

The Humboldt Digital Library: Exploring Innovative Structures

Prof. Dr. rer. nat. Detlev Doherr

Badstr. 24
77652 Offenburg
Tel. 0781 205-281
E-Mail:
doherr@fh-offenburg.de

IV.18 The Humboldt Digital Library: Exploring Innovative Structures

Prof. Dr. rer. nat. Detlev Doherr

Summary

Humboldt's travel narratives suggest innovative features in a digital library. His work is relevant today, and a digital library has to explore the originality of his narratives and the significance of his diverse data. Because his holistic vision of the nature cannot be presented in a proper way in XML-based documents, advanced information technologies are necessary to provide access to Humboldt's legacy. An Internet-based information system can integrate Web services into the Open Archives Initiative for digital libraries. This system will create a global network of information that contains multimedial and multilingual data, together with interactive maps.

Introduction

Alexander von Humboldt was a scientist whose precision in reporting of his observations and findings make him unique and still relevant for science today. His unorthodox combination of text, precise data, and images reflects his search for details, totality and global interconnectedness. The challenge is to open the path to a global information network that reflects the way Humboldt worked and understood nature. He apologized for his unsystematic organization in the "Researches Concerning the Institutions & Monuments" (originally the *Vues des Cordillères*) by writing that it "might have been preferable to have

1953: Geboren 3.11. in Göttingen

1960–1972: Humanistisches Gymnasium in Göttingen, Abitur

1973–1980: Geowissenschaftliches Studium an der Georg-August-
Uni Göttingen

1980–1983: Diplom, Promotion als **Dr. rer. nat.** nach DAAD-
Stipendium an der Uni Göttingen

1983–1990: Projektleiter in deutschem Bergbauunternehmen,
dort ab 1986 Referatsleiter für die Entwicklung eines Geoin-
formationssystems für den Bereich Bergbau-Geologie in
Partnerschaft mit IBM

seit 1990: Professor an der Hochschule Offenburg für Bereich
Umweltinformatik

seit 1993: Wissenschaftlicher Leiter des Hochschulrechenzentrums

seit 1993: Leiter des Steinbeis-Transferzentrums Umweltinformatik, Offenburg

Sonstige Tätigkeiten: Adjunct Research Associate im Hall Center, University of Kansas,
Lawrence, USA



arranged the materials in a geographical order; but the difficulty of collecting has prevented me from following this method."

The standards for digital libraries are defined by the Open Archives Initiative [1], which supports projects with a concern about interoperability. It provides a focus for the international implementation of the metadata harvesting protocol (OAI-PMH), which is based on metadata standards for global search and exchange of data, including specifications for library, formats and community [2].

Obviously there are many digital documents available on the Internet about Humboldt and his legacy, some of them with open access [See: Gallica: <http://gallica.bnf.fr/>]. But all of them show the same inherent system insufficiency: the fixed document type that allows for the identification of the document and a search for keywords. They do not provide interactive and dynamic linkage to other documents, integrating multimedial data like text in different languages, images, and maps with Web services and data mining.

Vision of a Global Information System

Today we have a unique opportunity to organize digital documents on the Web by using Internet technologies and information systems. This can be accomplished for a digital library with document services and global metadata standards. A very practical example of the limits of digital libraries is the question how we could compare and correlate text paragraphs from Humboldt's volumes in different languages. This leads to the tech-

nical concept for a digital archive as an information system. This system provides an interconnected set of information resources, including all of the means and mechanisms for data receipt, processing, storage, retrieval, and analysis. Information Systems can be designed for storage and dissemination of a variety of data products—including primary data sets and both intermediate and final analyses—and for an interface providing connections to external computers, external data banks, and system users.

Another aspect of our project is the role of the geographic information in Humboldt's travel narrative. The rich content of his work reveals the obvious advantages of a geographical information system (GIS) and suggests an innovative feature for our archive. A GIS can illustrate the spatial connections between Humboldt's travels and observations and open the navigation through the database.

Concepts for the Server Platform

With the digital library project the platform should be defined to create a network based on technology with deep semantic interoperability through use of metadata and object definitions [3] but also for text search, linkage and comparison with other text paragraphs in the same or different languages. Using the OAI Dublin core standards [1] the archive opens interoperability with other global digital libraries. In addition, a distributed data base environment is to be established on the Internet, using state of the art database servers with relational and object oriented data structures in connection with the Web server. Our

information system can be described mainly as a relational database, based on the SQL standards. This database contains object oriented information, which reflects the intrinsic representation and storage of real world entities no matter of the data type like text, image, and map. The object-oriented programming develops a software system that is modelled as a set of objects that interact with each other. The challenge is not simply the collection of data from 29 volumes in four languages. We need to show the interconnectedness of the content of Humboldt's volumes and the different object types of information, such as the precise descriptions, drawings, scientific reports, holistic visions, texts in four languages, images, geographic and thematic maps, cross sections, and travel routes.

