

Project Gutenberg

From Wikipedia, the free encyclopedia.

Project Gutenberg (PG) was launched by Michael Hart in 1971 in order to provide a library on the Internet of free electronic versions (sometimes called e-texts) of physically existing books. The texts provided are mostly in the public domain, either because they were never under copyright, or because their copyrights have expired. There are also a few copyrighted texts that Gutenberg has made available with the authors' permission. The project was named after the 15th-century German printer Johannes Gutenberg who propelled the movable type printing press revolution.

For the most part, Project Gutenberg concentrates on historically significant literature and reference works. The slogan of the project is "break down the bars of ignorance and illiteracy", chosen because the project hopes to continue the work of spreading public literacy and appreciation for our literary heritage that public libraries began in the early 20th century. All Gutenberg releases are available in plain ASCII text, and occasionally in other file formats as well. Because maximum availability is a goal, the project eschews prettier but bulkier and not-universally-compatible data formats such as PDF.

All Project Gutenberg texts may be obtained and redistributed by readers for no fee: the only restriction placed on redistribution is that the unaltered text must contain the Project Gutenberg header. If the redistributed text has been modified, the file must not be labelled as a Gutenberg text.

As of 2003, the project has released over ten thousand electronic books, almost entirely produced by volunteers, and remains active. Anyone can become a proofreader by signing up to the Distributed Proofreaders site [1], and volunteering for pages one by one from a variety of texts. While most are in English, there are some in French, Latin, Old English, and a few in other languages. Proofreading does not actually demand knowledge of the language.

History

In 1971, Michael Hart was attending the University of Illinois. Hart obtained access to a Xerox Sigma V mainframe computer in the university's Materials Research Lab, as his best friend and his brother's best friend were two of the four operators of that particular machine. He was given an operator's account with a virtually unlimited amount of computer time; that access has since been variously estimated to have been worth \$100,000 or \$100,000,000. Hart spent the next hour and a half trying to think of something to do with the computer that would be worth that much money. This particular computer happened to be one of the 15 nodes on the computer network that would become the Internet. Hart believed that computers would one day be accessible to the

general public and decided to make works of literature available for free in electronic form. He happened to have a copy of the United States Declaration of Independence in his backpack, and this became the first Project Gutenberg e-text.

By the time U. of I. stopped hosting Project Gutenberg in the mid-1990s, Hart was running it from Illinois Benedictine College. Later he came to a similar arrangement with Carnegie Mellon University, which agreed to administer Project Gutenberg's finances. It was not until the year 2000 that Project Gutenberg was formally organized as an independent legal entity, and it is now a non-profit corporation chartered in Mississippi with an IRS ruling that donations to it are tax-deductible.

Since the Project's early days, the time required to digitize a book has decreased dramatically. Books are generally not typed in, but are instead converted into text with the aid of optical character recognition (OCR) software. Despite these advances, books still need to be heavily proofread and edited before they can be added to the collection.

Other projects inspired by Project Gutenberg

In 2000, Charles Franks founded Distributed Proofreaders, which allows the proofreading of scanned texts to be distributed among many volunteers over the Internet. To make this possible, volunteers scan and run optical character recognition software on books, then place the results on a website for volunteer "proofers" to check. With thousands of volunteers each working on one or more pages, a reasonably-sized book can be proofed in several hours.

The Million Book Project aims to digitize a million public domain books by 2005. In order to process such a large number of books in such a short time, they generally skip the time-consuming transcription process and store their books as compressed image files.

Project Gutenberg
www.gutenberg.net

The Internet Archive an 'Internet library,' with the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format. Founded in 1996 and located in the Presidio of San Francisco, the Archive has been receiving data donations from Alexa Internet and others.
www.archive.org

The Million Book Project the goal of The Million Book Project is to digitize a million books by 2005. The task will be accomplished by scanning the books and indexing their full text with OCR technology. The undertaking will create a free-to-read, searchable digital library the approximate size of the combined libraries at Carnegie Mellon University, and one much bigger than the holdings of any high school library. The project is part of the Internet Archive (see above)

International Children's Digital Library: will be a collection of more than 10,000 books in at least 100 languages that is freely available to children, teachers, librarians, parents, and scholars throughout the world via the Internet
www.icdlbooks.org

Children's Books Online "the largest collection of illustrated antique children's books on line...we think"

www.childrensbooksonline.org

The Rosetta Project is a global collaboration of language specialists and native speakers working to develop a contemporary counterpart of the historic Rosetta Stone. In this updated iteration, our goal is a meaningful survey and near permanent archive of 1,000 languages. Our intention is to create a unique platform for comparative linguistic research and education as well as a functional linguistic tool that might help in the recovery or revitalization of lost languages in unknown futures.
www.rosettaproject.org

The Mutoxia project attempts to do for music what Project Gutenberg does for literary works.

www.mutopiaproject.org

Amazon.com's Search Inside the Book allows you to search millions of pages to find exactly the book you want : your search results will surface titles based on every word inside the book.

www.amazon.com