COMMUNICATIONS

Edward A. Galloway and Gabrielle V. Michalek, The Heinz Electronic Library Interactive Online System (HELIOS): Building a Digital Archive Using Imaging, OCR, and Natural Language Processing Technologies

In February 1994, Carnegie Mellon University (CMU) embarked on a project to convert one million pages of the congressional papers of Senator John Heinz (R-PA) into digital format and to provide access to these papers through information retrieval software developed at CMU. To create the HELIOS database, documents are scanned, converted to ASCII form via OCR, verified and organized, and indexed using the CLARIT natural language processing software. The HELIOS project will develop three graphical user interfaces in a Microsoft Windows environment: a scanning interface, an archivist/verification interface, and an end-user interface. HELIOS will allow researchers to access the Heinz papers through the CMU campus network as well as through the Internet.

   o   HTML file

       World-Wide Web:

       http://info.lib.uh.edu/pr/v6/n4/gall6n4.html

   o   ASCII file

       Gopher:

       gopher://info.lib.uh.edu:70/00/articles/e-journals/
       uhlibrary/pacsreview/v6/n4/galloway.6n4

       List Server:

       Send the e-mail message GET GALLOWAY PRV6N4 F=MAIL to
       listserv@uhupvm1.uh.edu.

+ Page 2 +

COLUMNS

Casting the Net

Priscilla Caplan, You Call It Corn, We Call It
Syntax-Independent Metadata for Document-Like Objects

Metadata really is nothing more than data about data; a catalog record is metadata; so is a TEI header, or any other form of description. The goal of the 1995 OCLC/NCSA Metadata Workshop was to define a set of data elements simple enough for authors and publishers to use in describing their own documents as they put them on the Net, but useful enough to facilitate discovery and retrieval of these documents by others. This simple metadata could then also be used by catalogers and other third parties as a starting point for creating more rigorous or more detailed descriptions.

- o   HTML file

    World-Wide Web:

    http://info.lib.uh.edu/pr/v6/n4/capl6n4.html

- o   ASCII file

    Gopher:

    gopher://info.lib.uh.edu:70/00/articles/e-journals/
    uhlibrary/pacsreview/v6/n4/caplan.6n4

    List Server:

    Send the e-mail message GET CAPLAN PRV6N4 F=MAIL to
    listserv@uhupvm1.uh.edu.

+ Page 3 +

Editor-in-Chief

Charles W. Bailey, Jr.
University Libraries
University of Houston
Houston, TX 77204-2091
(713) 743-9804
cbailey@uh.edu

+ Page 4 +

List Server Technical Support

List server technical support is provided by the Information Technology Division, University of Houston. Tahereh Jafari is the primary support person.

Publication Information

The Public-Access Computer Systems Review is an electronic journal that is distributed on the Internet and on other computer networks. It is published on an irregular basis by the University Libraries, University of Houston. There is no subscription fee.

To subscribe, send an e-mail message to listserv@uhupvm1.uh.edu that says: SUBSCRIBE PACS-P First Name Last Name.

Circulation

PACS-L@UHUPVM1.UH.EDU: 9,454 subscribers in 72 countries (PACS-L
is estimated to have 10,000 additional USENET subscribers).
PACS-P@UHUPVM1.UH.EDU: 3,349 subscribers in 61 countries.

Electronic Distribution

Each article is initially distributed in both ASCII and HTML
formats.

ASCII files are paginated. They are available from the following
servers:

- o List Server:

    Send the e-mail message GET INDEX PR F=MAIL to
    listserv@uhupvm1.uh.edu.

- o Gopher:

    gopher://info.lib.uh.edu:70/11/articles/e-journals/
    uhlibrary/pacsreview

HTML files are not paginated. HTML files may have linked GIF files. HTML files may have internal links, external links, or both. The editors do not maintain external links.

HTML files are available from the following server:

- o World-Wide Web:

    http://info.lib.uh.edu/pacsrev.html

+ Page 5 +

In consultation with article authors, the editors determine whether an article is updated, whether both ASCII and HTML files are created for updated articles, and whether all prior versions

of an article are retained.

Print Distribution

The first four volumes of The Public-Access Computer Systems Review are also available in book form from the American Library Association's Library and Information Technology Association (LITA). (Volume five is in process.) To order, contact: American Library Association, Book Order Fulfillment, 155 N. Wacker Dr., Chicago, IL 60606, (800) 545-2433.

Copyright