In this paper we present the PDLib concept: personal digital libraries with universal access. A universally accessible personal digital library system provides each user with a customizable, general purpose document repository (i.e. a personal digital library) and the means to access it from anywhere at any time from most computing devices connected to the Internet, including mobile phones, PDAs and laptops, therefore granting access from anywhere at anytime. Personal digital libraries provide traditional digital library services such as document submission, full-text and metadata indexing and document search and retrieval; augmented with innovative services for the moment to moment information management needs of the individual user and adapting the services to the device that is being used for interaction with the library. These innovations include flexible collection and metadata management, interaction with other digital libraries (whether personal digital libraries of other users or traditional digital repositories) and sharing of digital content with other users. These requirements are addressed by a personal digital library system architecture that considers traditional digital library challenges and mobile computing challenges. The main components of the architecture are the Client-Side Applications, the Data Server and the Mobile Communication Middleware (MCM). The current architecture targets to support multiple device access and support for mobile devices is also considered while traditional services have to be adapted to cope with the restrictions of mobile devices. A proof-of-concept prototype implementation of the PDLib System that exemplifies many of the main concepts is available and an overview of this prototype is presented in this paper.

ABSTRACT

In this paper we present the PDLib concept: personal digital libraries with universal access. A universally accessible personal digital library system provides each user with a customizable, general purpose document repository (i.e. a personal digital library) and the means to access it from anywhere at any time from most computing devices connected to the Internet, including mobile phones, PDAs and laptops, therefore granting access from anywhere at anytime. Personal digital libraries provide traditional digital library services such as document submission, full-text and metadata indexing and document search and retrieval; augmented with innovative services for the moment to moment information management needs of the individual user and adapting the services to the device that is being used for interaction with the library. These innovations include flexible collection and metadata management, interaction with other digital libraries (whether personal digital libraries of other users or traditional digital repositories) and sharing of digital content with other users. These requirements are addressed by a personal digital library system architecture that considers traditional digital library challenges and mobile computing challenges. The main components of the architecture are the Client-Side Applications, the Data Server and the Mobile Communication Middleware (MCM). The current architecture targets to support multiple device access and support for mobile devices is also considered while traditional services have to be adapted to cope with the restrictions of mobile devices. A proof-of-concept prototype implementation of the PDLib System that exemplifies many of the main concepts is available and an overview of this prototype is presented in this paper.

Categories and Subject Descriptors

D.2 [Software Engineering]: Software Architectures; D.2.11 [Software Architectures]: Domain-specific architectures—personal digital library system
H.3 [Information Storage and Retrieval]: Digital Libraries; H.3.7 [Digital Libraries]: Systems issues—personal libraries, universal access, mobile environments

General Terms

mobile personal digital library system architecture

Keywords

personal libraries, universal access, mobile environment

1. INTRODUCTION

We propose a universally available personal digital library system. It is “personal” in the sense that each user is provided with a general purpose document repository (i.e. a personal digital library). It is “universally available” in the sense that it allows the user to access her/his personal personal digital library from most computing devices connected to the Internet, including mobile phones, PDAs and laptops, therefore granting access “from anywhere at anytime.”

Personal digital libraries provide traditional digital library services such as document submission, full-text and metadata indexing and document search and retrieval; augmented with innovative services for the moment to moment information management needs of the individual user. These innovations include provisions to customize the classification of documents, interact with other digital libraries (whether personal or collective) and support user-to-user exchange of generic digital content.

Another distinction between our personal library system and traditional digital library systems is that it strives to provide its users with universal access, the capability of accessing the personal digital libraries “from anywhere at anytime” [19]. Consequently, we have designed our system for mobile environments.

Since our system extends traditional library services for the mobile environment in order to realize the abstraction of personal libraries, we must cope with the technological challenges imposed by the implementation of digital library services [4, 1, 7], the mobile environment [3, 13, 20] and the specific requirements of personal digital libraries.

The effectiveness of the document retrieval process from personal digital libraries must be assured with the provision
of an intuitive classification schema and the indexing of documents. A personal digital library system should allow users to define classification schemas according to their needs.

The creation of the personal digital library implies the submission of digital documents and their placement on the personal digital library under user-defined classification schemas.

The documents of a personal digital library must be accessible via a mechanism capable of providing meaningful answers to user’s queries. In a personal digital library system, search and retrieval mechanisms must adapt to the personalized classification schema defined by the users of the system.

The documents of the personal digital library must be readily available in response to user demand. This is quite a challenge in mobile environments, since connection variability is likely to interrupt document transfers. Additionally, the great diversity of mobile devices imposes a multi-platform approach to the design of mobile digital library systems.

A personal digital library must provide each user with mechanisms to restrict unauthorized access to their personal digital library and with administrative tools to manage digital library content.

