

《海外观察》

PDLib----个人数字图书馆普遍存取的系统框架*

王立贤 李书宁 编译

1.引言

随着无线通信的出现和发展,用户即使从一个地区到另一个地区仍旧能够连接到网络,这种现象被称为手机或者移动(*nomadic*)计算。移动计算正在改变着人们交流、工作和使用技术的方式。传统数字图书馆的服务一直借助于网络搜索界面,随着移动通信技术的广泛使用,可供数千万用户移动设备访问的数字图书馆系统在不久的将来一定会出现。这种系统需求的变化要求传统的数字图书馆框架做出一定的调整。

本文提出了一个基于普遍存取(*universal access*)的个人数字图书馆系统。个人数字图书馆不但需要提供诸如文献传递、全文和元数据标引、文档查询和检索等传统数字图书馆服务项目,还需要提供诸如个性化文本分类、与其他数字图书馆交互以及支持用户之间的数字文档交换等创新服务。这里“普遍存取”是指该系统允许用户从可以连接到网络的大多数处理设备(包括移动电话、掌上电脑(PDA)、手提电脑等)登陆到自己的个人数字图书馆,从而达到能够在任何时间、任何地点为用户提供服务的目的。

个人数字图书馆在移动环境下拓展传统图书馆的服务,不但必须解决数字图书馆服务以及移动环境具体实施过程中的技术问题,还要充分考虑个人数字图书馆的特殊要求,这些要求包括:(1)个人数字图书馆文献检索效果需要由固有的分类机制和文献标引机制来保证,因此个人数字图书馆应该允许用户根据自己的需要定义分类方案。(2)创建个人数字图书馆就意味将提交数字文献,具体的存储位置就是用户自定义分类方案下的个人数字图书馆。(3)个人数字图书馆文献的获得则是通过与系统用户自定义的个性化分类机制相适应的查询检索机制获得。(4)个人数字图书馆必须保证文献的可阅读性,因为移动环境下连接的不稳定很容易中断文献的传输。此外,移动工具的多样性给移动数字图书馆系统的设计提供了一个多平台途径。(5)个人数字图书馆不但必须为每一位用户提供一种限制个人数字图书馆未经授权访问的机制,还要提供一种管理工具来管理数字图书馆的内容。

2、普遍存取的个人数字图书馆

传统数字图书馆系统向一组用户提供数字图书馆的访问,而个人数字图书馆系统向每个用户提供一个数字图书馆。本文提及的 PDLib 就是允许用户设计并在任何时间任何地点采用几乎任何工具访问(即:普遍存取)的个人数字图书馆系统。为实现普遍存取的功能,PDLib 应满足如下条件:

(1) 灵活的资源 and 元数据管理:系统必须为资源提供一种文献分类机制,用户自己也可以定义描述每一种资源内容的元数据集。

(2) 数字文献提交:系统应该支持多种提交设备,也应该为同一文献提供多种文献格式。

(3) 信息查询检索机制与系统用户定义的个人分类方案相适应。

(4) 普遍存取:系统必须考虑移动环境下与移动和固定主机相适合的各种客户端应用软件的类型(如:胖客户端和瘦客户端,固定客户端和移动客户端)。

(5) 管理和授权控制: PDLib 拥有者不仅可以自由的获取其个人数字内容,还可以授

*本文论及的系统是墨西哥 Tecnológico de Monterrey 大学主持的个人数字图书馆开放源代码项目的成果 PDLib,实现了无线通信环境下的个人数字图书馆访问和获取。译文有删节,结构也稍有改动。

权其他用户获取这些资源。

(6) 互操作：使用标准互操作协议与其他用户和 PDLib 系统进行交互操作。

为了解释普遍存取的个人数字图书馆的概念，本文假设三个综合背景下切实可行的场景，第一个和第二个分别描述了个人数字图书馆和普遍存取机制；第三种则主要说明 PDLib 系统已经获得的移动连接支持机制。

