

ContextMiner: A Tool for Digital Library Curator

Chirag Shah
School of Information and Library Science
University of North Carolina
Chapel Hill NC 27599, USA
chirag@unc.edu

Gary Marchionini
School of Information and Library Science
University of North Carolina
Chapel Hill NC 27599, USA
march@ils.unc.edu

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces - *Interaction styles*

General Terms: Design, Human Factors, Management

Keywords: Digital curation, Digital Preservation, Contextual information

ContextMiner - Similar records			
Relevance ordering by title		Relevance ordering by source	
650	So They Tell Me	3270	Henry Wallace Speech
1447	Ask Me, Don't Tell Me	3414	Shopping Can Be Fun: A New Concept in Merchandising
3557	They Grow Up So Fast	1059	Human Beginnings
3857	Soundie - A Little Robin Told Me So		
1816	Kiss me		
771	Precisely So (Part II)		
773	Precisely So (Part I)		
3516	Cities How They Grow		
3812	Cities: Why They Grow		
3662	You Can Tell by the Teller		
Relevance ordering by description		Relevance ordering by keywords	
650	So They Tell Me	650	So They Tell Me
802	Peg-Leg Pedro	1476	America Goes Over (Part V)
967	Make Mine Freedom	1477	America Goes Over (Part IV)
3482	Make Mine Freedom	1478	America Goes Over (Part III)
1476	America Goes Over (Part V)	1479	America Goes Over (Part II)
1477	America Goes Over (Part IV)	1480	America Goes Over (Part I)
1478	America Goes Over (Part III)	701	San Francisco Liberty Loan Parade
1479	America Goes Over (Part II)	742	Remember These Faces
1480	America Goes Over (Part I)	1247	Divide and Conquer (Part IV)
3514	Information Machine, The	1248	Divide and Conquer (Part III)

Figure 1: *ContextMiner*: rank lists of relevant items based on different criteria

The Vidarch Project¹ aims to develop policies and tools that help video curators discover and add contextual elements that will help future generations not only find but also make sense of video content.

In order to help the curator in fetching and evaluating information for its relevance, we have developed a prototype system called *ContextMiner*². This web-based tool, implemented primarily using PHP and MySQL, helps the curator in looking for information on the Web or in specialized databases, compile it, and finally store in the repository. A typical flow of the *ContextMiner* system is given below.

1. The curator asks the system to grab a record from some specialized database. In our case, this is the Prelinger collection.
2. The record that the curator requested is displayed. Now the curator can ask the system to find similar items from some other sources.

¹<http://www.ils.unc.edu/vidarch/>

²Available from <http://www.ils.unc.edu/vidarch/>

Copyright is held by the author/owner(s).

JCDL'07, June 17–22, 2007, Vancouver, British Columbia, Canada.
ACM 978-1-59593-644-8/07/0006.

The screenshot shows a web form with several sections:

- Metadata:** Title (Angry Boy: Stephen Colbert's Challenge), Description (My submission to Stephen Colbert's lightsaber challenge...), Source (YouTube: starpilot), Keywords, Amount of motion, Color (Yes/No), Sound (Yes/No), Genre, Number of frames, Creation date (2006-09-04), Duration (1:90), Transcript.
- Spatial context:** Where.
- Temporal context:** When.
- Situational context:** Event.
- Social context:** Tags (Stephen Colbert lightsaber challenge green screen angry boy con), Rating (4.03), Comments (definitely too long that was incredibly original, but also very random and stupid. its amazing what people have done with this crap).
- Cognitive context:** Related, Novelty, % Quality.

Figure 2: *ContextMiner*: form for filling in various metadata and context fields. Many fields are automatically filled in.

3. At present, the *ContextMiner* system searches in OpenVideo collection to find the similar items. It performs search on four fields: *title*, *source*, *description*, and *keywords*. Thus, four independent searches are performed. *ContextMiner* uses Okapi retrieval model implemented in Lemur³ for retrieval and ranking.
4. Four independent rank lists are displayed with their IDs and titles (Figure 1).
5. The curator can select any record from any of the lists and view full details of it. The original record obtained from Prelinger collection is also shown for ready comparison.
6. Alternatively, the curator can issue a search query to specialized databases (e.g., Prelinger or OpenVideo), specialized websites (e.g., YouTube), or the entire Web. The results are brought back and presented as a rank list according to their relevance.
7. Selecting a result brings up a form with various metadata and context fields (Figure 2). *ContextMiner* fills in as much information as possible. The curator can add additional information and/or edit the information provided by *ContextMiner* as required.
8. When the form is submitted, a new record is added in the repository with an ID assigned automatically. The *ContextMiner* system also has an interface for viewing and updating the records in the repository.

³<http://www.lemurproject.org/>