Despite the apparent contradiction of the personalization goal of a personal digital library system with the communication standardization efforts of traditional digital library systems, interoperability is a challenge worth facing for the added value it provides to personal digital library services.

In this paper, we present a concept architecture and current status of our Personal Digital Library System (PDLib for short) that addresses these challenges with the integration of information storage and retrieval, database and mobile client technologies to provide universally available personal digital libraries.

The rest of this paper is organized as follows. In Section 2 we describe the concept of universally available personal digital libraries. Based on this discussion, Section 3 describes the architecture and implementation issues of the PDLib system. In Section 4, we compare PDLib with other work in the field. Finally, we conclude in Section 5 with directions of future work.

2. UNIVERSALLY AVAILABLE PERSONAL DIGITAL LIBRARIES

Traditional digital library systems grant a group of users access to a digital library. Personal digital library systems provide a digital library to each user (i.e. a personal digital library). PDLib is a personal library system that allows the user to shape and access her/his digital library from anyplace at anytime using nearly any computing device (i.e. universal access). To realize universally available digital libraries, the following requirements were defined for the PDLib system:

- **Flexible collection and metadata management.** Collections must be provided as a mechanism for document classification. Users should be provided with the ability to define the metadata set that will be used to describe the contents of each collection. These interactions will allow the user to customize her/his personal library as desired.

- **Digital document submission.** The user should be able to add any digital document to a personal digital library. Submission from several device types should be supported. The personal digital library must be able to accommodate several document formats for the same document. In addition, since provision for library content personalization is permitted by user-defined collections with customizable metadata sets, while submitting a document the user must be able to: (a) select or create the collection that will contain the document and (b) provide metadata information for the document according to the collection’s metadata set.

- **Search and retrieval.** Search and retrieval mechanisms must adapt to the personalized classification schema defined by the users of the system.

- **Universal access.** In order to provide universal access to the documents and services of the personal digital library, several client application types suitable for mobile and fixed hosts of mobile environments [3, 13] must be considered. Software clients can be classified according to their client-side architecture into: (a) thin clients and (b) thick clients; and agreeing with their mobility into: (a) fixed clients and (b) mobile clients (Figure 1). In thin clients, the application is delivered on a browser or microbrowser; while on thick clients both code and data reside on the device. Thick and thin clients can coexist in a single device. Mobile clients are those with wireless connections to the system’s network; while fixed clients have a reliable—wired connection—with the network.

- **Administration and access control.** The owner of a personal digital library always has unrestricted access to her/his personal digital library content. In addition, the owner should be provided with the capability to provide other users with access to her/his personal digital library content.

- **Interoperability.** Interoperability with other personal digital library users and with other digital library systems using well-known interoperability protocols.

Personal digital libraries are composed of collections. Collections contain, in turn, other collections and/or documents.

![Figure 1: Mobile Environment Clients](image-url)
Users can interact with personal digital libraries by creating and deleting collections and submitting, moving, copying or downloading documents. In addition, users can define the metadata set that will be used to describe the contents of each collection. These interactions allow the user to customize her/his personal library as desired.

To illustrate the concept of universally available personal digital libraries, we provide two hypothetical scenarios below. We have deliberately chosen scenarios feasible in a university context because we believe the concept of a personal digital library to be an interesting mobile application. These scenarios exemplify the two key ideas of our universally available personal digital library system: personal libraries and universal access.

Scenario 1: Personal Libraries

Shaping the Personal Library. Sarah, a visual arts undergraduate student, has been working on a two-student writing assignment for her Comparative Mythology class. Sarah and Aidan, her team partner, have chosen the comparison of the Norse and Greek mythological systems as their assignment topic. Sarah is using her Personal Digital Library (PDLib) from her desktop PC to store and classify documents (e.g. text, audio, video and/or image files) regarding the Norse myths. She has decided to arrange her documents under a collection named “Norse Myths” created for this purpose. Within the “Norse Myths” collection, she has created other collections to group documents under certain themes (e.g. “Theogonic Myths”, “Creation of the Universe”) or has placed the documents directly under “Norse Myths”. She has also created a “Vikings Online” collection to store useful Internet references. In order to simplify the classification of the “Vikings Online” collection and ease its population, Sarah has customized the metadata set definition of the “Vikings Online” collection so that it contains just a “Title” and “URL” fields. Finally, Sarah has set the permission rights of her “Norse Myths” collection so that Aidan has access to her documents. The files stored by Sarah will stay in her PDLib even after she graduates for further use.

Scenario 1 focuses on the composition of our personal libraries and the services they provide. While storing references for her writing assignment, Sarah interacted with objects of a personal library. The objects explicit in the scenario are: (a) documents, (b) collections, (c) metadata sets and (d) permissions. In addition, implicit in the scenario we encounter the following objects: (a) libraries, (b) document formats and (c) document metadata. Therefore, in addition to digital documents, the content of a personal digital library also includes user-defined collections, metadata—both user-defined field definitions (i.e. metadata sets) and document information (i.e. document metadata)—and permission rights. In Section 3 we present the model that supports this.

