

Realization of Four Important Principles in Cross-Cultural Digital Library Development

J. Stephen Downie

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
217-351-5037

jdownie@uiuc.edu

ABSTRACT

In this White Paper¹, the importance of four digital library (DL) design features are outlined: 1) intensive and extensive community involvement; 2) multi-level annotation; 3) synchronization of media; and, 4) content-based relevance feedback. These four items have significant ramifications for those developing cross-cultural DLs. This paper explicates these ramifications in light of their social and technological impacts.

Categories and Subject Descriptors

D.3.5 [Information Storage and Retrieval]: Digital Libraries – Dissemination, System issues, User issues..

General Terms

Design, Human Factors, Verification.

Keywords

Cross-cultural Digital Libraries; Multimedia Retrieval; Participatory Action Research; Digital Library Design and Evaluation.

1. INTRODUCTION

This author is currently part of a multinational, multidisciplinary research team at the University of Illinois at Urbana-Champaign (UIUC). This group is currently laying the foundations for a “Digital Library for Cultural Heritage Preservation (DLCHP)”. Headed by Professor Narendra Ahuja at UIUC’s Beckman Institute, the research team at Illinois has representatives from anthropology (Native American studies, linguistics), computer science (data mining, artificial intelligence, multimedia databases), engineering (audio and signal processing, machine vision), NCSA (data management, metadata standards), and library science (music retrieval, learning technologies, digital libraries). UIUC’s international research partners are drawn from the International Institute of Information Technology, Hyderabad (IIITH), India along with researchers from Central University of Hyderabad, Jawaharlal Nehru University, New Delhi, and Rashtriya Sanskrit Vidyapeeth.

The primary goal of the DLCHP project is to establish a large-scale DL testbed designed to facilitate and enhance the preservation of endangered cultural knowledge. As our test cases we are working within two distinct cultural milieus: 1) North American First Nations; and, 2) India. From each of these two

sets² of cultures, we are striving to build tools to capture the endangered knowledge pertaining to their respective Movement Arts (i.e., sign languages, dance). These two culture sets were chosen because both are undergoing considerable loss of cultural knowledge through such means as attrition (i.e., the death of traditional knowledge holders) and community assimilation of Western cultural values. The Movement Arts component of the project was chosen because both cultural groups have long traditions of cultural communication via the Movement Arts, and because the Movement Arts represent interesting multimedia (i.e., audio, video, musical, textual, notational, etc.) challenges that must be addressed before a robust DL can be constructed.

2. FIRST PRINCIPLES

The purpose of this White Paper is to highlight some first principles and key features of the DLCHP project. I see our project as being constantly informed by the following four central ideas:

1. the importance of intensive and extensive community involvement;
2. the importance of multi-level annotation;
3. the importance of synchronization of media; and,
4. the importance of content-based relevance feedback.

The issues raised, and the solutions proposed, in the achievement of these principles will be of interest to those creating cross-cultural DL systems. Let us now explicate each in turn.

2.1 Intensive and Extensive Community Involvement

Meaningful digital libraries should be more than glorified hypertextual encyclopedias. Imbedded in the “encyclopedic model” of DL development is a top-down imposition of cultural values. That is, experts gather, filter, organize, validate and then fix that which *they* believe to be important. In a sense, the encyclopedic model fossilizes the cultural objects it encapsulates.

Cultures are living, and thus, dynamic systems. If ones goal is the preservation of endangered cultural knowledge, then one must come up with a means of capturing the inherent dynamism that gives life to culture. This dynamism includes variations of presentation, local and regional cultural dialects, cultural discourse and disagreements about both form and meaning. Furthermore, this dynamism is generated by all those involved—in any way—with the culture, regardless of age, social status, or

¹ Originally presented at the JCDL 03 Workshop: Cross-Cultural Usability for Digital Libraries.

