

Fourth-Phase Digital Libraries: Pacing, Linking, Annotating and Citing in Multimedia Collections

J. Alfredo Sánchez

Libraries Division and CENTIA

Universidad de las Américas, Puebla, México

alfredo@mail.udlap.mx

J. Aníbal Arias

ICT - CENTIA

Universidad de las Américas, Puebla, México

anibal@ieee.org

Abstract

We discuss the implications of the use of current multimedia collections and posit that it is possible to build what we term fourth-phase digital libraries (4PDLs). In 4PDLs users can take advantage of both the powerful audiovisual channels and the proven practices developed for media such as text. We demonstrate how various technologies can be integrated to produce a 4PDL.

1. Introduction

The printed word has proven to be an effective means for supporting the generation and dissemination of new knowledge. We posit that multimedia digital libraries can be designed and built as integrative environments in which diverse media are interwoven to overcome their limitations and provide even more effective support for knowledge-intensive tasks. We refer to this sort of digital libraries as fourth-phase digital libraries (4PDLs).

2. The third phase

Simone [4] compares and contrasts the cognitive processes underlying the activities of *reading* and *viewing* documents to conclude that humankind has entered a “third phase” in the history of knowledge: The First Phase would have coincided with the invention of writing, which made it possible to persistently record the individual or collective human memory. A Second Phase started twenty centuries later with the invention of printing, which made books and knowledge accessible to a vast public, changing social and cultural life in profound ways. During the last twenty years, we moved into a Third Phase, in which the we have learned from reading much less than in the past and well-established learning practices and ways of knowing are being lost. Whereas writing allowed for more refined and complex forms of expression, knowledge expressed or acquired from audiovisual media tends to be less articulate and less subtle.

Seven traits of printed and audiovisual documents that impact the way they are used in knowledge access and generation are also discussed in [4]: (1) **Pace**: The pace of reading is determined by the reader; video authors “push” images at will towards the viewer. (2) **Corrigibility**: A reader may stop at any point in time to reflect on the text just read;

a viewer cannot do this easily. (3) **Encyclopedic references**: Reading allows users to stop and use complementary sources; doing that would disrupt the intended rhythm of video. (4) **Citability**: A text can be easily cited or even quoted; video segments do not exhibit this property. (5) **Conviviality**: Reading usually is a quiet, isolated, intent activity; viewing requires less attention and can be done collectively. (6) **Multi-sensoriality**: Reading is mono-sensorial, audiovisual materials reach the user through both sight and hearing. Finally, (7) **iconicity**: Images carry meaning even for the uninitiated; characters and sounds that represent words have zero iconicity.

3. Towards a fourth phase

If the Third Phase is characterized by the prevalence of an audiovisual culture, a Fourth Phase should integrate media, beyond just building multimedia collections, and make it possible for people to take advantage of the best of each of them. The digital substrate is rich and malleable enough to actually recreate existing media and give rise to improved interaction mechanisms. In order to continue to build on the ways of learning and knowing that have been developed for centuries, it is necessary and possible to reshape media such as audio and video so they acquire desirable features. Digital libraries of the fourth phase (4PDLs) should integrate technologies to produce knowledge centers in which reading and viewing occur seamlessly for all available media.

4. The VUDLA 4PDL

In order to explore the notion of a 4PDL we designed and prototyped VUDLA, a multimedia digital library that integrates image and speech processing, video streaming, information retrieval and multimedia databases. In VUDLA, users may create and explore digital video collections. More importantly, as illustrated in Figure 1, they can determine the pace of viewing, search for specific scenes based on their contents, add textual or graphical annotations or specify web links that will be attached to materials at specific points in time. Every time a video document is added, speech and image processing systems generate metadata (terms, color and texture indexes) that will be used to support video querying and viewing. VUDLA relies on information retrieval, video streaming and speech

processing components that have been developed by various groups in our research center. We are interested in developing video collections in Spanish and exploring mechanisms for retrieval that are particularly appropriate for them. A complete description of VUDLA is available at [3]. Relevant related projects are [1] and [5].