Library services are built around personal digital library objects. The services provided by a personal digital library are shown in Table 1.

The services provided by a personal digital library are:

- **CRUD Operations.** Creation, Retrieval, Update and Deletion (CRUD for short) operations to manage library, collection, document, document format and metadata sets objects. In the case of document objects, the creation operation is the document’s submission to the personal digital library.
- **Copy/Move Documents or Collections.** Copy and Move operations are provided to duplicate or re-locate document and collection objects within the hierarchical structure of the personal digital library.
- **Document Search.** Boolean search and ranked search of documents according to their full-text content and/or associated metadata. The following variants of the search operation have been defined:
  - **Basic Search.** Returns documents according to a search expression.
  - **Advanced Search.** Returns documents according to several metadata-specific search expressions.
  - **Full-Text Search.** Returns documents according to their textual contents.
  - **Recursive Search.** The search scope is performed in all the collection hierarchy.
  - **Multi-Library Search.** The search scope includes libraries from other users or other digital library systems.
- **Send Document.** Sends a document to the personal digital library of another user. Document metadata and, optionally, document formats are sent as well.
- **Email Document.** Sends document and metadata via email to any user with an email account.
- **Document Format Conversion.** Perform document format conversion (e.g. plain text document from a PDF document, a PNG image from a JPEG image).
- **Get/Set Collection Metadata Set.** Retrieve the metadata set of a given collection or assign a metadata set to a collection.
- **Update Document Metadata.** Updates document metadata first provided at document submission (i.e. creation of the document).
- **Get/Set Document or Collection Permissions.** Allow the user to grant/revoke permission rights to other users to her/his personal documents and collections.

### Table 1: Personal Digital Library Services

<table>
<thead>
<tr>
<th>Personal Digital Library (PDLib)</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library</td>
<td>(a) Creation, (b) Retrieval, (c) Update and (d) Deletion.</td>
</tr>
<tr>
<td>Collection</td>
<td>(a) Creation, (b) Retrieval, (c) Update, (d) Deletion, (e) Copy and (f) Move.</td>
</tr>
<tr>
<td>Document</td>
<td>(a) Submission, (b) Retrieval, (c) Update, (d) Deletion, (e) Search, (f) Email, (g) Copy and (h) Move.</td>
</tr>
<tr>
<td>Document Format</td>
<td>(a) Submission, (b) Retrieval, (c) Update, (d) Deletion and (e) Conversion.</td>
</tr>
<tr>
<td>Metadata Set</td>
<td>(a) Creation, (b) Retrieval, (c) Update, (d) Deletion and (e) Get/Set Collection Metadata Set.</td>
</tr>
<tr>
<td>Document Metadata</td>
<td>(a) Get/Set Collection Permission and (b) Get/Set Document Permission.</td>
</tr>
</tbody>
</table>
Scenario 2: Universal Access

Always there for Me. Aidan, currently working on campus at the Visual Arts Center and off campus as a graphic designer, is constantly moving from one location to another. He also changes his working computer frequently since he has been assigned a PC laptop at the Visual Arts Center and a desktop Macintosh at the design agency. Luckily, his PDLib is there for him. During his precious free time at both works, Aidan is able to browse the references found by his roommate Sarah for their shared homework using his PDLib to access hers. As a complement to Sarah’s research, Aidan has to find useful information regarding Greek mythology. Aidan uses his unoccupied time at work to gather references relevant to his topic. Aidan has been storing his findings at his PDLib under a collection called “Greek Myths”. On one occasion, he ran into his friend Matthew at the campus walkways. After briefing each other on their latest doings, they found out that Matthew—who is working on an essay for his Classical Mythology class—could use one of Aidan’s references. Since Matthew does not have a PDLib, Aidan uses the email document feature of his PDLib from his mobile phone to send Matthew a copy of the reference. A few days before the due date of his written assignment with Sarah, Aidan bought himself a PDA. Having written the joint essay and stored a copy of both the finished essay and its related presentation in their PDLibs, Sarah and Aidan decide to retrieve the presentation using PDLib in preparation for their in-class lecture. Later, at the lecture’s Q&A section, both Sarah and Aidan were able to access their PDLibs to search, retrieve and display some of the documents quoted in their work from Aidan’s PDA. After being praised for the quality of their presentation, Sarah and Aidan were able to send the essay via PDLib to their professor and classmates.