场景 1：个人数字图书馆

视觉艺术专业的大学生 Sarah 和 Aidan 两人一组正在做“比较神话学”的课程作业，已经选择了挪威和希腊两个神话系统作为作业的主题进行比较。Sarah 正在使用她个人电脑上的个人数字图书馆系统 PDLib 存储和类分关于挪威神话的文档（如文本、音频、视频和图像文件），她决定把这些文档存放在名为“挪威神话”的文档集（collection）中。在这个文档集中，她已经创建了其它特定主题的文档集用于对文档进行分组，或直接将文档放在“挪威神话”中。她还创建了一个“海盗在线”的文档集来存储有用的网络资源。为了详细说明“海盗在线”文档集的分类，Sarah 定制了该文档集的元数据集，包含“题名”和“URL”两个字段属性。最后，Sarah 设置了“挪威神话”文档集的访问权限，使 Aidan 有权访问她的文档。Sarah 存储的文件将会保存在她的 PDLib 中甚至直到她毕业后进一步使用。

场景 1 描述了个人数字图书馆的组成（文档、文档集、元数据集和访问准许等）及其提供的服务（参见表 1）。

表 1 个人数字图书馆服务

Personal Digital Library (PDLib)	
Objects	Services
Library	(a) Creation, (b) Retrieval, (c) Update and (d) Deletion.
Collection	(a) Creation, (b) Retrieval, (c) Update, (d) Deletion, (e) Copy and (f) Move.
Document	(a) Submission, (b) Retrieval, (c) Update, (d) Deletion, (e) Send, (f) Search, (g) Email, (h) Copy and (i) Move.
Document Format	(a) Submission, (b) Retrieval, (c) Update, (d) Deletion and (e) Conversion.
Metadata Set	(a) Creation, (b) Retrieval, (c) Update, (d) Deletion and (e) Get/Set Collection Metadata Set.
Document Metadata	(a) Update Document Metadata.
Permission	(a) Get/Set Collection Permission and (b) Get/Set Document Permission.

场景 2：普遍存取

Adian 校内任视觉艺术中心工作，校外作图像设计师，经常在不同地方之间变换，而且他也经常更换电脑，在视觉艺术中心使用笔记本电脑，在设计机构使用台式电脑。幸运的是，他的 PDLib 总是在为他服务。在他两份工作之余，Adian 则通过自己的 PDLib 访问 Sarah 的 PDLib，浏览 Sarah 找到的参考资料。随着对 Sarah 研究的补充，Adian 必须查找与希腊神话有关的有用信息，他利用自己的闲暇时间收集相关资料，并将找到的资料存储在他 PDLib 上名为“希腊神话”的文档集中。Adian 在校园的人行道上遇到朋友 Matthew，发现 Matthew 可以用到 Adian 的一个资料。由于 Matthew 没有 PDLib，Adian 通过自己的 PDLib 的电子邮件文献传递功能从他的移动电话上发给 Matthew 一个该资料的副本。在作业完成的前几天，Adian 买了个 PDA。他们的 PDLib 中存储了一些已经写完的论文以及相关的陈述，Sarah 和 Adian 决定检索 PDLib 中的陈述资料，为课堂上的报告作准备。在课堂上，Sarah 和 Adian 通过 Adian 的 PDA 方便的访问他们的 PDLib 来查找、检索和展示报告中引用的文档。由于他们高质量的陈述而受到表扬，之后，Sarah 和 Adian 通过 PDLib 将他们的论文发送给教授和同学。

场景 2 描绘了普遍存取的便利性。PDLib 系统的构建能够为连接在网络上的大多数处理

设备提供数字图书馆服务。另外，这里个人数字图书馆的概念考虑到了不同数字图书馆系统之间的交互作用：客户端应用程序能够用于浏览其它数字图书馆系统的内容，同时向用户提供个人数字图书馆服务的子集，与其它图书馆系统的内容之间交互信息。在个人数字图书馆系统的构建过程中可以通过内置 OAI-PMH 适应模块实现了系统间的互操作。

