorting documents by date, producing the html whose display is shown in Figure 9. The user now sorts and emphasizes by score, resulting in an ordering as shown in Figure 8. No new imagery is necessary, as the thumbnail image data has already been cached by the browser. Suppose the user now accesses "Present page by VIBE view" which requires the VIBE XSL style sheet to be downloaded the first time it is referenced. The style sheet is less than 2 KB in size, and is available in the browser's cache for quick reuse without the need to contact the web server the next time it is needed. Style sheets will generally be very small compared to the XML document. A vastly different VIBE presentation (Figure 4) of the document set utilizing match information is shown to the user with this style sheet, without needing to retrieve additional XML or data from the Informedia database.

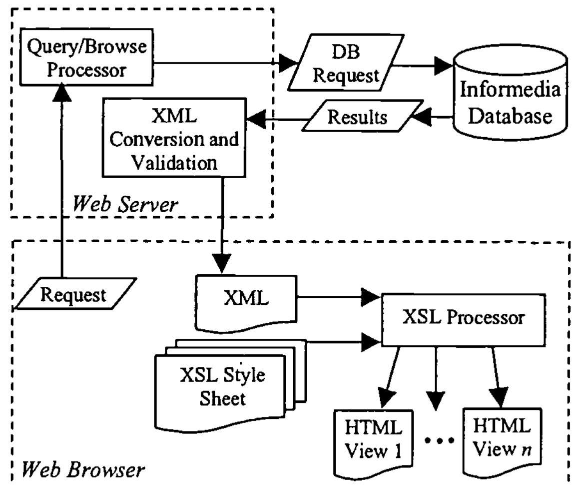

Figure 11. Overview of client-side XSL processing, where user interaction can produce multiple HTML views without Web server involvement.

3.4 Flexibility via XML and XSLT

Figure 7 shows already processed XML data being sent to clients. This architecture is useful for those clients with very focused or well-articulated needs. For example, another library service may need an Informedia document set expressed as Dublin Core elements, and the document set can be translated into that format by the Informedia Web server and sent to that service.

By contrast, Figure 11 shows XML data, along with XSL style sheets being communicated to clients. This allows clients to modify the views dynamically, offering flexibility to address the "informationitis" issues for multimedia libraries discussed in Jain's editorial [11]. Users can vary the views dynamically: those interested in image-rich overviews by date can be satisfied, as can users interested in query-specific set manipulation offered through VIBE. Given the numerous attributes and views into video collections, and the potential of each view to inform the user about specific characteristics like date, length, or geographic coverage, this architecture delays final rendering (in HTML or whatever form) of the semantic XML data until decisions made within the Web browser. In the examples used here, decisions are made through the "Present page" option.

4. CONCLUSIONS AND FUTURE WORK

Much work remains to be done in order to provide interoperable, tailored Web browser views into the Informedia library as expressive as those of the stand-alone Informedia library application. The W3C recommendations provide the ideal framework for creating these views, given the W3C's charter, broad industry support, and momentum from other National Science Foundation DLI-2 and NSDL projects also moving toward XML and XSLT; see for example DLESE [6] and the ACM SIGGRAPH Education Committee Digital Library [13]. XML, XSLT and related technologies XPATH and XML schemas allow semantics to be recorded, navigated, validated and translated in standard ways.

A necessary condition for widespread interoperability amongst digital video collections is agreement on a common metadata framework, as discussed in the usage guide for Dublin Core [8]. A common video metadata framework can be supported by Informedia and other video libraries through a default XSLT transforming the libraries' XML into this framework's XML. In all likelihood this framework would be an extension of Dublin Core, much as other groups such as metadata committees for learning, education and training are exploring use of Dublin Core as a foundation. A small subset of what such a minimal framework would look like for an Informedia document is as follows:

```
<?xml version="1.0" ?>
<!DOCTYPE rdf:RDF SYSTEM
  "http://purl.org/dc/schemas/dcmes-xml-
  20000714.dtd">
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
   syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description
   about="http://informedia.org/seg160814.mpg">
<dc:title>CNN World Today</dc:title>
<dc:description>On Monday that cold air in place over
   upper midwest and great lakes with showers
   over midwest and snow in great lakes
   ...</dc:description>
<dc:date>1999-1-14</dc:date>
<dc:format>video/mpeg</dc:format>
<dc:language>en</dc:language>
<dc:publisher>Cable News Network</dc:publisher>
<dc:contributor>Carnegie Mellon University
   Informedia Project</dc:contributor>
{Many more descriptors needed, e.g., coverage is from
   49:22 to 51:53 of the hour-long "World Today" show.}
</rdf:Description>
</rdf:RDF>
```