² “Sets” is used deliberately as each broad milieu comprises many distinct nations, cultures and traditions.

education. The “community” just defined includes schoolchildren, students, scholars, experts, elders, practitioners (e.g., dancers, musicians, storytellers, etc.), and non-practitioners—though, perhaps, audience members—of the gamut of ideas, feelings, gestures and artifacts that make up cultural knowledge.

So, rather than fossilize the culture heritage we hope to preserve, we have repudiated the “encyclopedic model” of DL development for one that is intrinsically community-based. To this end, we are incorporating the technologies of the Inquiry Page project (<http://inquiry.uiuc.edu/>) [3]. “Inquiry is an approach to learning that involves a process of exploring the natural or material world, that leads to asking questions and making discoveries in the search for new understandings” [17]. Users of all levels engage the *Inquiry Unit Generator* to produce and share online lesson plans, project outlines, project and research reports, and so on [4]. Through this mechanism, all members of the community can participate in the building of a dynamic digital repository of cultural knowledge.

Closely related to our incorporation of the Inquiry Page technologies is the evaluation framework the Inquiry Page researchers have integrated into their design and evaluation paradigm. This evaluative framework [2]

represents a reconsideration of traditional approaches to DL design and evaluation research [1]. To engage seriously with the social practice disenfranchised users, we are incorporating ideas and techniques from two domains that are not often folded into technology evaluations. One domain is *participatory action research* (PAR), which claims social practice as its fundamental object of study and explicitly pursues agenda focused on improving conditions for disenfranchised members of society [21, 27]. The other domain encompasses *inquiry-based learning* [6]. Here we find that framing usability research as a collaborative “community inquiry” process helps in integrating the knowledge and views of diverse participants in the development of digital libraries. The notion “community inquiry” frames DL research as a democratic process in which everyone can learn from each other [5].

To summarize: In the same sense that culture heritage *is* the living dynamic community that it envelopes, we are conceiving the “users” of our digital library as its “creators” *and* “users”. Our role in this endeavour is simply to develop the necessary set of tools and practices to make this possible.

2.2 Multi-level Annotation

One part of the necessary tool set involves the creation of annotation tools that can be used by all members of the community. This implies that the tools must simultaneously be capable of supporting both the creation and extraction of cultural knowledge in a wide variety of contexts. For example, the tools developed must allow for the audio/video digitization and subsequent annotation of, say, a story told in Native American sign language, along with its Native American verbalization or explication. A digital video and audio representation of an Indian classical dance could also be the object of annotation. To these basic objects, users/creators must be able to add personal and scholarly interpretations, transcriptions (in both native and translated scripts) or perhaps variations on the story in a variety of media (i.e., audio, video, text). For example, a possible annotation

might be a user’s/creator’s digitization of his/her own competing version of the story or dance.

It is important to note that we envision some annotation sets to be “very scholarly” and others to be, in the opinions of some, “silly”, “misinformed” or even simply “wrong”. I believe that this situation is not detrimental to the viability of the DL for three reasons. First, in a digital environment, annotations are non-destructive. That is, if one conceives of the aforementioned dance video, for example, as a kind of “base object”, the attachment of one or more annotations to the base object in no way destroys the object itself. Second, annotations themselves can also be considered to be a kind of “base object”. That is, annotations, once entered into the system, become the object of other annotations that can add further interpretations, refutations, corrections, and so on. Third, and most importantly, the act of annotating annotations is, in fact, a kind of cultural discourse and as such, keeps cultural knowledge alive through its inherent dynamism.

2.3 Synchronization of Media

Let us revisit the example of Indian traditional dance. A more or less complete base representation of this in the DL would involve:

1. video of the dance itself
2. audio of the music
3. audio of storyline, interpretation(s) and explication(s)
4. text of storyline, interpretation(s) and explication(s)
5. translations of #3 and #5
6. transcriptions of the music in various symbolic forms (i.e., Indian and Western versions)
7. transcriptions of the dance in symbolic form(s) (e.g., Labanotation)
8. computer generated version(s) based upon #7
9. and so on.

