

A Lightweight Digital Library Framework

Lighton Phiri

Supervisor: A/Prof Hussein Suleman

October 3, 2011




Introduction

What is a digital library system?

- Digital Library System (DLS)
 - Information system
 - Collection of digital objects
 - Storage, management, access and preservation of digital objects
- Architectural Design
 - User interface
 - Services
 - Repository

Institution repositories

- UCT CS Document Archive



UCT CS Research Document Archive [Home](#) || [About](#) || [Browse](#) || [Search](#) || [Register](#) || [User Area](#) || [Help](#)

Welcome to UCT CS Research Document Archive



Welcome to the [UCT Computer Science](#) Research Document Archive, which archives and makes accessible documents that are products and by-products of research in the department.

Search the Titles, Authors, Abstracts and Keywords :

[Browse](#)

Browse the archive by [Subject](#), [Year](#), [Lab](#) or [Type](#).

[Latest Additions](#)

View items added to the archive in the past week.

(<http://pubs.cs.uct.ac.za/>)

Cultural heritage collections

- Bleek and Lloyd Digital Collection

THE DIGITAL BLEEK AND LLOYD **HOME**

The Digital Bleek and Lloyd includes scans of every page of the 110 Lucy Lloyd !xam notebooks, 17 Lloyd (mostly !kun notebooks and 28 Wilhelm Bleek !xam notebooks. It also includes Jemima Bleek's solitary Korana and !kun notebook and four Lloyd Korana notebooks in the Maingard collection of the [Library](#) at the University of South Africa, as well as Dorothea Bleek's 32 notebooks. All the drawings and watercolours made by Jhan#kass'o, Dialkwain, Tamme, Juma, !nanni and Da are also in the digital collection. Currently the Bleek and Lloyd dictionaries are being digitised and will become available online in 2011. This represents the coming together of an archive dispersed across four institutions: [the National Library](#), [Iziko South African Museum](#), [Unisa](#) and the [University of Cape Town](#). The digital archive includes a 280 000-word searchable index, cross-referenced and including notes and summaries for each of the stories listed. Notes in italics are direct quotes from the reports of Bleek and Lloyd in which they detailed the progress of their research. The project was initiated by [Pippa Skotnes](#), assisted by Eustacia Riley, [Thomas Cartwright](#), [Cara van der Westhuizen](#) and [Fazlin van der Schyff](#) and is regularly updated. Please refer to the CCA website at <http://www.cca.uct.ac.za> for updates.

KEYWORD SEARCH

Books
Cover to cover | Contributor | Story | Category | Keyword

Drawings/Watercolours
Contributor | Category | Keyword | !xam | !kun | Researchers

