

The Google Book Search copyright settlement and the future of information—Part 1

K. Reed
12 August 2009

The following is the first of a two-part article on the Google Book Search settlement. The concluding part will be posted August 13.

In George Orwell's *1984*, protagonist Winston Smith secretly obtains a copy of what is known as "the book" in his quest to learn the truth about the dystopian world in which he lives. His inquiry is cut short, however, when he is arrested by the Thought Police for reading the forbidden text.

Paradoxically, if you locate *1984* with Google Book Search and try to read the novel online, you will likely experience an Orwellian moment. After viewing just a few pages, the following warning appears: "You have either reached a page that is unavailable for viewing or reached your viewing limit for this book." While nothing found on Google Book Search is banned (not yet, as far as we know), obviously someone, or more accurately something, is watching you!

As it turns out, Google's copy of Orwell's 1949 classic is a digital duplicate of the 1990 Signet Classic edition and is republished under copyright restrictions. In this case, Book Search tracks the number of pages viewed and, when the limit stipulated by the copyright holder is reached, Google cuts you off. If you wish to find out the fate of our poor Winston Smith, you will have to purchase a copy of *1984* or borrow it from your local library.

Google provides the following explanation: "Many of the books you can preview on Google Book Search are still in copyright, and are displayed with the permission of publishers and authors. You can browse these 'limited preview' titles just as you would in a bookstore, but you won't be able to see more pages than the copyright holder has made available.

"When you've accessed the maximum number of pages allowed for a book, any remaining pages will be omitted from your preview. You can order full copies of any book using the 'Buy this book' links to the right of the preview page." Google also provides links to libraries that can lend you the title.

"Limited preview" of copyrighted works is just one of the ways books can be viewed on Google Book Search. Titles in the public domain are displayed in "full view" and can even be downloaded onto local computers for later offline reading. Two other options are: "snippet view," in which basic information about the book is displayed along with a few snippets of text that show the search phrase in context; and "no preview," in which only the basic information is displayed like in a library catalog.

What is Google Book Search?

Google Book Search is a system that marries the scanned images of printed book pages with the conversion of their content into searchable text. When a book in Google's database is found in a Web

search (using Google, of course), the Book Search item will appear as one of the results in the list. Google promises that its Book Search entry will not appear above the fourth item. In the *1984* example above, a search for "George Orwell 1984" yields the Book Search link as the fifth item.

The company changed the name of its service from the original Google Print to Google Book Search because of the way people actually used the technology. Most often, readers come across a title within Book Search during a standard Web search of a word or phrase. When Google locates the search phrase in the text of one of its online books, the title is shown in the results list. When the title is clicked, the search phrase is highlighted throughout the viewable portion of the book in the Book Search window.

Google's Book Search is no doubt a significant technological achievement. Announced in 2004 in partnership with Harvard, the University of Michigan, the New York Public Library, Oxford and Stanford, and called the Google Print Library Project, it was established to digitize the entire collections of these institutions. The combined number of volumes in these libraries is estimated at more than 15 million titles. As of October 2008, Google reported that 7 million books had been scanned and made available through Book Search.

Indeed, the creation of the universal digital library—access to everything ever written, by everyone, everywhere, at any time—will be a significant social accomplishment. The technical implications were summed up creatively by Kevin Kelly in his May 14, 2006, *New York Times Magazine* article entitled "Scan This Book!": "All this material is currently contained in all the libraries and archives of the world. When fully digitized, the whole lot could be compressed (at current technological rates) onto 50 petabyte hard disks [1 petabyte is equal to 1000 terabytes]. Today you need a building about the size of a small-town library to house 50 petabytes. With tomorrow's technology, it will all fit onto your iPod. When that happens, the library of all libraries will ride in your purse or wallet—if it doesn't plug directly into your brain with thin white cords."

But is this really what Google has in mind? Google's Book Search is not the only online library of printed books. It is distinct from government or non-profit online library projects like American Memory, Project Gutenberg or Internet Archive in that it is privately owned and funded by the most powerful of Internet companies. From its inception, the Book Search project has been a priority for the founders of Google, Sergey Brin and Larry Page. They see it as a critical component of Google's mission to "organize the world's information and make it universally accessible and useful."

As one of the world's largest publicly traded Internet technology

companies with a market capitalization of approximately \$140 billion, Google has been the subject of considerable criticism. Despite the company's efforts to prove itself harmless (the informal corporate slogan is "Don't be evil"), Google has been attacked on a number of fronts: for violations of privacy by tracking and logging user search activity, for collaborating with the Chinese government's censorship and surveillance program and for gathering user data to feed its enormous advertising revenue requirements. It has even been taken to task for the excessive amounts of energy required to power its estimated 700,000 servers with up to 8 petabytes of memory.