Scenario 2 portrays the convenience of universal access. Despite moving from one place to another or changing computing devices, Aidan was able to reach his PDLib whenever and whenever he needed to. This was possible since PDLib’s architecture can provide digital library services to most computing devices with an Internet connection. The architecture contemplates different several client application types and the great diversity of hosts in a mobile environment. Client applications are responsible of adapting PDLib’s services according to the interface and computing resources of client devices. Mobile thick client applications also adapt to the limitations of the mobile environment. Specialized connection middleware is used to support the adaptation measures undertaken by mobile thick clients.

Our personal digital library concept considers the interaction with other digital library systems. This means that our client applications can be used to browse the contents of other digital library systems while providing the user with a subset of the personal digital library services to interact with the content of other library systems. Interoperability has been made possible by the inclusion of OAI-MHP [15] compliant modules in our personal digital library system architecture.

3. SYSTEM ARCHITECTURE

Our goal for PDLib is to realize the concept of universally available personal digital libraries. As presented in Section 2, a universally available personal digital library system seeks to provide the abstraction of individual digital libraries in a mobile environment [17]. An overview of the PDLib system is shown in Figure 2 and its architecture is provided in Figure 3.

The main component of the system is the PDLib Data Server. The data server stores the objects of the personal digital libraries (see Table 1). The data server also provides the client applications with a communication module that enables remote access to the personal digital library content. In order to support the requirements introduced in Section 2, we have developed and architecture and a prototype that provide digital library services to several types of devices according to their computing capacity (e.g. desktop, laptop, PDA and mobile phone) with multiple operating systems (e.g. Microsoft Windows, Linux, Mac OS, Palm Os and Microsoft Windows CE). The driving idea behind this architecture is to provide each device with the best way to interact with the services offered by the data server.

The architecture of PDLib also supports the interaction with other digital library systems via the OAI-MHP. This means that a PDLib user can navigate in other digital library systems compliant with the OAI-MHP using PDLib’s infrastructure. Figure 2 shows a general overview of the PDLib system. The system is composed of three layers:

1. **Client Tier.** Includes the variety of devices with which an user can interact with PDLib;

2. **Server Tier.** Shows the server system infrastructure that provide services to clients: (a) Data Server, (b) Mobile Connection Middleware (MCM) and (c) Web Front-end.

3. **Interoperability Tier.** Includes other (PDLib) data servers and OAI-MHP compliant digital library systems.

The devices of the client tier communicate with the server tier to access PDLib digital library services. The access type of the client tier with the server tier varies according to the client device’s capabilities. The architecture of PDLib reflects the following access types:

1. **Middleware Access.** Supports mobile devices, especially those with limited computing resources (e.g. HTTP-enabled mobile phones, PDA).
2. **Web Access.** Provides HTTP-access to any device that includes a Web browser (e.g. WML/HTML microbrowser-enabled mobile phones).

3. **Direct Access.** Applications with very particular requirements can access the data server directly. In Section 3 we provide an example of such an application.

The following sections describe the components of the PDLib system and their interactions.

### 3.1 Clients

As mentioned before, the software clients of a mobile environment can be classified according to their client-side architecture and mobility. We elaborate on this classification with support of Figure 1. Figure 1 plots client-side architecture on the vertical axis and client mobility on the horizontal axis. The quadrants contain stereotypical protocols and development platforms for each client category. Mobile thick clients are especially important to achieve PDLib’s goals since mobile applications should center on mobile thick clients—a key difference with Web applications, which typically focus on fixed thin clients—since thick clients make offline operation possible. A technological migration trend that goes from Web applications to mobile applications is suggested. This trend reflects recent software application development tendencies and implies that the technologies available in an earlier stage are present in subsequent stages as well.

It is possible to map the mobile client application types to the devices supported by PDLib as follows (refer to Figure 2 and Figure 1):

- **According to client-side architecture.** Thin client applications connect to the Data Server through the Web Front-end while thick client applications communicate with the Data Server via the Mobile Connection Middleware.

- **According to mobility.** Mobile applications connect to the server tier via unreliable wireless connections while fixed applications connect to the server tier via a reliable wired connections. Figure 2 shows wireless connections as dashed lines and wired connections as solid lines.

With the objective of providing an access medium appropriate to the devices shown in Figure 2, the following client application types were defined:

- **Mobile clients (mobile thick clients).** Mobile clients were designed to cope with the limitations of the mobile environment. This means that mobile clients redefine the functionality provided by PDLib’s fixed clients to provide the abstraction of a personal digital library on a mobile device. A middleware is required to help mobile clients achieve their mobile connection adaptation goals.

- **Web clients (fixed and mobile thin clients).** This category includes those devices with a browser or microbrowser capable of displaying HTML (Figure 6) or WML pages. Despite the fact that Web clients do not have all the features available in thick clients, they provide basic interaction support with PDLib. Web clients connect with PDLib’s Web Front-end.

