Efficient Visualization of Content and Contextual Information of an online Multimedia Digital Library for Effective Browsing

Surjeet Mishra    Amarendra Gorai    Tavleen Oberoi    Hiranmay Ghosh
TCS Innovation Labs, Delhi
Tata Consultancy Services Ltd., India.
{surjeet.mishra, amarendra.gorai, tavleen.oberoi, hiranmay.ghosh}@tcs.com

Abstract

In this paper, we present a few innovative techniques for visualization of content and contextual information of a multimedia digital library for effective browsing. A traditional collection visualization portal often depicts some metadata or a short synopsis, which is quite inadequate for assessing the documents. We have designed a novel web portal that incorporates a few preview facilities to disclose an abstract of the contents. Moreover, we place the documents on Google Maps to make its geographical context explicit. A semantic network, created automatically around the collection, brings out other contextual information from external knowledge resources like Wikipedia which is used for navigating collection. This paper also reports economical hosting techniques using Amazon Cloud.

1. Introduction

The amount of multimedia contents in the internet is increasing hastily with each passing day. Digital libraries organize these documents and allow users to browse through the collection. But information associated with each document is limited to a short synopsis (and sometimes an image) in today’s portals like BlinkxTV.com. A media instance (a still image or a video) is a complex entity. Videos, perhaps the most complex, follow a hierarchy of shots and scenes, and the information present in short synopsis is inadequate in concluding the relevance of the document to users’ need. The mise-en-scene objects in a multimedia data-stream interact with each other in space and time to connote meaningful events. Earlier works on collection visualization [1,2] (re-)organize the collection to create different views of it based on classical distance measures of information retrieval theories. These distances can be computed for text documents alone and the algorithm cannot be extended to multimedia collections of digital libraries like European Navigator. The content and contextual information, which is of more significance to the user, remains unexplored.

To overcome these limitations we present some visualization techniques for presenting a vast content and context information that is associated with media documents. To illustrate the content information of a video, the portal is equipped with features like Scene Grid, Story Board and Excerpt. The central theme and location of documents can be used to correlate them. We visualize geographic context using Google Maps and thematic context using domain ontology graph. Earlier works have been done to visualize multimedia collection using manually-created ontology [3] which imposes the burden on content creator and the ontology gets biased to creators’ perception of the collection. We automatically construct an ontological graph for the collection by analyzing the links of large public knowledge resource (Wikipedia) and map information nodes (individual document) on it, thereby making the implicit knowledge organization in the collection explicit. Several large graph visualization approaches have been presented in [4]. Our approach is unique as it presents abstract views of the knowledge structure at different levels, does not clutter the display with too many nodes and enables flexible navigation operations. We allow the user to have a scene level search to satisfy his/her requirement. Though the features are proficient in bringing out the information hidden in documents, there needs to be a well formed content model around each document. This can be done using our content creation tool (Editing Tool) which takes the content creator on a tour of content enrichment for each document and allows user to go through the workflow as many time till satisfied with the content created.

The paper is organized as follows. Section 2 describes the system in details. Section 3 discusses deployment on cloud and section4 concludes the paper.
2. Description of work

Effective interaction with any multimedia content like video requires creation of a comprehensive content description. The effectiveness is related to the overall user experience with the content usage, satisfying user needs. Our system is built over the Multimedia Explorer framework [5] which relies on creation content description for videos for effective utilization. The framework has been modified (see Fig.1). The system consists of two parts: Editing Tool for enriched content creation around documents and Web Portal for its efficient visualization as described below.

2.1. Editing Tool

During editing phase each video goes through two phases: video segmentation and content creation. Initially the document level content and context information is entered by the content creator. The description and other basic information like director, author, etc. constitute the content information, and the context information is stored by means of geographical coordinates (latitude, longitude) and by creating domain ontology. Once document level information is stored in the content model, the documents (videos) needs to be ingested and automatically processed.

2.1.1. Automatic Video Segmentation. When a video is ingested into the system, this module extracts the structural information and some feature descriptors, which are required later for content attachment and for retrieval in Web Portal. First, the video is processed to extract the frames and detect the shot boundaries. The shots are detected on the basis of difference in visual features of successive frames using adaptive thresholding. Information like representative frame- an image which describes the shot, free text annotation, visual features of frames, is extracted. This structural information of document is saved as per the multimedia content model proposed in MPEG-7 standards [6]. As these shots are computer generated they do not convey a complete story or message and are of no semantic use until meaningful scenes are created by merging them.

2.1.2. Content Creation. Since different scenes do not contain any relation (similarity or dissimilarity in particular) based on their physical properties, an automated system isn’t ideal for scene detection. To bring out the semantics, human interference is required. The tool allows content creator to play shots separately, find the information flow and create meaningful scenes. It also allows the user to select individual shots that would be showed to portal user as excerpt.

Once the excerpt and scenes are defined the user can attach the textual description - annotation and tags, to each scene. Annotation is a short textual description of the scene. Tag is a single word or phrase that depicts the concepts in scenes. Tags are later used in searching the document based on users’ queries. Tags are used to create ontology around the collection by associating them with information nodes (topics) from Wikipedia.