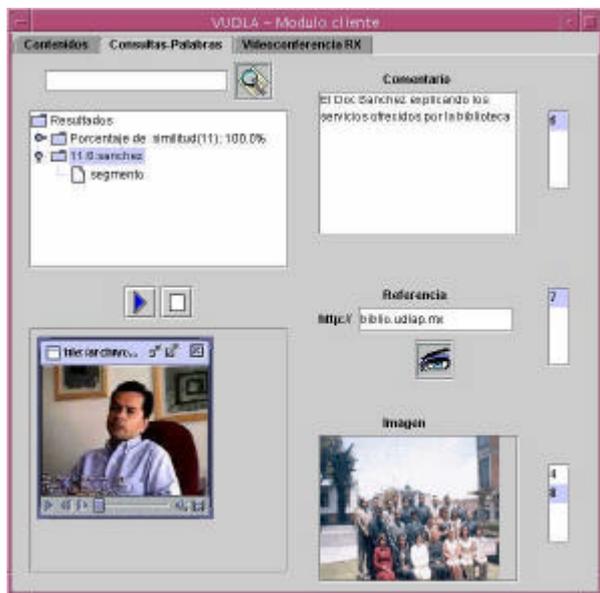


Figure 1. Main user interface of VUDLA.

5. Preliminary results

The VUDLA prototype has been undergoing tests in two major areas: functionality and usability. In the first area, we have successfully integrated diverse technologies. In the second area, we are demonstrating that VUDLA enables the use of digital video in such a way that users are able to perform tasks involving multimedia resources and overcoming the limitations referred to in Section 2. Our video collection is academic, in Spanish, and it includes interviews with researchers from various areas, lectures and software. We have been able to observe users interact with the new multimedia collections and we report here results with respect to relevant media traits introduced in Section 2, which guided our design:

Pace: In VUDLA, the interaction with resources in all available media is completely under the user's control. The user determines the pace at which temporal media are "played". **Corrigibility:** Users of VUDLA may stop at any point in time to reflect on resources being examined, regardless of their type. **Encyclopedic references:** Not only can complementary sources be used while examining multimedia resources in VUDLA, but those resources may be associated directly with a section of a document and available immediately for perusal. **Citability.** Content-based

search over digital video using text or images, as well as the possibility to point to any instant within a video segment make multimedia collections in VUDLA citable and quotable. **Conviviality, multi-sensoriality and iconicity:** A 4PDL such as VUDLA redefines the concepts of reading and viewing. In the digital library, reading may be an isolated activity, but the reader may also choose to be aware of the presence of other readers and decide to interact with them by taking advantage of the library's collaboration facilities. In addition to videoconferencing, other communication and awareness facilities are available in the overarching digital library [2]. Reading may involve video illustrations just like viewing may lead to textual annotations. In a way, the multi-sensorial and iconic properties of multimedia are imparted to text.

There is ample room for improving VUDLA. We plan to provide formatting and personalization options for textual annotations as well as *frame hotspots*, which should map specific frame regions to annotations or web links.

6. Conclusions

One of the concerns regarding the prevalence of video and the abandonment of reading is that video appears as a regression from the structured, analytical, sequential reasoning promoted by reading to more primitive, sensorial forms of intelligence. In a 4PDL, multimedia resources are experienced by the user's senses but associative thinking is also promoted. While still some of the mind-structuring properties of knowledge construction based primarily on reading may be at risk, there are sure new, more powerful cognitive processes triggered by 4PDLs.

Acknowledgments

This work is supported by Conacyt (Projects No. 35804-A and G33009-A). Thanks are also due to our colleagues at ICT and the Tlatoa and Videoconferencing Groups.

References

- [1] Lyu, M., Yau, E., Sze, S. 2002. A multilingual, multimodal digital video library system. *Proceedings of JCDL 2002*. 145-153.
- [2] Reyes-Farfán, N., Sánchez, J. A. 2003. Personal spaces in the context of OAI. *Proceedings of JCDL 2003* (this volume).
- [3] Sánchez, J. A., Arias, J. A. 2003. VUDLA: A fourth-phase digital library. Tech. Report ICT-2003-2. Laboratory of Interactive and Cooperative Technologies, Universidad de las Américas, Puebla, México. <http://ict.udlap.mx>.
- [4] Simone, R. *La Terza Fase*. Laterza. Rome, Italy. 2000
- [5] Wactlar, H. D., Kanade, T., Smith, M. A. Stevens, S. M. Intelligent access to digital video: Informedia Project. *IEEE Computer* 29, 3 (May). 1996. 46-52.