场景 3：移动连接支持

Matthew 刚刚得到了一个 PDA，并在朋友 Adian 的建议下获得了一个 PDLib。昨晚在家里台式电脑上为自己的 PDLib 创建了几个文档集，能用 PDA 通过校园的无线网络访问 PDLib。随后，他在校园中漫步时用 PDLib 进行试验，发现尽管网络不稳定，却仍能平稳的在图书馆中浏览，Matthew 很高兴地发现 PDLib 用几乎不变的速度回答了自己的提问。经过进一步观察，他意识到当他重新追踪自己的浏览记录时，PDLib 的回应时间甚至更快了。Matthew 还注意到从 PDLib 往 PDA 中取回文献时，在失去网络连接甚至关闭 PDA 后，一旦连接恢复，文献传递仍然能够成功的完成。

场景 3 描述了由包含普遍存取设备的个人数字图书馆通过移动瘦客户端所提供的用户体验，设备包括一个移动连接中间件，提供如下服务：（1）稳定的导航回应速率（constant navigation rate），在一个相对稳定的时间内向用户提供尽可能多的数据记录；（2）快速后向导航（back navigation speed up），建立缓存机制快速追踪用户的浏览记录；（3）基于连接感知的文献传递（connection aware document transfers）：用户能够在移动工具和 PDLib 之间传递文件，这种传递是连接感知的，一旦连接建立就可以进行文件传递，传递速率根据网络状况自行调整。

为了实现上述情景描述的功能，我们根据不同设备（如台式机、笔记本、PDA 和移动电话）的处理能力以及不同操作系统（如 Windows、Linux、Mac OS、Palm OS 和 Windows CE）的要求设计系统的架构和原型。

3、系统构建

PDLib 的目标是实现普遍存取的数字图书馆。如第二部分所述，普遍存取的数字图书馆系统致力于提供移动环境下的个人数字图书馆的抽象模型。图 1 和图 2 分别是 PDLib 系统概况和系统框架。系统的主要组成部分 PDLib 数据服务器，它存储了个人数字图书馆的对象（见表 1），还提供客户端应用程序，通过交流模块确保个人数字图书馆内容的远程获取。PDLib 的架构还支持通过 OAI-PMH 与其它数字图书馆系统之间的交互，这就意味着，一个 PDLib 用户能够使用 PDLib 的设备在其它支持 OAI-PMH 协议的数字图书馆系统中浏览。

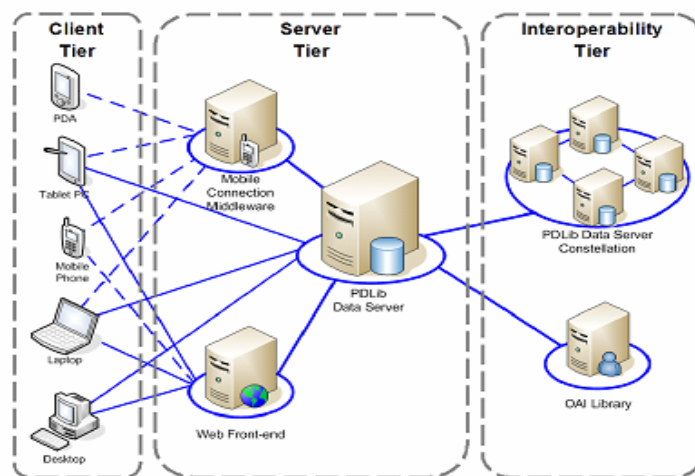


图 1 PDLib 系统概况

图 1 展示了 PDLib 系统的概况，它由三个部分组成：(1) 客户端层 (Client Tier)：包括用户与 PDLib 交互的各种设备；(2) 服务器层 (Server Tier)：向客户端提供服务，包括数据服务器、移动连接中间件 (MCM) 和网络前端；(3) 互操作层 (Interoperability Tier)：包括其它 PDLib 数据服务器和兼容 OAI-PMH 协议的数字图书馆系统。