Prototype

How can the user gain access to Humboldt's world? We have chosen his "plant geography" with the famous visual image of the Chimborazo mountain (see fig. IV.18-1), a masterpiece of a pioneer-

ing graphic representation that shows botanical distribution in relationship to altitude, as an example for the digital test bed [4]. In accordance with standards established by the Open Archives Initiative, the digitized texts become accessible documents in a digital library. In addition, organized according to volume, chapter, page, and paragraph, the texts will become part of the information system.

To use real data we digitized English texts of the "Relation Historique" and the *Vues des Cordillères* by text recognition at the Max-Kade Center in Lawrence, Kansas. The formatted text were inserted as a PDF document to the digital library Eprints in Offenburg, Germany, where we use the standards of the OAI. To provide a view to the original scan for further research the scanned pages from each volume were stored to the database and linked to the related text paragraphs.

For the presentation of geographic or thematic maps an interactive map server on the Internet is used. This map server

contains digitized information and provides an interface to the text archive. We developed access to texts with the Chimborazo map and tested navigation features. We defined map objects that have attributes such as themes, names, or notes. This experiment revealed the usefulness of SQL (Structured Query Language) structures for the correlation of objects inside the maps and the harvesting of information from the digital archive.

The interactive map allows zoom functions, presents some points of interest, such as geographic localities from plants, and provides further functions to search information. We defined the link for some localities as name of plants from Chimborazo in the map and developed an entry point and search functions in the database. The result can be presented as a map, image, or a text paragraph in the Web page (see fig. IV.18-2). The connection between the text and spatial data will be established by interactive maps containing graphic objects with attributes. The focus of our work will be on the identification of text attributes in



Fig. IV.18-1: Image of the Chimborazo in the Web portal, representing an interactive map with further navigation features and data mining beyond the image

the database and the definition of hyperlinks to access the archive.

The complex graphic display of the Chimborazo image and its plant geography offers many possibilities of innovative design within the digital library. It is a model presentation of Humboldt's concept of interconnectedness. Thus, the Chimborazo graphic can serve as an example for the way in which our system can access information within Humboldt's writings and then become linked to data bases outside it.

References

- [1] <http://www.openarchives.org> -> Open Archives Initiative
- [2] Thomas Stacker: „Digitalisierung alter Drucke – Auswahl, Technik, Metadaten“.- Herzog August Bibliothek Wolfenbuttel, Wolfenbuttel, 2003; url: www.goethe.de/oe/mos/ts.doc
- [3] see: Hsinchun Chen (1999): Semantic research for Digital Libraries, Vol. 5, D-Lib Magazine, ISSN 1082-9873, 1999
- [4] <http://www.avhumboldt.net> -> Prototyp Humboldt Digital Library

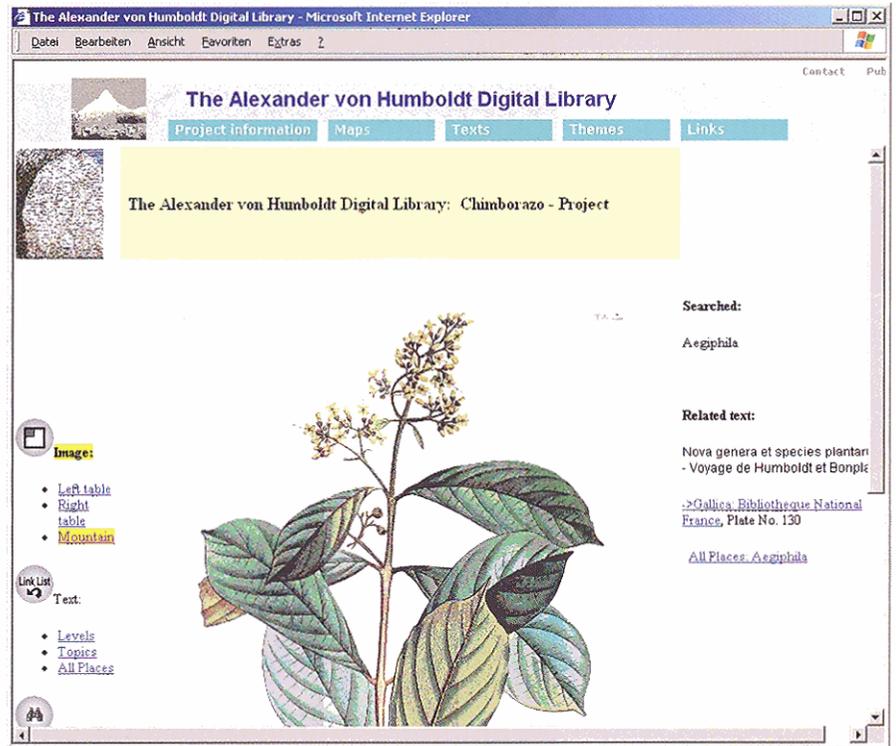


Fig. IV.18-2: The plant images and the texts with their detailed descriptions are easily accessible by a click on the plant names in the Chimborazo graphic.