- **Application clients (fixed thick clients).** Application clients are intended to run in desktop and laptop devices with few computing resources constraints. Application clients are desirable because they provide a richer interface than Web clients. In addition, application clients can communicate directly with the data server, which leads to a more efficient communication of fixed clients.
The previous client types were defined because each type has very particular characteristics. Application clients are intended to run on resource unconstrained devices such as desktop or laptop. On the other hand, a mobile user can interact with PDLib using a mobile client. Mobile clients allow the interaction with the personal library offering a subset of the functionality provided by fixed thick clients. Web clients, on the other hand, were designed to endow the user with access to their library without the prerequisite of a client application installed on a computing device. One of the main challenges is in mobile clients, since they have to deal with the limitations of the mobile environment. To accomplish their task, mobile clients perform the following functions:

- **Local Storage Mechanism.** Store documents in the mobile device for offline viewing.
- **Connection Adaptation Mechanism.** This mechanism provides a constant response time despite the connection variability of wireless connections. A constant response time can be achieved by calculating the size of the data-transfer window and predicting network state [6].
- **User Interaction Support.** Graphical interface that allows the user to manage the personal digital library content from its mobile device. Mobile clients read results received from the MCM and display them on screen.

### 3.2 Web Front-end

The Web Front-end transforms personal digital library services into a web application. In order to support fixed and mobile thin clients, the Web Front-end is capable of delivering WML or HTML according to the requesting device (i.e. browser or microbrowser). The Web Front-end provides the following functionality:

- **Session Handling.** Session handling is used to maintain the interaction of a thin client with the Web Front-end beyond a single HTTP request.
- **Browser Markup Language Support.** Verifies the markup languages supported by the requesting browser and serves HTML or WML accordingly.

### 3.3 Mobile Connection Middleware

One of the main problems to solve in order to provide the data server services is the fact that the data server has been designed to be used by a wide range of devices and not just mobile devices. However, there is a clear difference in computing resources between mobile (e.g. PDA) and fixed (e.g desktop) devices. This computing resource disparity makes it difficult to adapt the data server to the capabilities of mobile devices; and signals the need of a middleware component to mediate the interaction of the mobile device and the data server. This arrangement also precludes the necessity of changing the previously established services. We call this middleware the Mobile Connection Middleware (MCM). The MCM is designed to provide following functionality:

- **Connection Support.** Is required by mobile clients in order to perform adaptation to the high bandwidth variability of the mobile environment and to cope with the frequent disconnections of mobile devices.
- **Process Delegation.** Execute functions that would demand an excessive (and quite possible, unavailable) amount of computing resources to the mobile device.
- **Mobility Support.** Operations such as prefetching can be performed to speed up the retrieval of documents from the data server and store them in cache servers that are closer to the user. When the user changes location, it would be necessary to support migration of the information between different instances of the MCM.
- **Device Interaction Support.** Performs adaptation of content according to the characteristics of the device on which it is desired to show the information.

### 3.4 Data Server

The central part of PDLib is the data server. The data server provides the services of a personal digital library, stores personal digital library data and supports interoperability via the OAI-MHP. The data server provides the following functionality:

- **Personal Library Services.** The data server offers creation, retrieval, update and deletion operations over the library objects stored in the personal data storage (collections, documents and metadata). The data server also provides services to copy and move documents or collections, search for documents in a personal library, send documents to the personal library of other users, email documents to any user with an email account, perform document format conversion, change collection metadata sets, edit document metadata and grant or revoke permission rights over personal library content.
- **Personal Data Storage.** Stores the data of the personal digital libraries. A text search engine is used to index the content of the personal digital libraries. A database is used to store the objects of the personal digital libraries since we require a structured model to represent personal library objects and the relations among them. The usage of a text search engine and a database in the data server permit the combined use of text-based queries and SQL datatype-based queries to support the hierarchical classification of personal library content. The hierarchical classification can greatly improve query performance since it allows queries to be specified against documents contained in individual collections with or without the inclusion of their nested collections to the scope of the query.
- **OAI-MHP Support.** The data server exposes the metadata of the personal digital library documents via the OAI-MHP. In addition, the data server harvests metadata of other OAI-MHP compliant library systems to provide users with a subset of the personal digital library services to interact with other (OAI-MHP compliant) library systems.

Our personal digital library data model shown in Figure 4 establishes that a library contains one or more collections. A collection contains documents or more collections and is associated with a metadata set. The metadata set is composed by one or more metadata definitions.
A metadata set is the metadata of the document metadata (e.g., name="Author", type="Text" and name="Title", type="Text"). The document metadata is the metadata of a particular document (e.g., name="Author", value="S. Sturluson" and name="Title", value="Prose Edda"). The fields that compose document metadata are determined by the metadata set of the collection. The metadata set of a collection defaults to the metadata set of the parent collection and can be redefined by the user on a per collection basis. Both collections and documents have an associated permission. Permissions are particularized in the following access rights: (a) Personal. Access is restricted to library owner; (b) Incoming. Other users are allowed to create documents and collections. This access right applies only to collections; (c) Outgoing. Other users are allowed to browse collections and retrieve documents; (d) Update. Other users are allowed to update documents and their associated document metadata and document formats. However, other users are not allowed to create documents nor delete them; and (e) Shared. Other users are granted full access to documents and collections. Permission access rights associated with documents contained in a collection defaults to the permission access right associated with their parent collection. A document can have one or more document formats.  