客户端层设备通过服务器层进行交流来获取 PDLib 数字图书馆的服务，访问类型因用户工具的不同而不同，PDLib 系统框架主要提供了三种访问类型：(1) 中间件访问：支持移动工具，特别是只有有限处理资源的设备，如支持 HTTP 的移动电话和 PDA；(2) 网络访问：向任何工具提供 HTTP 访问，包括网络浏览器，如 WML/HTML 微浏览器支持的移动电话；(3) 直接访问：某些有特殊需求的软件能够直接访问数据服务器。

下面详细描述了 PDLib 系统的各个组成部分以及各个组成部分之间的交互过程。

3.1 客户端

移动环境下的软件客户端设备可以按照它们的客户端架构和移动性进行分类，图 3 详细描述了这一分类。图 3 中纵轴表示客户端构架，横轴表示客户端的移动性。移动胖客户端对于实现 PDLib 的目标特别重要，因为移动应用程序应该以移动胖客户端为中心，这也是它与网络应用程序的关键不同点。网络应用程序更加注重固定瘦客户端，因为胖客户端有可能实现脱机操作。技术转移正在从网络应用程序向移动应用程序的方向发展，这种趋势反应了近来软件应用程序的发展趋势，预示着早期可利用的技术在接下来的阶段也会出现。

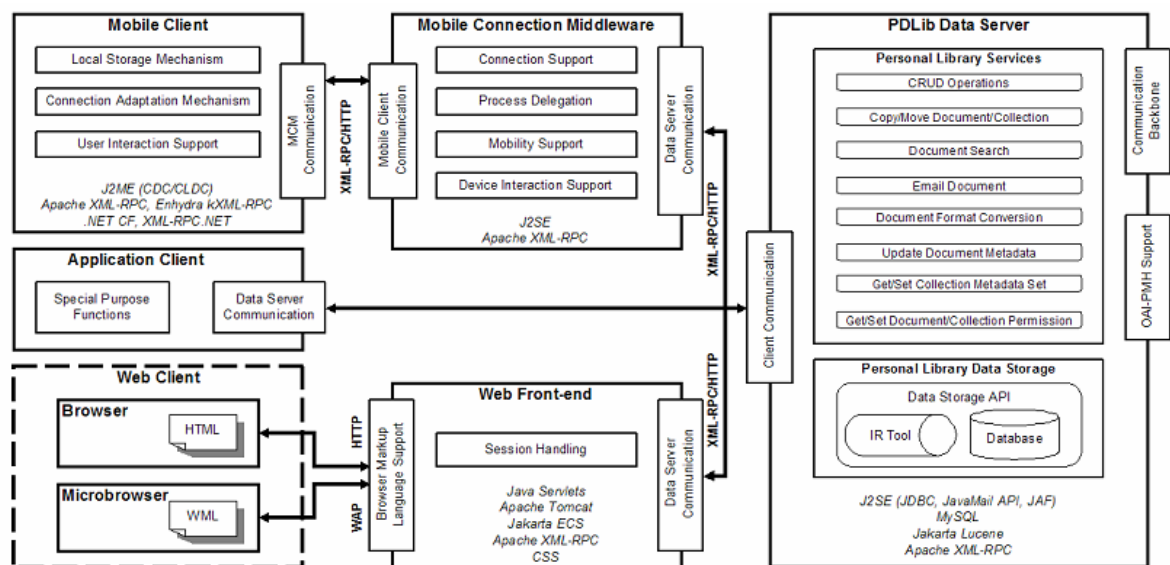


图 2 PDLib 系统框架

与 PDLib 支持的用户工具相对应的移动客户端应用软件类型可以划分如下 (参见图 1 和图 3)：

(1) 根据客户端的结构，可以分为胖客户端和瘦客户端，瘦客户端通过网络前端连接到数据服务器，而胖客户端通过移动中间件连接到数据服务器。

(2) 根据移动性，可以分为移动应用程序和固定应用程序。移动应用程序通过不可靠的无线网络连接到服务层，固定应用程序则通过可靠的有线网络连接到服务层。图 1 中使用虚线表示了无线连接，使用了实线表示了有线连接。