In this example, the “meaning” of the dance resides in the sum of its constituent parts. Since meaning is at the heart of cultural life, it is important that the DL capture and preserve this meaning as fully as possible. Therefore, it is our goal to develop a set of tools to allow for the fine-grained synchronization of these disparate representations for each digital instance of the Movement Arts in the collection

To this end, we will be drawing upon on research at UIUC in music information retrieval (MIR) [8, 9; 11, 12, 14], speech recognition in noisy backgrounds [18], and multimodal dialog systems [20]. Sophisticated machine-learning techniques involving such technologies as Hidden Markov modeling and advanced signal processing will be explored as possible mechanisms for synchronizing audio, visual and symbolic events. For example, there have been significant breakthroughs in recent years in the domain of beat and rhythm detection from digital audio sources (e.g., [7, 13, 25]). We plan on utilizing these breakthroughs as part of our synchronization process. Of particular interest to us, is the work being conducted in Barcelona (Music Technology Group) and Paris (IRCAM) by Guyon and Meudic [15] on low-level *tractus* detection (i.e., fundamental pulse detection) and higher-level meter detection (i.e., grouping low-level detected pulses into rhythmic units, such as $\frac{3}{4}$ or $\frac{2}{4}$ meters, etc.). While ideas of persistent meter are not directly applicable to many non-Western musics (e.g., Indian and Native American musics), we

believe, however, that we can exploit the notion of *tractus* detection to help us establish synchronization points in the audio source files. The identification and subsequent marking of the *tractus* points in the audio source files will allow us to make fine-grained linkages between audio, video, symbolic and metadata elements.

I believe work on the synchronization of representations will reward the significant effort that will be necessary to make it a reality. First, the synchronization process in effect generates a set of extremely powerful multi-dimensional indexes. That is, one representation becomes the entry, or access point, for any of the other representations. For example, the Labanotation becomes the entry to a specific video segment. Likewise, a music audio segment provides access to the corresponding event in the Western score, and so on. Second, very few, if any, members of the user/creator community will be conversant in all possible representations of a given artifact. That is, some might be able to interact with the storyline in, say, Hindi, the audio version of the music and its Western score representations, but be totally ignorant of its Labanotation and its English language interpretation. Because synchronization affords so many different ways to approach any given artifact, creators/users will be able to exploit their literacies in some representations to overcome their illiteracies in others. Third, users/creators will be able to annotate a widest possible range of depth. That is, they can point to entire pieces, or to select passages and/or events both within and across different artifacts. This pointing, linking and crossing of representational and artifactual boundaries is the stuff of new ideas and knowledge being born.

2.4 Content-Based Relevance Feedback

Content-based relevance feedback (CBRF) extends relevance feedback (RF) to include relevance judgments based on features detected in the content of multimedia objects such as colors in images, melodies in music, structures in architectural graphics, and choreographic units in dance. RF techniques have a long history of successful deployment in the traditional text-based environments [24]. Popular search engines now routinely include RF; google.com, for example, includes a “similar pages” retrieval option. CBRF tools are particularly effective, perhaps even necessary [8], in a multimedia DLs because general users do not typically possess the domain-specific (usually text-based) vocabularies to express their multimedia information needs. In the domains of image (primarily photographic) and video retrieval, great CBRF advances are being—and continue to be—made with [16, 22, 23, 26] representing but a small subset of this growing body of literature. In the domains of dance, movement arts and music, there exist, however, no extant mechanisms for achieving CBRF. One reason for the lack of CBRF techniques in these domains is the problem of *disambiguation*: what set of features, components or facets of the object(s) is the user deeming to be relevant and which irrelevant? In music, for example, is it the tempo, melodic line, orchestration, lyrics and/or rhythm of a given work upon which the user is basing the relevance assessment? Again, we will be exploring a variety of machine learning and signal processing techniques to determine which provide the most useful sets of disambiguated, medium-specific, features. Once each medium has been disambiguated, CBRF interfaces will be developed that will allow users to see and/or hear the constituent components of the object(s) of interest so

further, more refined, content-based query specifications can be submitted to the appropriate retrieval systems.