We will track closely the work of other digital libraries like DLESE that manage video resources, as well as the industry initiatives such as the work within the Association of Moving Image Archivists, as they address a common video metadata framework. In addition to providing a minimal but broadly applicable view (Figure 7), we also have the goal of migrating Informedia surrogates and visualizers to HTML-based expressions, so that they can be generated dynamically through XSL processing against XML within Web clients (see Figure 11). Hence, we will have a more detailed, "Informedia-rich" XML schema capable of supporting such enhanced views as those shown in Figures 1 through 6.

Work to date has addressed thumbnail grids, ordering, and query word-based views, including VIBE. Work is ongoing to provide interactive map interfaces, where zooming, panning, and map layer highlighting can be performed dynamically and efficiently. These features are required to provide a map visualization service like that shown in Figure 6, where countries highlight in different colors based on the user dragging a time period indicator across a scroll bar. We are currently investigating another W3C format, the Scalable Vector Graphics (SVG) format available as a Candidate Recommendation as of early 2001. SVG will allow quick map updating in the browser, as well as allow VIBE rendering to be more efficient so that greater numbers of documents can be shown simultaneously.

Improving summarization and visualization across video document sets is an ongoing research activity within the Informedia Project [10], and as new techniques become available, they will be added to the set of XSL style sheets available to the Informedia library patron. For example, work continues to identify faces within the video library, and name those faces with proper names. An interesting visualization along the lines of Figures 4 through 6 would be a key person/player view showing people's faces who dominate the news for particular time periods or for a specific text, image or geographic query.

We will continue implementing XSL style sheets and updating the Informedia-rich XML to allow users to have multiple views into the Informedia document sets. Future work includes usability tests on these views to investigate their utility and to determine the costs and benefits in supporting client-side XSL processing. Informedia metadata in particular is unusual compared to other libraries in that it is errorful, produced through automatic means without manual cataloging. Studies will need to be run to determine the effects of errorful metadata on subsequent XSL transformations and ultimately on the user's experience.

We will need to revisit the architecture of Figure 11 over time to see whether multiple style sheets operate on the same XML, or whether each style sheet has unique requirements for additional metadata from the Informedia database, and hence must contact the Web server anyway. If each XSL style sheet is essentially independent, requiring contacting the Web server, then there is no advantage to client-side XSLT. However, our first trials using XML with match information (Section 3.2) shows that the same XML supports diverse views, from thumbnails to plain text to VIBE. By adding match information to the "Informedia-rich" XML set, a match-specific view such as VIBE can be implemented through client-side XSLT. When a map view is added, metadata about geographic coverage for each Informedia document will need to be added to the XML. Should named faces be added, that metadata will need to be added to the XML as well. The same base XML can be grown to cover all the views, so that it is downloaded once and then operated on in the browser, an option that may be feasible given the expense of video data.

Video streaming is only now starting to reach a broader audience on the web. Video still requires comparatively large bandwidth and network integrity, and playback of web video beyond the tiny postage stamp window requires patience from even the well-connected university user on a T1 line. Users therefore may be willing to wait seconds to download lots of XML and associated XSL style sheets, so that they can then quickly browse through metadata representing hundreds of hours of video and megabytes or terabytes of actual video data. The views from XSLT allow a careful exploration of that material before investing in minutes or longer of video download time. Through the tailoring techniques described here, video library patrons can browse and explore video assets with minimal time commitments through surrogates and visualizations. These interfaces are rendered through W3C standards for increased potential to work within and across other digital video collections on the Web.

