

Managing personal documents with a digital library

Imene Jaballah, Sally Jo Cunningham, Ian H. Witten

Department of Computer Science, University of Waikato, Hamilton, New Zealand

ibaj1@cs.waikato.ac.nz

sallyjo@cs.waikato.ac.nz

ihw@cs.waikato.ac.nz

Abstract: This paper presents a desktop system for managing personal documents. The documents can be of many types—text, spreadsheets, images, multimedia—and are organized in a personal “digital library”. The interface supports browsing over a wide variety of document metadata, as well as full-text searching. This extensive browsing facility addresses a significant flaw in digital library and file management software, both of which typically provide less support for browsing than for searching, and support relatively inflexible browsing methods. Usability studies of a prototype were conducted to suggest design refinements, which were then incorporated into the final system.

Key words: browsing, digital library, document management, metadata, personal information management.

1 Introduction

For nearly four decades, personal computers have been using the desktop and folder system metaphors. These metaphors use a hierarchical structure to allow users to store and access documents in their personal file space. This approach worked quite well as long as the number of items was in the range of hundreds, but it does not scale to thousands or ten of thousands of files. The challenge has shifted from deciding what to keep, to finding specific documents when they are needed (Soules & al. 2003). The result is too many folders for the users to organize, remember and access when seeking information within their personal collection of files.

Currently, the ability of users to browse and search through their files is limited by conventional hierarchical structure and location-based browsing. Strict hierarchies map poorly to user needs. The restriction that a document can appear only in one place at any given time, and using document locations as the principle of organization structure, forces computer users to create strict categorizations for their files. Previous studies of filing practices of computer users have suggested that such restrictions to a hierarchical structure can hinder rather than help users in quickly finding desired documents. Providing other means for browsing would give users more flexibility when looking for information in their personal electronic collections.

The system described here is an attempt to provide better support for information seeking within personal

information collections, through a Desktop Digital Library (DDL). Although the DDL supports both searching and browsing, the emphasis is on browsing based on document properties and document contents—that is, those features of a document that are meaningful to users. The implementation is based on a digital library solution, Greenstone, and uses a metadata-based approach.

Previous attempts to provide different and better ways of browsing include Tree-maps, which present the relationships between two dimensional images and their representation in hierarchical tree structures (Shneiderman 1992). Alternatively, Boardman (Boardman 2001) proposed a technique to organize resources at the workspace level, by sharing one hierarchy between all applications. Freeman and Gelernter have proposed the Lifestreams project which provides a complete file management system based on time stamps (Freeman & al. 1996). Lifestreams generates a visualization of documents organized by time, forming a personal history. However, all these solutions escape one fixed organizational scheme, the folder-hierarchy, to fall into another, such as the time-line. Users need not be restricted to two dimensional representations, hierarchical structures or temporal organizations.

The closest related work to this project is UpLib (Janssen & al. 2003)—a personal digital library system. The system could be accessed through an active agent via a Web interface (similar to the Greenstone’s collection access method). In addition, like Greenstone it provides a full-text index of the collection documents. The system uses both document