We further propose to explore CBRF possibilities beyond the intrinsic features found *within* each media type to also include extrinsic similarities *across* media types. Because we will be performing extensive synchronization of media objects, it should, in theory, be possible to provide the users/creators with a CBRF interface that presents the option to retrieve “similar” (in some sense) objects in different media formats, perhaps even different domains. That is, users/creators might be prompted to consider linkages between objects that are not initially obvious. For example, three pieces of music might have no melodic sequences, no rhythms, no harmonies nor lyrics in common, but still might be deemed to be in some sense “similar” and “relevant” because all three have had “similar” dance gestures associated with them.

To further explicate the importance and significance of creating powerful CBRF interfaces for these media, consider the following two “use scenarios”.

Use scenario #1: Imagine that a scholar wishes to explore notions of cross-cultural interpretations of gesture. This scholar has been presented by the interface with a video in which a Native American dancer makes a wide circular sweeping motion with outstretched arms. Using the CBRF tools we developed, the scholar isolates the specific segment of interest and submits the selected extract as the query to the system. The library’s retrieval tools then gather up all “similar” gestures and presents the resultant set to the scholar for continued investigation. Because of the extensive synchronization of information that has taken place, the scholar then further refines the query by asking the system to cluster the retrieved objects based upon some other similarity factors, such as textual metadata (i.e., type of dance: celebratory, funereal, etc.), tempo of underlying music, instrumentation, melodic shape, etc. The scholar then presents his/her findings via the system’s Inquiry Page, complete with pointers to the specific example segments and annotations of their significance. Others in the community would then comment upon the scholars findings via annotations of their own that also contain fine-grained, multimedia example segments to illustrate their opinions.

Use scenario #2: In a similar way, imagine that a grade school student is interested in exploring notions of cross-cultural “affect” (i.e., how emotional elements are expressed). While browsing the DL, the student identifies a dance segment from India that is accompanied by music that the student finds particularly “sad.” The notion of “sadness” is, of course, highly subjective. Notwithstanding its subjectivity, however, it is a very real and perfectly valid way of experiencing music and dance (see [8, 10, 19]). To begin her explorations into the cross-cultural expression of “sadness”, the student would engage in an interactive session with the CBRF interface to identify those features within the music that give rise to its perceived “sadness”. Perhaps, it is the tempo, the rhythm or the harmonies present, or some user-specified mixture of all of these. Once the student has refined the set of features that, for her, embody “sadness”, she can then submit the feature set to the system to examine whether or not these features do indeed retrieve other examples of “sad” music and dance. Again, her findings would then be gathered up as a set of annotations and presented to the community for discussion and debate.

3. CONCLUDING REMARKS

There are two primary themes running through this White Paper that those involved in the cross-cultural DL research might wish to consider. One is social, the other technological. The first—on the social side—is the notion of extensive and intensive community involvement in all aspects of their projects including the creation and evaluation of the content and the systems themselves. That is, the reconceptualization of “users” as iterative “creators *and* users”. The second—on the technological side—is the deployment of sophisticated content-based tools to enable the annotation, synchronization and retrieval of the system’s multimedia content in an effort to capture and exploit the important interrelationships within, and between, the DL’s collection of cultural objects.

