

U-M's Entire Library to be Put in Google Billion-Dollar Project Not the First to Digitize Books

by Karin Nead

On December 14, 2004, the University of Michigan announced an agreement with search engine Google under which the complete text of all seven million volumes in U-M's library will be turned into a computer-readable format and made instantly searchable by anyone using the search engine.

Articles run across the country tout this effort as being the largest such digital scanning project ever undertaken, the benefits of which will take online searching beyond the traditional search of web pages and their static information. However, what was not mentioned in these articles is this exact type of project was spearheaded many years ago by Michael Hart in 1971. His brainchild, Project Gutenberg, is the Internet's oldest producer of free electronic books.

Hart began his efforts in response to a huge operator's account he had been given by the University of Illinois—\$100,000,000. The obscenely large account represented the fact that there was more computer time than people knew what to do with, and the operators were encouraged to do whatever they wanted with that time in the hopes they would learn more for their job proficiency. Hart decided the greatest contribution he could make would not be through actual computing, but through the storage, searching and retrieval of information stored in public libraries.

For Google's effort with U-M, digitizing the university's collection is part of an effort called Google Print (<http://print.google.com>), in which the company is working to create digital databases of books, reports, manuscripts and other printed materials. The goal is for Web users accessing the search site to be able to type in a phrase or key words and be presented with direct access to in-depth research and literary material.

This is essentially what Project Gutenberg has been doing for 33 years through its e-texts and eBooks. The information digitized and made available by the Project is unique in that it is available in the simplest, easiest to use forms. 99% of the hardware and software a person is likely to use will be able to read and search these files. No special downloads or programs are required. And, the information is just as easily accessible by someone using a Mac or running UNIX as it is for someone using a PC.

"U-M's Entire Library to be Put in Google" is an original article by Karin Nead.

Copyright 2005 by Canville Communications.

In addition, the books digitized by the Project are those deemed most likely to be researched by the general public. According to the Project Gutenberg website (<http://www.gutenberg.org>), “Project Gutenberg selects etexts targeted a bit on the ‘bang for the buck’ philosophy. . .we choose etexts we hope extremely large portions of the audience will want and use frequently. We are constantly asked to prepare etext from out of print editions of esoteric materials, but this does not provide for usage by the audience we have targeted, 99% of the general public.” The Project Gutenberg eBooks are comprised of light literature, such as *Alice in Wonderland* and *Aesop’s Fables*, heavy literature, such as religious documents and *Moby Dick*, and reference materials like dictionaries and almanacs. Currently, Project Gutenberg publishes texts in 35 different languages, with new languages joining regularly.

The Project produced over 10,000 eBooks by October 2003, making 350 new books available every month, and has mirror sites on every continent, including Antarctica. They’ve been joined in their efforts by Distributed Proofreaders (www.pgdp.net) whose site provides a web-based method of easing the proofreading work associated with the digitization of books into e-books for Project Gutenberg. Distributed Proofreaders has added an additional 6,200 eBooks to the Project.

Even if the national news circuit didn’t take notice of the Project’s accomplishments when announcing the Google effort, the international community has. In 2002, “The Project Gutenberg Literary Archive Foundation” was presented the Stockholm Challenge trophy that symbolizes their technological progress.

It may be the only real news in the U-M/Google announcement is the use of new technology developed by Google that speeds up the digitizing process – which they refuse to say much about. Google also refuses to say how many people will be at the University doing the digitizing work. Project Gutenberg utilizes existing technology and thousands of volunteers to proofread and encourage the creation and distribution of the free eBooks.

Google’s goal with their project has been said to be ambitious. Google says all seven million volumes of the U-M library should be digitized into their database sometime shortly after 2010 at a cost of millions of dollars. In 2002, Project Gutenberg had announced a new goal of getting one million eBook titles to one billion people, for a total of one quadrillion eBooks to be hopefully given away by the end of the year 2015—all for free. In perspective, Project Gutenberg’s efforts can certainly be considered just as ambitious as Google’s.

Karin Nead is the president of Midwest Biz Solutions, Inc. in Belton, Missouri. Her company provides small business owners with outsourcing options for administrative support, bookkeeping, desktop publishing and graphic design.

“U-M’s Entire Library to be Put in Google” is an original article by Karin Nead.

Copyright 2005 by Canville Communications.