3.5 System Status  
As of the writing of this paper, the following prototype implementations of the PDLib components have been developed:

- **A mobile thick client application:** PDLib Mobile Client. The Mobile Client application allows the user to: 1. Glance through the contents of his/her personal digital library (Figure 5 (a)); 2. Retrieve and store a local copy of selected digital library documents on the mobile device; 3. Manage her/his personal digital library by: i. Moving/copying documents or collections and ii. Setting access rights to documents or collections (Figure 5 (b)); 4. Share their digital library content by: i. Sending documents via email and ii. Sending a document to another PDLib user (Figure 5 (b)); 5. Search for documents in her/his personal digital library (Figure 5 (c)); 6. View and edit document metadata (Figure 5 (d)); and 7. Browse the contents of other users’ personal digital libraries. Prototype implementations of PDLib’s Mobile Client have been developed using CLDC/MIDP J2ME [5] (Figure 5) The prototype applications communicate with the MCM via the XML-RPC protocol libraries of the Java/XML Enhydra Application Server [11].

- **A fixed client application:** The PDLib BeUp! Bin. The BeUp! Bin is a prototype client application that uploads files to a special collection where all unclassified documents are stored. The current implementation of this application supports the addition of files via “drag-and-drop” operations, the queueing of upload requests and direct file selection. The BeUp! Bin is an example of an application client with direct access to the Data Server (as explained earlier in this section). The application communicates directly with the PDLib Data Server using the Apache XML-RPC client library [2].

- **Data Server.** PDLib’s Data Server provides personal digital library content storage and retrieval services. The prototype implementation of the Data Server supports the creation, retrieval, update and deletion operations over the library objects stored in the personal data storage (collections, documents and metadata). The data server also provides services to copy and move documents or collections, search for documents in a personal library, send documents to the personal library of other users, email documents to any user with an email account, change collection metadata sets and edit document metadata. The data server implementation exposes the metadata of the personal digital library documents via the OAI-MHP and harvests metadata of other OAI-MHP compliant library systems in order to allow PDLib clients to browse their contents. A prototype implementation of the data server has been created using J2SE [8]. MySQL is used
to store all personal digital library data [14]. We are currently working to incorporate the Lucene full-text search engine [12] to improve performance and reduce index storage requirements.

- **MCM (Mobile Connection Middleware).** The MCM interacts with the Mobile Client to provide the following connection adaptation services: (a) Session handling over the stateless XML-RPC protocol; (b) A constant collection navigation response time despite the connection variability of wireless connections; (c) Partition document transfers and support disconnections while transferring documents. In addition to proving connection adaptation, the MCM also stores query responses received from the data server in order to accelerate the responses it serves to the Mobile Client. In the current implementation, the MCM services are accessed via XML-RPC. The Apache XML-RPC implementation is used to provide the services [2].

- **Web Front-end. (a.k.a. PDLib Web Client)** The Web Front-end was developed using Java Servlets, Cascading Style Sheets (CSS), HTML and WML. The Web Front-end verifies the display capabilities of the requesting browser and answers accordingly (e.g. serving a WML page instead of HTML if the requesting browser is incapable of displaying HTML) (Figure 6). The Web Front-end runs on the Tomcat servlet container [23].

### 4. RELATED WORK

This section compares PDLib with other digital library and mobile database projects with goals related to the universally available personal digital libraries concept. This comparison distinguishes PDLib from the projects presented in the remainder of this section.

**DSpace** is a Web-based institutional digital library system that captures, stores, indexes, preserves and redistributes the intellectual output of a university’s research faculty in digital formats [22, 21]. DSpace differs from PDLib in several aspects such as: (a) Purpose. DSpace focuses on institutional (i.e. collective) digital libraries while PDLib is concerned with personal digital libraries. (b) Client Application Type. DSpace is a Web-based application and targets fixed-thin clients while PDLib addresses thick and thin clients on mobile and fixed devices. (c) Structure. DSpace follows a Web-based three-layer architecture divided in storage, business, and application layers. PDLib follows a multilayered peer to peer architecture to support Web-based applications and mobile applications. (d) Data Distribution. DSpace centralizes digital library data in a single data store while the PDLib system architecture considers: (i) Integrating the data of distributed data stores; and (ii) Communicating digital library data across several data server subsystems.