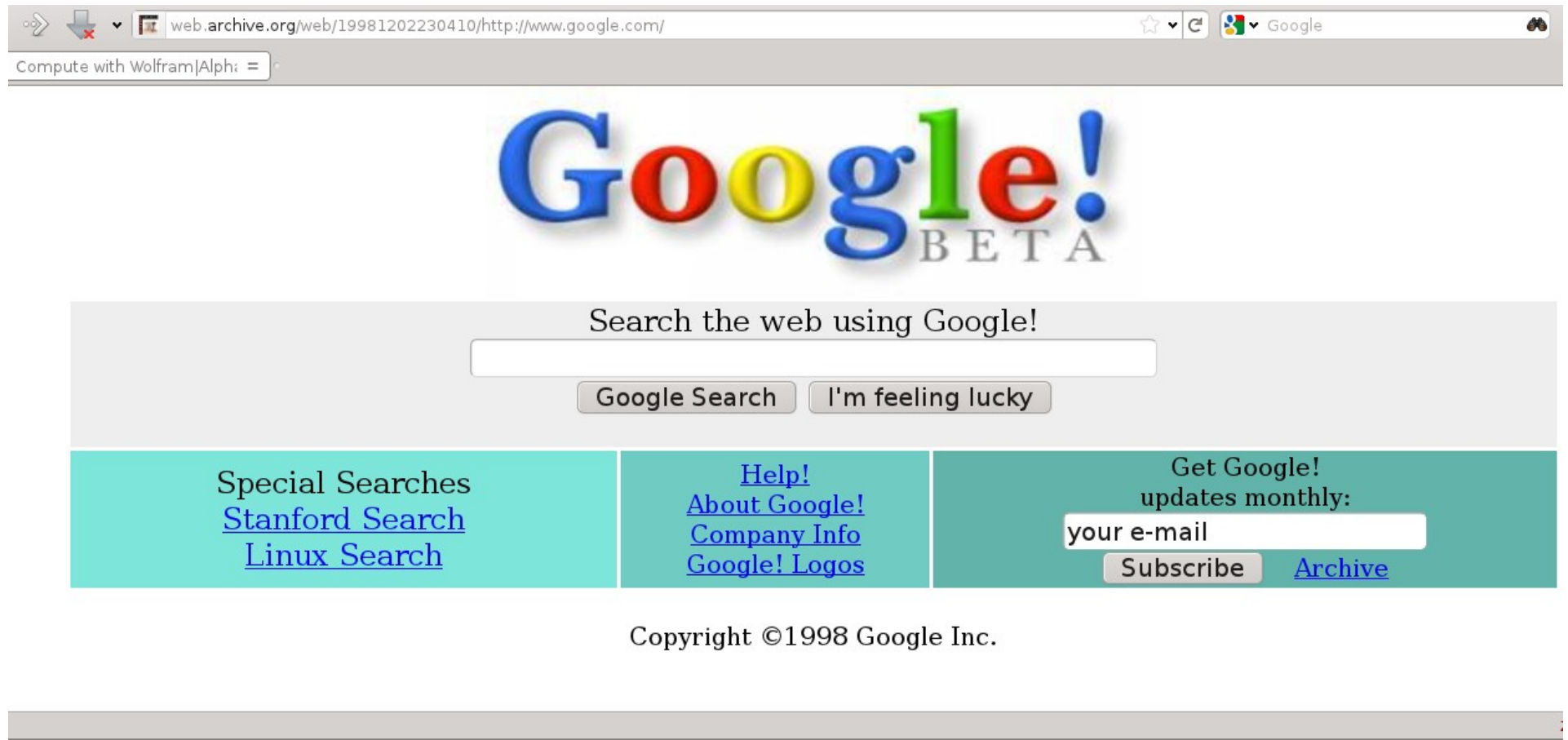
Contributors

This digital publication is part of a Llapec project to digitise, research and publish the Bleek and Lloyd Archive. Llapec (the Lucy Lloyd Archive, Resource and Exhibition

(<http://lloydbleekcollection.cs.uct.ac.za/>)

Large scale preservation

- Internet Archive



(<http://www.archive.org/>)

Problem Statement

Motivation

- Complexity of existing solutions
 - Preservation costs
 - Initial, maintenance and recovery costs
 - Funding
 - Resource limitation
 - Skills and education
 - Internet Bandwidth
 - Expensive in developing countries



Related Work

Frameworks

- DELOS DL reference model
 - Abstract framework
- Kahn/Wilensky Framework
 - Generic IS framework
 - DSpace and Fedora
- 5S Framework
 - Formal mathematical model

Digital Library System tools

- **Popular FOSS** tools for creating digital collections



Research Questions

Previous observations

- 2007
 - Digital Libraries Without Databases
 - in-Browser Digital Library Services
- 2008
 - An African Perspective on Digital Preservation
- 2010
 - Hybrid Online-Offline Collections

Previous observations (cont.)

- simplyCT principles
 - Any metadata, any objects, any services
 - Minimalism
 - Preservation by copying
 - Repeatability
 - Superimposed information
 - Explicitly avoid the use of APIs

Hypotheses

- A formal simplistic abstract framework for Digital Library System design can be derived
- A Digital Library System architectural design based on a simple framework can be easily adopted and managed over time

Research questions

- Is it feasible to implement a Digital Library System using simple architectures?
 - How should simplicity for Digital Library System storage and service architectures be defined?
 - What are the comparative advantages and disadvantages between simple and complex architectures?

Methodology

Approach

- Pilot study
 - Digital library collection prototype
- Theory derivation
 - simplyCT principles
 - Case study of existing tools
- Design and Implementation
 - Extension of prototype DLS
 - Collection integration

Evaluation

- Efficiency
 - Controlled experiments
- Effectiveness
 - Real world case studies
- Usefulness
 - User studies

Work in
Progress

Pilot study

- Bleek and Lloyd collection
 - Dublin Core Application profile
 - Fully qualified Dublin Core
 - Virtual and physical objects
- Metadata Object Relationships
 - <dcterms:hasPart>
 - <dcterms:isPartOf>
- Rudimentary digital collection
 - Search and browse

Anticipated Outcome

Anticipated outcome

- A formal definition of simplicity for digital library system design
- A proof of concept Digital Library System
- A set of principles behind the simplyCT framework

Thank you