

# A Visual Dictionary for an Extinct Language

Kyle Williams, Sanvir Manilal, Lebogang Molwantoa, and Hussein Suleman

Department of Computer Science  
University of Cape Town  
Private Bag X3, Rondebosch, 7701  
{kwilliams, smanilal, lmolwantoa, hussein}@cs.uct.ac.za

**Abstract.** Cultural heritage artefacts are often digitised in order to allow for them to be easily accessed by researchers and scholars. In the case of the Bleek and Lloyd dictionary of the |xam Bushman language, 14000 pages were digitised. These pages could not be transcribed, however, because the language and script are both extinct. A custom digital library system was therefore created to manage and provide access to this collection as a purely “visual dictionary”. Results from user testing showed that users found the system to be interesting, simple, efficient and informative.

**Key words:** Digital preservation, repository, dictionary, Web interfaces

## 1 Introduction

Cultural heritage in South Africa, like in most developing countries, is subject to degradation and inaccessibility. There is a clear need to preserve cultural heritage and make it accessible for the future. A snapshot of South African heritage would be incomplete without mentioning the Bushman people - one of the oldest known ethnic groups in the world. With the rapid influence of Western culture, there are now only a handful of these Bushmen people left in South Africa [1]. It is estimated that in a few years the entire generation of Bushman will have passed on, thereby creating a need to preserve whatever artefacts and knowledge exist from the Bushmen people in order to allow for them to be accessed in the future.

The Bleek and Lloyd Collection is a collection of artefacts that document the life, language and culture of the Bushman people of Southern Africa. The collection is primarily made up of notebooks that contain Bushman stories, narratives and artwork. Included in this collection is a dictionary that contains English words and their corresponding |xam Bushman language translations. This dictionary can be used to assist researchers in understanding and interpreting the |xam Bushman language. However, the script used to represent the |xam language can not be represented using modern data encoding techniques. An accessible and preservable archive for a “visual dictionary,” where words are represented by images and definitions are extracted from viewing the images, known as the Bushman OnLine Dictionary (BOLD), was therefore built. Using this visual dictionary, users are able to browse, search and interact with the |xam words based on their English translations.

## 2 Related Work

The Native Languages of the Americas is a non-profit organisation that works towards preserving Native American languages by making use of Web technology [2]. The Nuer Field Notes Project is an attempt to preserve and make available a set of linguistic field notes recorded by Eleanor Vandevort, a missionary in South Sudan between 1949 and 1963, using modern data encoding techniques [3]. In addition to these specific attempts at preserving languages, there also are a number of general preservation attempts, such as the Contemporary African Music and Art Archive [4] and the Amarius archive [5]. In 2007, Suleman [1] devised an XML-centric approach to manage the notebooks and artwork in Bleek and Lloyd Collection, showing the XML-centric approach to be more efficient than the traditional database model. All of the above-mentioned systems allow users to readily access the collection of digital cultural artefacts. However, none of these systems preserve and make accessible a dictionary, which cannot be represented using modern data encoding techniques, as a live reference for researchers and other people who access the archives.

## 3 Design and Implementation

The BOLD system was built using Fedora Commons as an underlying digital repository system. Figure 1 shows the Web interface with which users interact with the visual dictionary. The Web interface shows the three classes of digital objects that make up the collection: envelopes, slips and inserts. For every word in the dictionary there is one or more envelopes, and each envelope contains one slip and one or more inserts. The Xam Bushman words and their corresponding English translations are written on the inserts. The following means of interaction with the visual dictionary are provided by the interface:



Fig. 1. The website users use to interact with the collection

**Core services:** Browsing and searching the visual dictionary based on the English translations of |xam words is the core functionality provided by the Web interface. The results of browsing and searching the visual dictionary should be such that they can assist researchers in understanding and interpreting the |xam language, even though the language cannot be represented using modern data encoding techniques. Users are able to browse the words in the dictionary by their English initial letters, or via a scrollable list. Users also can make use of an AJAX live search to find specific words in the dictionary.

**Enhancement for understanding:** There are a number of services that attempt to assist the user in their understanding of the collection. The first of these services comes in the form of links to the notebooks of the Bleek and Lloyd Collection for each word displayed, thereby assisting in contextualising words. In addition to this, the definition of a word appears below it when it is clicked on and spelling correction suggestion takes place when a search is performed.