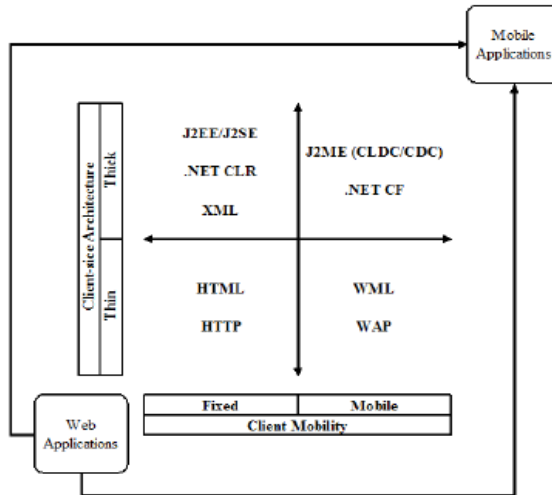


图 3 移动环境客户端

为了图 1 中所展示的用户设备提供相应的访问中介, 本文对客户端应用程序类型进行了如下定义:

(1) 移动客户端 (移动胖客户端): 设计移动客户端是为了应对移动通信环境的限制, 用于重新定义固定客户端提供的功能, 在移动设备上提供一个抽象的个人数字图书馆。这需要中间件来协助移动客户端完成移动连接调整的功能。

(2) 网络客户端(固定和移动的瘦客户端): 包括那些具有浏览器或微浏览器显示 HTML 或 WML 页面的设备。网络客户端提供了与 PDLib 基本的交互功能, 与 PDLib 的网络前端进行连接。

(3) 应用软件客户端(固定瘦客户端): 在台式电脑和便携电脑上运行, 没有资源限制, 能够直接与数据服务器交流信息。

应用软件客户端目的是在没有资源限制的工具如台式电脑和笔记本电脑上运行。移动用户通过移动客户端与 PDLib 交互信息。移动客户端允许与个人图书馆的交互, 提供由固定胖客户端提供的功能的子集。网络客户端的设计用于允许用户不必事先在处理工具上安装就能够访问他们的图书馆。由于必须处理移动环境的限制, 因此其面临的主要挑战之一就是移动客户端。为了完成任务, 移动客户端要履行如下功能:

(1) 本地存储机制: 在移动工具中存储文档以便于脱机浏览; (2) 连接适应机制: 尽管无线连接不稳定, 但这一机制能够提供稳定的回应时间, 通过计算数据传输窗口的大小和预测网络状况得以实现; (3) 用户交互支持: 图画式界面允许用户从移动工具管理个人数字图书馆的内容, 移动客户端阅读从 MCM 接收的内容, 并将它们显示在屏幕上。

3.2 网络前端

网络前端能够将个人数字图书馆服务转换到网络应用程序上。为了支持固定客户端和移动瘦客户端, 网络前端能够根据请求工具 (如浏览器或微浏览器) 传递 WML 或 HTML 文件, 提供如下功能: (1) 时域控制 (session handling), 用于维持瘦客户端与网络前端之间除单一 HTTP 请求之外的交互服务; (2) 浏览器置标语言支持, 用于识别请求服务的浏览器的标识语言并提供相应 HTML 或 WML 服务。

3.3 移动连接中间件 (MCM)

数据服务器提供的服务是针对大量设备工具, 而不仅仅是移动设备。但是移动设备和固定设备处理的资源存在明显的不同, 这种不同使得数据服务器难于适应移动工具的性能, 中间件就用来协调移动工具和数据服务器之间的交互, 我们称之为移动连接中间件 MCM, 它具有如下功能: (1) 连接支持: 使移动用户能够适应动态变化的移动环境, 处理移动工具