Clearly, these aspects of Google's corporate operations are examples of the negative social and political implications of the world's information organizer run on the basis of private ownership and the profit motive. This is what imparts to the Google Book Search system its bittersweet character.

The contradictions embedded in the Book Search technology are examples of a fact of modern life: digital media technologies are outstripping the legal and social forms within which they were born. As with the controversies over sharing recorded music (Napster) and uploading copied videos (YouTube), the inevitable conversion of the entire back catalog of the printed word into digital information on the Internet has thrown the publishing business, the copyright system, the government and Google itself into turmoil.

Our digital world is revolutionizing the things we have up-to-now taken as given in our lives: the book, book publishing and the library. These critically important developments in the history of human culture are being merged together and transformed into something entirely new. As with Gutenberg's invention of the mass production of printing type (and the consequent expansion of literacy) and its intersection with the emergence of capitalist society in the Middle Ages, so too the current digital media revolution represents one of the technological foundations of the emerging socialist transformation of society. However, there looms the very real possibility that this technological achievement will be harnessed for destructive and reactionary purposes instead of being used for progress and the betterment of humanity as a whole.

What are the issues in the 2005 copyright dispute?

Initially, Google's interpretation of copyright law came into conflict with that of the American Association of Publishers (AAP) and the Authors Guild (AG). Both of these organizations filed independent class-action infringement suits in 2005 (*The McGraw-Hill Companies et al. v. Google Inc.* and *The Authors Guild et al. v. Google Inc.*) aimed at halting the Book Search project.

The plaintiffs claimed that Google had violated copyright law by not obtaining authorization from copyright holders of literary works before digitally duplicating their books and putting them online. They additionally sought injunctive relief for what they considered "Google's planned unauthorized commercial use" of the books.

Google defended its program on the grounds that creating searchable copies of books on the Internet constitutes "fair use" and is in keeping with the intent of US copyright law. As explained on their Book Search site: "Copyright law is supposed to ensure that authors and publishers have an incentive to create new work, not stop people from finding out that the work exists. By helping people find books, we believe we can increase the incentive to publish them. After all, if a book isn't discovered, it won't be bought.... That's why we firmly believe that this project is good news for everybody who reads, writes, publishes and sells books."

After nearly two years of behind-the-scenes legal work, an out-of-

court deal was reached on October 28, 2008. Google announced that the settlement with the AAP and AG satisfied the interests of all three parties. The agreement (available [here](#)) establishes a framework for online republication of out-of-print books that are both in- and out-of-copyright as well as in-copyright works that are readily available in bookstores. A compensation structure called the Book Rights Registry is planned for participating authors and publishers who agree to opt-in to the Book Search program.

The settlement defines a book as a published or publicly distributed set of written or printed sheets of paper bound together in a hard copy. It specifically excludes periodicals, unpublished personal papers such as diaries or letters, or works with a large amount of musical notation and lyrics.

Google would have everyone believe that the settlement is very simple and, by satisfying the interests of the AG and AAP, the public interest has also been safeguarded. However, the reality is that the agreement runs directly across the interests of nearly everyone not included in it, especially the public at large. Meanwhile, the settlement is so complicated (134 pages and 15 appendices) that the Federal Court overseeing the case had to extend the deadline for authors and publishers to opt in by an additional four months.

There are no doubt enormous financial interests at stake in the struggle for online publishing supremacy. The American Association of Publishers is an organization that represents a multi-billion-dollar industry with over 300 member companies, including some of the largest publishers in the world such as McGraw-Hill (lead plaintiff in the Book Search lawsuit), Houghton Mifflin Harcourt and Simon & Schuster.

McGraw-Hill is a worldwide leader in the production and distribution of materials in the "financial services, education and business services markets." The corporation had annual sales in 2008 of \$6.3 billion and a net profit of \$800 million. The company Chairman, President and CEO Harold "Terry" McGraw III earned over \$6 million last year.

The book publishing industry has suffered significant losses in recent years due to the combined impact of digital technologies, changes in buying behavior and the economic crisis. For McGraw-Hill that meant a 6 percent decline in sales in 2008 and 21 percent decline in net profit during the same period. The company recently announced 550 layoffs, as it has been hit hard by the reduction of education spending for textbooks as a result of budget cuts.

Learning from the crises of the recorded music and video industries, the book publishers and authors associations are working with Google to ensure the erection of a profitable business model that protects digital assets from illegal duplication. Thus, a paramount concern of the Google Book Search settlement is the prevention of piracy and unauthorized online distribution of copyrighted works.

To be continued



To contact the WSWS and the
Socialist Equality Party visit:

wsws.org/contact