**Greenstone** is a software system for the creation of digital library collections. The Greenstone system seeks to provide a unified framework for searching and browsing library contents. The Greenstone system differs from the PDLib system in the following: (a) Purpose. Greenstone realizes collective digital libraries while PDLib is concerned with personal digital libraries; (b) Client Application Type. Greenstone has a Web-based interface and facilities to create standalone CD-ROM working versions of the digital library. On the other hand, PDLib addresses thick and thin clients on mobile and fixed devices. (c) Build Process. Greenstone building process can be cumbersome for large collections, taking a day or two to complete [24]. PDLib does not require a building process, since its search engine component supports the addition and removal of documents to existing indexes (i.e. user actions effectively “build” the digital library interactively). (d) Data Distribution. The Greenstone system centralizes the digital library data in a single data store while the PDLib system architecture considers integrating the data of distributed data stores and communicating digital library data across several data server subsystems.

**Owl Infract Engine** (Owl) is a multi-user Web-based document repository system for publishing documents of a corporation, small business, group of people, or an individual [16]. Owl presents the following differences with the PDLib system: (a) Individual Libraries. Owl is a multi-user environment where an individual digital library is defined with permission policies. PDLib provides personal digital libraries with access control policies that allow the user to share her/his library data with other users; (b) Information Storage and Retrieval. Both Owl and PDLib support keyword-based full-text search on document content and document metadata. However, the metadata definition of Owl is fixed while PDLib allows its users to specify custom metadata definitions. In addition, PDLib extends the keyword-based full-text search with standard database queries. (c) Client Application Type. Owl is a Web-based application and targets fixed-thin clients while PDLib addresses thick and thin clients on mobile and fixed devices. (d) Data Distribution. Owl centralizes the digital library data in a single data store while the PDLib system architecture considers integrating the data of distributed data stores and communicating digital library data across several data server subsystems.

The **Phronesis** system is a practical and efficient tool for the creation, administration and maintenance of distributed digital libraries on the Internet [7, 18]. Phronesis digital libraries are grouped in collections. A Phronesis collection consists of documents relevant to a knowledge area or relevant to a group of people. The Phronesis system, like Greenstone, builds on the Managing Gigabytes program [25]. Phronesis includes the following distinctive features [18]: (a) Full-text (text, PostScript, html, pdf and rtf documents) and metadata indexing and searching based on the MG...
(Managing Gigabytes) system [25], (b) Metadata based on the international standard Dublin Core. (c) Searching and retrieval of English and Spanish documents. (d) Compression of digital library content. (e) Parallel search in remote Phronesis Collections. (f) Spanish and English Web-based interface. (g) Access control policies. Some differences between Phronesis and PDLib follow: (a) Purpose. Phronesis focuses on collective digital libraries while PDLib is concerned with personal digital libraries. (b) Client Application Type. Phronesis is a Web-based application and targets fixed-thin clients while PDLib addresses thick and thin clients on mobile and fixed devices. (c) Build Process. As in the case of the Greenstone building process, Phronesis building process can be cumbersome for large collections. PDLib does not require a building process, since its search engine component supports the addition and removal of documents to existing indexes. (d) Data Distribution. Phronesis centralizes the digital library data in a single data store while the PDLib system architecture considers integrating the data of distributed data stores and communicating digital library data across several data server subsystems. However, while Phronesis data stores are all centralized, Phronesis provides support for parallel searches in remote collections.

UpLib is a Web-based full-text indexed repository accessed through an active agent. The UpLib system has been specifically designed for secure use by a single individual [10]. Collaborative operation of multiple UpLib repositories is possible with extensions that facilitate community building around individual document collections [9]. UpLib is “universal” in the sense that documents are represented as projections into the text and image domains. UpLib interface is predominantly based on page images. Therefore, UpLib can handle any document format which can be rendered as pages. UpLib includes mechanisms to provide alternative representations of documents besides their textual and image representations.

While both UpLib and PDLib define themselves as being universal personal digital library systems, there are fundamental differences between the motivating factors behind each system. These differences follow: (a) Universality Notion. UpLib provides “universality” by representing documents as text and images. UpLib is a universal digital library system since it provides a uniform alternative to digital library systems with specific language extensions (like those provided by Phronesis) or domain specific descriptions (such as a well-defined metadata set for documents in a particular knowledge area). Conversely, PDLib attempts to provide “universal access”—i.e. from anyplace at any time [19]—to personal digital libraries. (b) Personal Libraries. UpLib is a personal digital library system since it was specifically designed for secure use by a single individual and to accommodate for user-defined extensions [10]. In contrast, PDLib declares itself a personal digital library system because it seeks to provide the abstraction of individual digital libraries in a mobile environment. Both UpLib and PDLib are digital library systems that store individual knowledge. However, UpLib provides a universal interface for library content while PDLib endeavors to allow personalization and customization of library content. PDLib supports individualization of library content with a data model designed to provide the notion of user-defined collections (i.e. possibly nested, document containment entities) with customizable metadata sets.