4. REFERENCES

- [1] Bishop, A. P., I. Bazzell, B. Mehra, and C. Smith. Afya: Social and digital technologies that reach across the digital divide. *First Monday* 6, 2001.
- [2] Bishop, A. P., and B. C. Bruce. Usability research as participative inquiry. In *JCDL workshop on Usability of Digital Libraries '02*, 2002.
- [3] Bruce, B. C., A. P. Bishop, P. B. Heidorn, K. J. Lunsford, S. Poulakos, and M. Won. The inquiry page: Learning with digital libraries. In *Joint Conference on Digital Libraries 03*. In press, 2003.
- [4] Bruce, B. C., A. P. Bishop, and J. Robins. The inquiry page: A collaboratory for curricular innovation. In *Computer Support for Collaborative Learning: Foundations for a CSCL Community*, p. 746, Hillsdale, NJ: Lawrence Erlbaum, 2002.
- [5] Bruce, C., and A. P. Bishop. Using the web to support inquiry-based literacy development. *Journal of Adolescent and Adult Literacy* 45, 2002.
- [6] Dewey, J. *Experience and education*. New York: MacMillan, 1938.
- [7] Dixon, S. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research* 30: 39-58, 2001.
- [8] Downie, J.S., and S. J. Cunningham. Toward a theory of music information retrieval queries: System design implications. In *3rd International Conference on Music Information Retrieval*: 299-300, 2002.
- [9] Downie, J.S. Music information retrieval. *Annual Review of Information Science and Technology* 37: 295-340, 2003.
- [10] Downie, J. S.. The MusiFind musical information retrieval project, phase II: User assessment survey. In *22nd Conference of the Canadian Association for Information Science*: 149-166, 1994.
- [11] Downie, J. S. *The Music Information Retrieval/Music Digital Library Evaluation Project White Paper Collection, Edition #2*. Champaign, IL: Graduate School of Library and Information Science, UIUC, 2002.
- [12] Downie, J. S. Music information retrieval annotated bibliography website project, phase I. In *2nd International Symposium on Music Information Retrieval*: 5-7, 2001.
- [13] Foote, J., M. Cooper, and U. Nam. Audio retrieval by rhythmic similarity. In *3rd International Conference on Music Information Retrieval*: 265-266, 2002.
- [14] Futrelle, J., and J. S. Downie. Interdisciplinary communities and research issues in music information retrieval. In *3rd International Conference on Music Information Retrieval*: 215-221, 2002.
- [15] Gouyon, F., and B. Meudic. Towards rhythmic content processing of musical signals: Fostering complementary approaches. *Journal of New Music Research*: In press, 2003.
- [16] Heidorn, P. B. Image retrieval as linguistic and nonlinguistic visual model matching. *Library Trends* 48 (2):303-325, 1999.
- [17] *Institute for Inquiry. Inquiry descriptions: Inquiry forum. Exploratorium Institute for Inquiry Resources*, 1996. <http://www.exploratorium.edu/IFI/resources/inquirydesc.html>.
- [18] Jing, Z., and M. Hasegawa-Johnson. Auditory-modeling inspired methods of feature extraction for robust automatic speech recognition. In *Proceedings of ICASSP '02*, 2002.
- [19] Kim, J.-Y., and N. J. Belkin. Categories of music description and search terms and phrases used by non-music experts. In *3rd International Conference on Music Information Retrieval*: 209-214, 2002.
- [20] Naphade, M., A. Garg, and T. Huang. Audio-visual event detection using duration dependent input output Markov models. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*: 39-43, 2001.
- [21] Reardon, K. M. Participatory action research as service learning. In R. A. Rhoads and J. P. F. Howard, editors, *Academic service learning: A pedagogy of action and reflection*: 57-64. San Francisco: Jossey-Bass, 1998.
- [22] Rui, Y., T. S. Huang, S. Mehrotra, and M. Ortega. A relevance feedback architecture in content-based multimedia information retrieval systems. In *IEEE Workshop on Content-based Access of Image and Video Libraries*: 82-89, 1997.
- [23] Rui, Y., T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Trans. on Circuits and Systems for Video Technology* 8 (5): 644-655, 1998.
- [24] Salton, G. *Automatic Information Organization and Retrieval*. New York: McGraw-Hill, 1968.
- [25] Scheirer, D. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America* 103:588-601, 1998.
- [26] Squire, W. Muller, and H. Muller. Relevance feedback and term weighting schemes for content-based image retrieval. *Visual Information and Information Systems*: 549-556, 1999.
- [27] Whyte, W. F., editor. *Participatory action research*. Newbury Park: Sage, 1999.