频繁断开的连接；(2) 委托处理：用于完成移动设备需要处理大量资源时所需的功能；(3) 移动支持：可以加快文档检索和存取速度，用户改变位置后还可以支持 MCM 在不同场合的信息移植；(4) 设备交互支持：根据不同工具的特性使内容与之相适应。

其中当前工作的重点就是移动连接支持，我们已经建立图 4 的原型构件系统，具有如下机制：(1) 通过时域机制，保持连接状态信息和用户工具特性的记录；(2) 通过分页机制，提供数据服务器的原始数据的分页，使其便于在移动工具显示屏上显示和文档片段的传输；

(3) 通过存储机制存储数据服务器回应的查询结果，加快移动客户端服务速度；(4) 通过连接调整机制，根据网络连接状态调整 MCM 和移动客户端数据交流的速度。

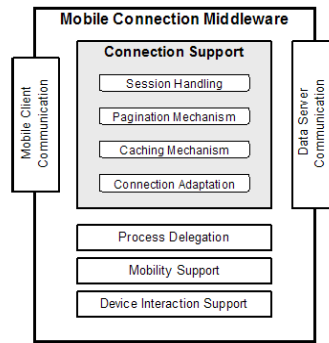


图 4 MCM 框架

移动连接支持是由以上机制和移动客户端协同完成的，MCM 针对这些机制的实施向移动客户端提供如下连接支持服务：(1) 建立在无状态 XML-RPC 协议上的时域控制；(2) 稳定的导航回应时间；(3) 快速后向导航；(4) 可靠的移动网络文档传输。

3.4 数据服务器 (data server)

数据服务器是 PDLib 的核心，提供个人数字图书馆服务，存储个人数字图书馆数据，支持基于 OAI-PMH 的交互操作。它提供如下功能：

(1) 个人图书馆服务：创建、检索、更新和删除个人存储资源（文档集、文档和元数据）；文档和文档集的复制和移动；在个人图书馆中查询文档；向个人图书馆上传其他用户的文档；通过电子邮件向任何拥有邮件账号的用户传递文档；完成文档格式转换；更改元数据集；编辑文档元数据；提供或终止 PDLib 内容的访问授权。

(2) 个人数据存储：提供 PDLib 的结构化存储和可升级的检索功能；使用搜索引擎等信息检索工具来索引个人数字内容，使用数据库来存储结构化数据对象和它们之间的关系。信息检索工具索引允许快速随机访问编入索引的内容，并用于查找存储在数据库中的信息。在数据服务器中使用文本搜索引擎和数据库，允许将基于文本的和基于 SQL 数据类型的查询结合起来使用，以支持个人数字内容的等级式分类。

(3) 通讯中枢：建立与其它数据服务器的通讯中枢，能够实现全球移动数据管理。全球数据管理解决的是诸如定位、选址、复制和广播这样的网络相关的问题。全球数据管理应用在 PDLib 的一个典型例子就是个人图书馆复制。当用户所处位置改变到距离另一个数据服务器更近的位置时，如果相当长的时间内没有再变化，MCM 会把用户的图书馆复制并定位到离他最近的数据服务器上。

(4) 支持 OAI-PMH：数据服务器通过 OAI-PMH 公开 PDLib 文档的元数据，并收割其它基于 OAI-PMH 的图书馆系统的元数据，向用户提供其服务子集，以方便与其它 OAI-PMH 图书馆系统进行交互。

