

Online library gives readers access to 1.5 million books

November 27 2007

The Million Book Project, an international venture led by Carnegie Mellon University in the United States, Zhejiang University in China, the Indian Institute of Science in India and the Library at Alexandria in Egypt, has completed the digitization of more than 1.5 million books, which are now available online.

For the first time since the project was initiated in 2002, all of the books, which range from Mark Twain's "A Connecticut Yankee in King Arthur's Court" to "The Analects of Confucius," are available through a single Web portal of the Universal Library (www.ulib.org), said Gloriana St. Clair, Carnegie Mellon's dean of libraries.

"Anyone who can get on the Internet now has access to a collection of books the size of a large university library," said Raj Reddy, professor of computer science and robotics at Carnegie Mellon. "This project brings us closer to the ideal of the Universal Library: making all published works available to anyone, anytime, in any language. The economic barriers to the distribution of knowledge are falling," said Reddy, who has spearheaded the Million Book Project.

Though Google, Microsoft and the Internet Archive all have launched major book digitization projects, the Million Book Project represents the world's largest, university-based digital library of freely accessible books. At least half of its books are out of copyright, or were digitized with the permission of the copyright holders, so the complete texts are or eventually will be available free.

The collection includes a large number of rare and orphan books. More than 20 languages are represented among the 1.5 million books, a little more than 1 percent of all of the world's books.

Many of the books, particularly those in Chinese and English, have been digitized — their text converted by optical character recognition methods into computer readable text. That allows these books to be searched and, eventually, reformatted for access by PDAs and other devices.

An outgrowth of Reddy's Universal Library, the Million Book Project received \$3.5 million in seed funding from the National Science Foundation and substantial in-kind contributions from hardware and software manufacturers. These funds were primarily used to purchase scanning equipment and for developing the scanning, digitization and cataloguing methods necessary for creating a large digital library.

The vast majority of the scanning, digitization and cataloguing has been performed at centers in China and India, where more than 1.1 million and 360,000 books have been scanned, respectively. The U.S., China and India provided \$10 million each in cash and in-kind contributions to the project. More recently, the Library at Alexandria, Egypt, has joined the effort. Now, about 7,000 books are scanned daily by more than 1,000 workers worldwide.

“We greatly value the participation of Bibliotheca Alexandrina,” said Michael Shamos, a Carnegie Mellon computer science professor and copyright lawyer. “Scholars everywhere regret the destruction of the Alexandria Library at various points in history, and we're willing to go to great lengths to see that no such destruction is ever possible in the future. Once books are on the Internet, they become immortal.”

Protecting and preserving texts is a major goal, said Pan Yunhe, the

leader of the Million Book Project in China. “Paper gets old and brittle, so books soon become so delicate that no one can read them without damaging them,” said Yunhe, the former president of Zhejiang University who is now vice president of the Chinese Academy of Engineering. “Artwork fades. But once we have digitized texts and illustrations, we can keep them in circulation indefinitely. And by storing them at multiple sites, we can minimize the risk that they be destroyed, as occurred in Alexandria.”

“This collection of books in multiple languages opens up unparalleled opportunities to bring Indian cultural material to everyone, and offers a huge range of possibilities in natural language research,” said N. Balakrishnan, associate director of the Indian Institute of Science in Bangalore, one of the partners in the project.

“Digital libraries constitute an essential part of the future of the developing world,” said Ismail Serageldin, director of Bibliotheca Alexandrina. “This requires that we approach conditions governing copyright, digital archiving and scientific databases with a view to creating two-tier systems of access to information that would allow access to such data from developing countries for a nominal fee or for free.”

Though the long-term goal of the Universal Library is to make books, artwork and other published works available online for free, about half of the current collection remains under copyright. Until the permission of the copyright holders can be documented, or copyright laws are amended, only 10 percent or less of those books can be accessed at no cost.

The project has surpassed one million books, but the participants are looking to expand to all countries and eventually every language. At the Third Annual International Conference on Universal Digital Library,

held at Carnegie Mellon Nov. 2-4, 2007, the partners in the Million Book Project agreed to continue scanning, to enlist more centers for the scanning of rare and unique materials, and to work on governmental solutions to the problem of books which are out of print but still in copyright.

Source: Carnegie Mellon University

Citation: Online library gives readers access to 1.5 million books (2007, November 27) retrieved 16 April 2024 from <https://phys.org/news/2007-11-online-library-readers-access-million.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.