For effective visualization and navigation, the documents in the collection need to be interrelated on a knowledge space. The ontology is a formal tool to represent a bounded knowledge domain. We have used Wikipedia to create an informal ontology, motivated by [7]. Every article in Wikipedia represents a topic, which belongs to one or more categories. However, we have created a constrained view of the ontology by restricting it to contain the category nodes that are just sufficient to cover the information nodes in the collection. The metadata associated with the documents is used to relate a document to one or more Wikipedia topics. The topics related to a document are sometimes explicitly given as tags. In absence of such explicit association, the related topics can be extracted from the annotation of documents using different text processing techniques. Once the topics pertaining to a document have been identified, we create ontology superstructure for the document in a bottom-up manner starting from a topic and going up the Wikipedia category hierarchy ladder using Depth First Traversal, till we reach ‘Main topic classifications’ node. This process is repeated for every topic in a document and for every document in the collection, merging the common sub-graph as soon as it is discovered. Thus, a document is associated with

---

1. A shot is a segment of the video that comprises of a sequence of visually similar frames of a video stream.

2. A scene is composed of series of shots stitched around a central theme that conveys a meaningful message or story.
one or more leaf category nodes in the ontology (Fig. 2). This automated process helps in updating the ontology in an incremental way for new documents.

![Multimedia Document Collection](image)

**Figure 2. Knowledge organization in Ontology**

For videos we allow the scene level content enrichment which is not required for other types of documents. This way the content is more detailed and a lot of scene level information is preserved which is otherwise lost at the video level synopsis. The user can go through the different phases in content creation till completely satisfied with it. Now the content model is complete and is uploaded into the web portal.

### 2.2. Web Portal

The web portal presents the content and contextual information associated with each document to the content user (portal user) by the means of document details page, some static features like Story Board & Excerpt and some interactive features like Ontological Graph & Google Maps.

#### 2.2.1. Content Visualization

The content information is presented by three features viz., scene grid, story board and excerpt. In the document details page, the basic information about the documents, like cover image, director’s/author’s name, synopsis, are presented along with the grid of representative images of each scene. The user can view scene description by bringing the mouse pointer over each representative image. Using the Scene Grid, the user can see all scene level information at a single place, but the flow of the story and the central message is yet to be presented. So, we developed Story Board to bring out the storyline and Excerpt to present the central theme of the document.

**STORY BOARD:** As it’s well known that every document meanders around different concepts/themes, a small synopsis will not serve purpose. Video is a sequence of scenes carrying different stories. To project the complete storyline we have Story Board where the representative image of each scene is shown along with its annotation. This sequential presentation helps the user to get a feel of the visual content of each scene and the complete storyline of the video.

**EXCERPT:** An excerpt is a visualization technique that shows some selected shots bringing out important messages to portal users. The shots marked by the content creator during content creation are stitched into a short trailer. The difference between conventional video trailer and the excerpt is that excerpts can be changed by just changing the shot selection. It provides flexibility to content creator to create a good essence for the complete video that brings out the message without using a commercial video editing tool.

![Scene Grid, Story Board, Excerpt, Google Map](image)

**Figure 3. Documentary Details**

#### 2.2.2. Context Visualization

Every document is created with a central theme and a central geographic location. These thematic and geographic contexts are inter-related for many documents and so different documents can be related with each other based on these context. For example many books and videos are created around the Indian Independence but on its different facets which users can use to relate them. We present the thematic context for each document using ontology graph and the geographic context by mapping them on Google Maps.

**GEOGRAPHIC CONTEXT:** We map documents over the atlas based on their central location using Google Maps (Fig. 3 top-right side). The idea is to allow user to view different documents based on a certain geographic location. As year is displayed along with each document, the user can view documents and compare them over the geographical and time domain.

**THEMATIC CONTEXT:** The main objective of creating the knowledge map (ontology) is to use it as the base for visualizing collection to digital library.
users. But the knowledge map so created contains thousands of category nodes, posing a challenge for visualization. For providing a better cognition to the user, we present an abstract view of ontology [8] by creating a sub graph that depicts the overall knowledge structure. We show the abstract views with few selected nodes and edges based on the user selected node.

The user can click on any node to navigate on the ontology graph. As a user zooms in, topics below the selected node gets magnified, topics in other regions shrink and the dynamically created document clusters split. The contextual view of the document collection is dynamically created as user navigates on it. A casual visitor is greeted with the overall structure of the collection, enabling a deep-dive into some broad areas of interest, progressively refine the information needs and contextually discover documents in the collection.

2.3. Search

The portal allows the user to search with names of director/authors and the documents corresponding to the query are displayed. Our portal gives user the provision to search in the content model of the documents with a list of tags. After searching, the corresponding document and scene detail is displayed. Scene level search is a unique feature that is provided and allows the user to drill down the search scope.

3. Deployment on the Cloud

Though our web portal is much lighter as compared to traditional video portals, the requirement for an economic and quick accessible storage is desired that is what a cloud computing service provider provides. Amazon Web Services resource charges are usage specific. We have also automated the start and stop of the cloud server rather than running it 24x7 to optimize resource usage. The server is invoked when a user accesses it and automatically shuts down after being idle for certain time, cutting down cost even more.

<table>
<thead>
<tr>
<th>Table 1. AWS cost saving using auto-control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resource</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Server (EC2)</td>
</tr>
<tr>
<td>Data Transfer (out)</td>
</tr>
<tr>
<td>Permanent Storage (S3)</td>
</tr>
<tr>
<td>Total Cost</td>
</tr>
</tbody>
</table>

Table 1 illustrates the projected cost saving for a video library (250 videos). The average data for each video is about 30MB since we do not store videos in our portal; we store only corresponding .swf files. Data downloaded is assumed to be 100GB/month. Using our automatic instance start/stop technique, we can achieve about 53% of cost even if it runs for 8 hours a day.

4. Conclusion

The portal stores only the contextual information regarding the videos and not the original video as it is a priced commodity. The content creator can provide detailed content description at the document as well as lower level. Based on the content model, visualization techniques that we present in this paper brings out vital information (both content and contextual) that is inherently associated with the multimedia document collection, helping the user to take an informed decision on selecting any document. This multimedia document visualization portal helps the online sellers to manage their multimedia collection efficiently, present it with a wide range of information and manage their business with better customer satisfaction.

5. References