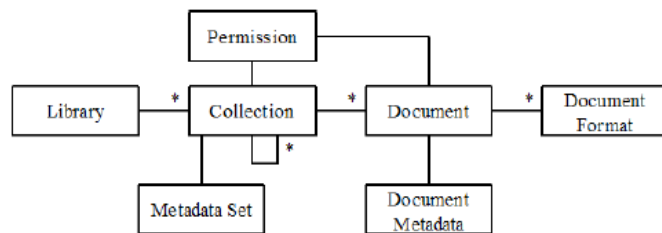


图 5 个人数字图书馆的数据模型

图 5 所示的个人数字图书馆数据模型表明一个图书馆包含一个或多个文档集，文档集会包含文档或更多的文档集，并与相应元数据集关联，这个元数据集则由一个或多个元数据定义组成。元数据集是文档元数据的元数据（如名称=“作者”，类型=“文本”和名称=“题名”，类型=“文本”）。文档元数据是某一特定文档的元数据（如名称=“作者”，属性值=“Snorri Sturluson”和名称=“题名”，属性值=“Prose Edda”）。组成文档元数据的域由文档集的元数据集决定。文档集元数据集的默认值与其父母文档集的元数据集相同，并能由用户重新定义。对于访问许可，则包含以下几种：（a）Personal 权限：访问许可仅限于图书馆拥有者；（b）Incoming 权限：允许其他用户创建文档和文档集，访问许可仅适用于文档集；（c）Outgoing 权限：允许其他用户浏览文档集和检索文档；（d）Update 权限：允许其他用户更新文档及与之相关联的文档元数据和文档格式，但不允许创建或删除文档；（e）Shared 权限：授权其他用户对文档和文档集的全部访问准许。一个文档集与所包含的文档相关联的访问权限许可与其上层文档集默认相同。一个文档也可以拥有一个或多个文档格式。

3.5 系统实现

下面介绍一下 PDLib 原型系统的具体实现：

(1) 移动客户端应用程序的实现

PDLib 移动客户端允许用户：①浏览自己 PDLib 的内容（图 6（a））；②检索和存储在移动工具上选定的数字图书馆文档的本地副本；③通过 i、移动、复制文档和馆藏；ii、定文档和馆藏的访问权限来管理个人数字图书馆（图 6（b））；④通过 i、利用 email 传递文档；ii、向其他 PDLib 用户传递文档来共享数字图书馆内容（图 6（b））；⑤在自己的个人数字图书馆中搜索文档（图 6（c））；⑥查看和编辑文档元数据（图 6（d））；⑦浏览其它用户的个人数字图书馆的内容（图 6）。

移动客户端原型使用 CLDC/MIDP J2ME 进行开发，移动客户端则通过开放源代码的 Java/XML Enhydra 应用服务器的 XML-RPC 协议集与 MCM 交流信息。

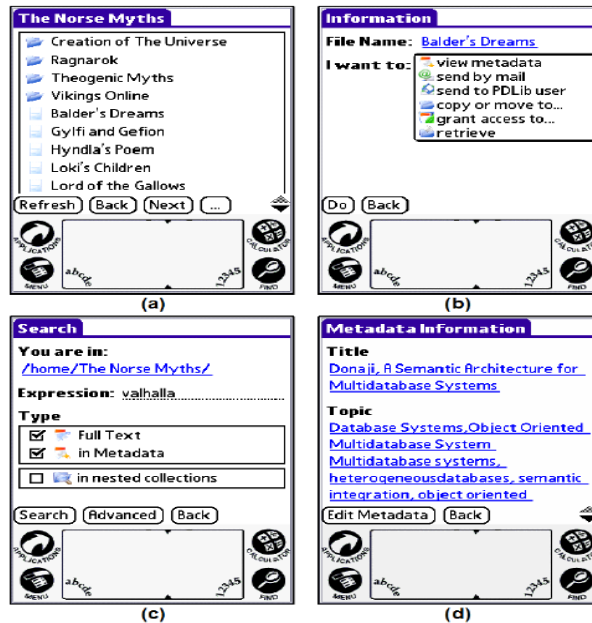


图 6 PDLib 移动客户端 (J2ME CLDC/MIDP)

(2) 固定客户端应用软件 (PDLib BeUp! Bin)