**Enhancement for experience:** To improve the experience of the user when interacting with the collection, a history of all the words that have been searched for is stored in a session cookie and there is an image zoom function that zooms an image to the size of the browser viewport when the image is clicked on, thereby allowing for closer inspection of the image.

Having briefly discussed the design and implementation of the BOLD visual dictionary, the next section will present an evaluation of the system.

## 4 Evaluation

The visual dictionary was evaluated using usability testing in which 22 users participated. User evaluation was conducted on a one-on-one basis in which users first answered a pre-test questionnaire. Users then were given an introduction to the Bleek and Lloyd Collection to read, after which they were asked to perform three tasks. They were then given free reign to experiment with the system and, lastly, were asked to answer a post-test questionnaire. The key findings from the evaluation are summarised here.

81% of users were 100% satisfied with the way in which they could search for a single word and 60% of users were 100% satisfied with the way they could search for multiple words. 72% of users were 100% satisfied with the way in which they could browse the system. Users noted that when searching for multiple words the system became slow and was not intuitive. However, overall the results show that the core services of searching and browsing were well received and resulted in high user satisfaction.

Users felt that the system was easy to navigate and that there was adequate information to help with navigation. Users also found the dictionary to be generally informative and useful and felt that they could find information quickly. Users felt that their errors were easy to correct while 17 users agreed that the system furthered their knowledge in African Cultural Heritage. All users agreed that the system is useful for researching the |xam language. Users found the enhancement for understanding useful and were surprised by the apparent com-

plexity of the |xam language script compared to English. Two users expressed that they felt like they did not know what they could do with the system. All but 2 users felt that the system was slow.

The most appealing aspects of the visual dictionary were: the thumbnail view when browsing; the multiple ways of searching and the AJAX search; the simple layout, lack of clutter, pretty design, ease of use and easy navigation; the quick access to resources; high resolution images; rich information; and links to the Bleek and Lloyd Collection. The least appealing aspects of the visual dictionary were: that it was slow; the links to the Bleek and Lloyd notebooks were not intuitive; the interface was too simple; and the correction of misspelled words to suggested words that were not in the dictionary.

The most prominent words used to describe the visual dictionary were: interesting, simple, effective and informative. Evaluation showed that users were generally happy with the system and felt as if they could use it in meaningful ways. In this sense, the BOLD Project shows that it has potential in meeting its goal of assisting researchers in interpreting and understanding the |xam Bushman language.

## 5 Conclusions

The Bushman OnLine Dictionary (BOLD) is a cultural heritage archive system for providing access to a visual dictionary for the |xam Bushman language, which forms part of the Bleek and Lloyd Collection. The system was built to assist researchers and scholars in understanding and interpreting |xam Bushman texts. The system contains core services for browsing and searching the archive, as well as services for enhancing user understanding and enhancing user experience. Evaluation showed that users had a positive experience using the system and were pleasantly surprised by many of its features. The BOLD Project is a first step in building a system that allows for meaningful interaction with a visual dictionary and has set the stage for future work to be done in this area.

## References

1. Suleman, H.: Digital Libraries Without Databases: The Bleek and Lloyd Collection. In: Kovacs, Laszlo, Norbert Fuhr and Carlo Meghini (eds). *Proceedings of Research and Advanced Technology for Digital Libraries, 11th European Conference (ECDL 2007)*. LNCS, vol. 4657, pp. 392-403. Springer, Berlin/Heidelberg (2007).
2. Native Languages of the Americas: Preserving and promoting American Indian languages, <http://www.native-languages.org/>
3. Nuer Field Notes, <http://www.dlib.indiana.edu/collections/nuer/>
4. Marsden, G., Malan, K., and Blake, E.: Using digital technology to access and store African art. In: *CHI '02 extended abstracts on Human factors in computing systems (CHI '02)*, pp. 258-259. ACM, New York, NY, USA (2002).
5. Doumat, R., Egyed-Zsigmond, E., Pinon, J., and Csiszar, E.: Online ancient documents: Armarius. In: *Proceeding of the Eighth ACM Symposium on Document Engineering (DocEng '08)*, pp. 127-130. ACM, New York, NY, USA (2008).