Besides their conceptual differences UpLib and PDLib have the following operational differences: (a) Client Application Type. UpLib is a Web-based application and targets fixed-thin clients while PDLib addresses thick and thin clients on mobile and fixed devices. (b) Variety of Formats. UpLib can handle any document format which can be rendered as pages and includes mechanisms to provide alternative representations of documents besides their textual and image representations. Alternatively, PDLib’s data server subsystem has been designed for the storage and retrieval of general purpose content and is not constrained by the visual representation (or lack of it) of document formats. Nonetheless, the data server’s keyword-based retrieval schema does require the text projection of documents. (c) Data Distribution. UpLib centralizes the digital library data in a single data store while the PDLib system architecture considers integrating the data of distributed data stores and communicating digital library data across several data server subsystems. However, collaborative operation of multiple UpLib repositories is possible with extensions that facilitate community building around individual document collections [9].

5. CONCLUSIONS AND FUTURE DIRECTIONS

We have presented the concept of personal digital libraries with universal access. A universally accessible personal digital library system provides each user with a customizable, general purpose document repository (i.e. a personal digital library) and the means to access it from anyplace at anytime from nearly any computing device. In particular, we have described the requirements, architecture and implementation of a personal digital library system with universal access: The PDLib system. To realize universally available personal digital libraries, the following requirements were characterized: (a) Flexible collection and metadata management; (b) Digital document submission; (c) Search and retrieval; (d) Universal access; (e) Administration and user control; and (f) Interoperability. These requirements were addressed by our personal digital library system architecture. The main components of the architecture are the Client-Side Applications, the Data Server and the Mobile Communication Middleware (MCM). The Client-Side Applications target mobile and fixed devices and depending on the device digital library services are adapted. The Data Server supports the services for data storage, indexing and retrieval of information from users. The MCM is an intermediary between mobile devices and the Data Server and provides functionality for mobility support like connection adaptation. Currently a proof-of-concept prototype implementation exists that exemplifies many of the main concepts presented on this paper.

The PDLib system is an ongoing research and development effort where interesting challenges related to digital libraries and mobile computing are presented. We are currently working on the following research topics: (a) Scalability A personal digital library system must cope with an increasing number of mobile users that will demand a large amount of distributed storage resources to hold an even greater quantity of personal data objects (e.g. documents): (b) Universal access and availability In addition to the diversity of the client applications, we have expanded the uni-
versal access criterion to include measured performance and availability metrics; support of digital library services during offline operation as well as ad-hoc interactions are being defined. Caching and prefetching techniques that can improve performance of mobile devices accessing digital library objects are also topics of interest for our work; (c) Content adaptation Content summarization and visualization methods of digital library information on limited screen size devices are being explored could provide an added value to mobile clients of our personal digital library system; and (d) Interoperability In addition to the current functionality there is the opportunity for exploring further interoperability with OAI-MHP and other systems like z39.50 and web search engines. The traditional concept of a library was a site where documents were stored and patrons had to attend the library site in order to access the documents. The information technology opened new opportunities allowing the access to traditional and new library services from remote locations. Personal digital libraries with universal access provide services that support the abstraction that a single user has his own library that he can carry with him anywhere he goes. He can manage and access documents anywhere and can also interact with personal digital libraries from other users or with traditional digital library repositories. University students can have their own Personal Digital Library where they store important papers during the time that they attend the university, after they graduate, they keep their Personal Digital Library and continue expanding their own collections according to their personal and professional interests. Faculty members and researchers can have their own Personal Digital Libraries, during a conference, they can add/retrieve documents to/from their personal library as they interact with colleagues. These are just some sample scenarios of the potential applications for this concept.

6. ACKNOWLEDGMENTS

We would like to thank Roberto Garcia-Sanchez, Adan Salinas-Flores, Luis Hurtado-Alvarado, Miguel Escoffie-Puerto, Eugenio Flores-Centeno, Marcos Guevara-Camacho and Eduardo Alvarez-Cavazos for their support in the elaboration of this paper. To all of them our most sincere gratitude.

7. ADDITIONAL AUTHORS

Lorena G. Gomez-Martinez (ITESM, Campus Monterrey, email: lgomez@itesm.mx) and Martha Sordia-Salinas (ITESM, Campus Monterrey, email: msordia@itesm.mx).

8. REFERENCES


[17] PDLib. The personal digital library project. URL copernico.mty.itesm.mx/pdlib/.
[18] Phronesis. The phronesis digital library project. URL copernico.mty.itesm.mx/phronesis/project/.