The BeUp! Bin 是一个客户端应用软件原型, 可以向特定文档集上传文件, 所有未分类的文档都存储在这个文档集中。这个软件当前的应用是通过拖放操作、上传请求排队和直接的文件选择来支持文档的增加。这是一个直接访问数据服务器的应用软件的例子, 通过 Apache XML-RPC 客户端图书馆直接与 PDLib 数据服务器交流信息。

(3) 数据服务器

PDLib 的数据服务器提供个人数字图书馆内容存储和检索服务。数据服务器原型支持对存储在个人数据仓储 (文档集、文档和元数据) 中的图书馆数字对象的创建、检索、更新和删除操作。数据服务器还提供其它服务, 如复制和移动文档和文档集, 在个人图书馆中搜索文档, 向其它个人图书馆传递文献, 向任何拥有电子邮件账户的用户发送文献, 改变文档集的元数据集合编辑文档的元数据等。数据服务器的实现, 通过 OAI-PMH 揭示个人数字图书馆文档的元数据, 并能收割其它适用 OAI-PMH 的图书馆系统, 允许 PDLib 客户端浏览它们的内容。它的原型实施已经通过 J2SE 实现, MySQL 用来存储所有的 PDLib 数据。我们现在所做的工作是整合 Lucene 全文搜索引擎以改善 PDLib 的性能并降低索引存储要求。

(4) 移动连接中间件

MCM 与移动客户端相结合提供如下连接支持服务: ① 建立在无状态 XML-RPC 协议上的时域控制; ② 稳定的导航回应时间; ③ 通过储备数据服务器对请求的回复提高导航速度; ④ 经由移动网络的可靠的文档传递, 通过三种方式实现: (a) 文档片段间的连接; (b) 文档片段的独立传输; (c) 时域支持的文档传输。在当前实现过程中, MCM 服务可以通过 XML-RPC 获得, Apache XML-RPC 就提供这一服务。

(5) 网络前端

网络前端使用 Java Servlets、层叠式样式表 (CSS)、HTML 和 WML 进行开发。网络前端检验请求端浏览器的显示性能, 并根据其做出回答 (如果请求端浏览器无法显示 HTML, 网络前端就提供 WML 页面) 网络前端运行在 Tomcat servlet container 上。

4、未来研究方向

PDLib 系统是一个不断发展着的系统, 目前和将来主要是关注以下几个方面: (a) 可测量性: PDLib 系统必须能够处理日益增加的移动用户所带来的大量分布式个人数据资源的存

储要求；(b)普遍存取和可用：除了用户应用程序的多样性以外，系统还必须扩展数据的访问标准，支持数字图书馆服务的脱机操作以及特别交互（一种计算机之间的通信模式，不需要接入点，各无线设备以点对点的方式进行直接通信）。另外，改善移动设备访问数字图书馆性能的缓存和预取技术也是一个研究热点；(c)内容调整：数字图书馆信息内容的摘要和在小屏幕访问设备上显示的方法会给 PDLib 系统移动客户端系统提供额外价值；(d) 交互性：除了最新交互功能以外，还要进一步开发基于 OAI-PMH 和其它基于 z39.50 或网络搜索引擎的系统。

编译自：

- 1、Francisco Alvarez-Cavazos, David A. Garza-Salazar, Juan C. Lavariega-Jarquín, Lorena G. Gomez-Martinez, Martha Sordia-Salinas. Universal Access Architecture for Personal Digital Libraries. <http://copernico.mty.itesm.mx/pdlib/docs/TR-PDLIB-01-ARCHITECTURE-3.pdf>
- 2、Francisco Alvarez-Cavazos, Roberto Garcia-Sanchez, David Garza-Salazar, Juan C. Lavariega, Lorena G. Gomez, Martha Sordia. Universal Access Architecture for Digital Libraries. <http://delivery.acm.org/10.1145/1110000/1105636/p12-alvarez-cavazos.pdf>