Infrastructure for Digital Repositories

Infrastructure for Digital Repositories
Richard Marisa, Cornell University

Two complementary technologies, Dienst and CUPID, facilitate serving, navigating, searching and printing digital documents.

Dienst is the protocol underlying NCSTRL (http://www.ncstrl.org/), the National Computer Science Technical Report Library. A new version (5.1) of the Dienst protocol extends the ability to manage metadata and to navigate "structured" documents (for example, the ability to request "chapter 2" of a book, or retrieving a page by its "native" page number).

A lightweight implementation of the Dienst 5.1 protocol runs under Windows NT and uses XML to represent metadata and document structure information, and to communicate with client applications. An example application built on Dienst 5.1 features full text searching of scanned historical law journals based on OCR data.

To facilitate production of printed reproductions of digital documents, Dienst 5.1 cooperates with CUPID, a printing architecture specified by the CNI CUPID "Consortium for University Printing and Information Distribution" (http://www.cni.org/docs/ima.ip-workshop/CUPID.html). We are using CUPID printshop clients to direct documents to local printers, the Cornell Digital Print Shop, Kinko's and a local offset printer. A "Dienst Printshop Client" planned for CUPID will allow users to virtually "print" an electronic document to a Dienst archive for viewing and subsequent printing by remote users.

Here is a sampling of the Electronic Printing and Publishing Initiatives in process at Cornell:

EZ-Publish II

> EZ-Publish II is a service built on the CUPID architecture which allows users to print certain types of common finished documents (e.g., booklets) at on- and off-campus printshops. An alpha version of the system was used in 1998 by the CIT/ATS Publications group to automatically send dozens of

titles for printing at the on-campus Digital Print Shop, to Kinko's and to an offset printer in Ithaca. A beta version of EZ-Publish II and evaluation by additional users is planned in 1999.

PubWeb Collaboration

PubWeb, Inc., (www.pubweb.com) of Woburn, MA, is using the CUPID architecture to facilitate the ordering, processing, printing and delivery of custom textbooks. PubWeb is attempting to secure rights to textbook content which may be combined with local content to produce textbooks designed for a specific course and printed on demand. Cornell (CIT/OIT/ATS) has executed a collaborative CUPID development agreement with PubWeb. A PubWeb system is being evaluated in Cornell's Digital Print Shop.

Hein Law Review Project

William S. Hein & Co., Inc. of Buffalo, NY, in collaboration with the Cornell Law Library and Cornell Information Technologies (OIT/ATS), are preparing an online collection of historical law review journals. Hein, a re-publisher of historical law materials which services research law libraries, is scanning the materials and providing content and structure metadata in XML. These are sent to Cornell where the page images are OCR'd and mounted in a Dienst- 5.1 based digital document archive. An application to browse and search the collection online is being developed in CIT; the Cornell Law Library is coordinating an evaluation of the system's capabilities and interface with input from other research law libraries. The Dienst-5 archive is being developed in CIT in coordination with Carl Lagoze of the Cornell University Computer Science Department. A CUPID gateway will facilitate printing excerpts of the online materials.

Net-Print

Net-Print (www.cit.cornell.edu/cit-pubs/net-print) is a fee-based laser printing service available in Academic Technology Services (ATS) Computer Labs, the Residence Hall Network (ResNet) and elsewhere on the Cornell campus. The system authenticates users and allows them to pay for printing requests via student bursar accounts, cash debit accounts or other special accounts. Net-Print allows

users to queue documents to any of dozens of laser printers in several locations, view a log of their print jobs and track monthly printing charges on their bursar bills from any properly configured workstation.

Synergies

Together, the CUPID, Dienst and Net-Print infrastructures allow us to contemplate new applications for authenticating and authorizing access to digital materials, and for viewing and printing materials in digital archives on demand. These technologies are targeted for several new projects currently in the development/specification stages.

Richard Marisa
Electronic Printing and Publishing Initiatives
Cornell Information Technologies
220a Computing and Communications Center
Garden Avenue
Cornell University

(607) 255-7636
rjm2@cornell.edu