5. ACKNOWLEDGMENTS

Partial support for this work is provided by the National Science Foundation's National Science, Mathematics, Engineering, and Technology Education Digital Library Program under grant DUE-0085834. This work is also supported by the National Science Foundation under Cooperative Agreement No. IRI 9817496.

6. REFERENCES

[1] Christel, M., Winkler, D., and Taylor, C.R. Improving Access to a Digital Video Library. In *Human-Computer Interaction: INTERACT97, the 6th IFIP Conf. On Human-Computer Interaction*, Chapman & Hall, London, 1997, pp. 524-531.

[2] Christel, M., Smith, M., Taylor, C.R., and Winkler, D. Evolving Video Skims into Useful Multimedia Abstractions. In *Proceedings of CHI '98* (Los Angeles, CA, April 1998), ACM Press, 171-178.

[3] Christel, M.G., Hauptmann, A.G., Warmack, A.S., and Crosby, S.A. Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library. In *Proc. ADL '99* (Baltimore MD, May 1999), IEEE Press, 98-104.

[4] Christel, M.G. Visual Digests for News Video Libraries. In *Proc. ACM Multimedia '99* (Orlando, FL, Nov. 1999), ACM Press, 303-311.

[5] Christel, M.G., Olligschlaeger, A.M, and Huang, C. Interactive Maps for a Digital Video Library. *IEEE MultiMedia* 7(1), 60-67.

[6] Ginger, K, web page maintainer. DLESE Metadata Working Group Homepage, Nov. 6, 2000, http://www.dlese.org/Metadata/index.htm.

[7] Green, D. Beyond Word and Image: Networking Moving Images: More Than Just the "Movies", *D-Lib Magazine*, http://www.dlib.org/dlib/july97/07green.html.

[8] Hillman, D. Using Dublin Core. July 16, 2000, http://purl.org/dc/documents/wd/usageguide-20000716.htm.

[9] Hunter, J. MPEG-7: Behind the Scenes. *D-Lib Magazine*, http://www.dlib.org/dlib/september99/hunter/09hunter.html.

[10] Informedia Project web site at Carnegie Mellon University, http://www.informedia.cs.cmu.edu.

[11] Jain, R. Informationitis. *IEEE MultiMedia* 7(4), 1, 5.

[12] Olsen, K. A., Korfhage, R. R., Sochats, K. M., Spring, M. B., and Williams, J. G. Visualization of a Document Collection: The VIBE System. *Information Processing & Management* 29(1), 69-81.

[13] Owen, G.S., Sunderraman, R., and Zhang, Y. Use of Database and XML Technology for Retrieval and Repurposing of DL Contents. Presentation at DLI-2 All Projects Meeting (Stratford-upon-Avon, U.K., June 2000), http://asec.cs.gsu.edu/asecdl-nsf-dli2/presentations/all-projects-meeting-6-2000-uk/asecdl-xml_files/frame.htm.

[14] Wactlar, H., Christel, M., Gong, Y., and Hauptmann, A. Lessons Learned from the Creation and Deployment of a Terabyte Digital Video Library. *IEEE Computer*, 32, 2 (Feb. 1999), 66-73.

[15] World Wide Web Consortium Synchronized Multimedia (SMIL) 1.0 Recommendation, June 1998, http://www.w3.org/AudioVideo/.

[16] World Wide Web Consortium Extensible Markup Language (XML) 1.0 Recommendation, Feb. 1998, http://www.w3.org/XML/.

[17] World Wide Web Consortium XML Path Language (XPATH) 1.0 Recommendation, Nov. 1999, http://www.w3.org/TR/xpath.html.

[18] World Wide Web Consortium XML Schema Candidate Recommendation, Oct. 2000, http://www.w3.org/XML/Schema.

[19] World Wide Web Consortium XSL Transformations (XSLT) 1.0 Recommendation, Nov. 1999, http://www.w3.org/TR/xslt.

U.S. Department of Education
Office of Educatonal Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE
(Specific Document)

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (9/97)