How Big a Problem is Copyright?

Michael I. Shamos, Ph.D., J.D.
Director, Universal Library
Institute for Software Research
Carnegie Mellon University
The Universal Library

- All published works online
- Instantly available
- In any language
- Anywhere in the world
- Searchable, browsable, navigable
- By humans and machines
- Sponsoring the Million Book Project
The Million Book Project

- Major partners: CMU, India, China
- Over 700,000 books scanned
- **China** >400,000 books
  - 18 scanning centers
- **India** >300,000 books
  - 22 scanning centers
- Egypt 20,000 books
- Capacity: 1 million pages per day
  - About 2000 books per day
Google Copyright Lawsuits

• Google News scans newspaper websites, posts digests of news, links to original stories
• Suits against Google News
  – Agence France Presse
  – Copiepresse (Belgium). Google LOST.

• Google Books has arrangements with libraries to scan and index copyrighted material
• Query responses include a page or “snippet” of surrounding text
• Fair use?
• Suits against Google Books:
  – Author’s Guild, McGraw-Hill, Pearson: pending
Outline

• Copyright principles
• International copyright treaties
• Copyright problems
• A radical solution
Classes of Works

- **Public Domain**
- **Out of Print**
- **Free Permission**
- **Paid Permission**
- **Orphans (Owner Unknown, Missing)**
- **In Copyright**
Public Domain

- Free of copyright restriction, available for unrestricted use by anyone
  - Expired copyright
  - Dedication to the public
  - Works “born public,” e.g. U.S. Government works
- Public Domain Enhancement Act (proposed)
  - abandoned works
Copyright Treaties

- **BERNE CONVENTION**
  - 160 NATIONS
  - USA
  - CHINA
  - INDIA

- **WIPO COPYRIGHT TREATY**
  - 56 NATIONS

- **TRIPS (TRADE-RELATED ASPECTS OF INTELLECTUAL PROPERTY)**
  - 148 NATIONS

- **UCC (UNIVERSAL COPYRIGHT CONVENTION)**
  - 64 NATIONS
Fair Use

• Berne Convention, Art. 10: “It shall be permissible to make quotations from a work … provided that their making is compatible with fair practice, and their extent does not exceed that justified by the purpose”

• In the U.S., commercial use is rarely fair unless for criticism or news reporting

• Is Google print fair?

• Not a defense that the use increases sales of the original – that is the copyright owner’s right
Compulsory Licensing

• Owner may not refuse to license. User must pay.
• Fees may be statutory, or set by an independent commission, sometimes case-by-case
• Example: U.S. compulsory license for sound recordings of non-dramatic musical works
  – Per-copy rate (2006): 9.1 cents or 1.75 cents per minute of playing time, whichever is greater
• Japan has an extensive compulsory licensing scheme
• Problem: for public uses, where does the money come from?
Public Lending Right (PLR)

- Book borrowed from library: author receives nothing. Reduces from initial sale only
- In the UK, Government provides a fixed annual pool of funds (since 1979)
- Number of times a book is checked out is counted. Pool divided pro rata to authors: 5.57 pence per loan, up to a maximum of £6,600 per work
- 2006 pool: £6.54M to about 18,500 authors
- Fewer books being checked out!
- No payment for in-library use
- Who benefits? J.K. Rowling (Harry Potter)
Access to Knowledge Treaty (A2K)

• WIPO effort, draft of May 2005
• Purposes
  – overcome differences in wealth, development and access to knowledge
  – reduce technological barriers to accessing information
  – open access to scientific research and data
• Access by Internet search engines is fair use
• Limits digital rights management measures
• No fee for lending
• Extensive compulsory licensing
• Treaty is being blocked by the U.S. and Japan
What Is the Answer?

- Copyright is a major barrier to scanning
- But: copyrights must be respected
- Solution: circumvent copyright legally

- What is really important for agriculture?
  - The information contained in the works
  - Not the works themselves
- But information is not protected by copyright
Key Provisions of TRIPS

• Art. 9, Sec. 2: “Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.”

• Art. 10, Sec. 2: Protection for “compilations of data or other material … shall not extend to the data or material itself.”

• China, India and the US are all TRIPS members

• In general, it is not a copyright infringement to extract information from works
Solution to the Copyright Problem

• Use language technologies such as
  – Test summarization
  – Synthetic document generation
  – Automated translation
    • Real-time speech translation announced last week by Alexander Waibel of CMU (Mandarin – English)
• Scan copyrighted works for their information content
• Build a corpus of information from copyrighted works
• Deliver it through automated translation in a wide variety of languages
Compulsory License

• Berne Convention nations may provide for compulsory licensing of certain types of works (e.g. musical works)

• May NOT provide for compulsory licensing of other types (e.g. cinema)

• Silent on other classes of works (e.g. books)
Berne Convention

• 1886 Treaty (amended many times) subscribed by 160 countries, including China, India & China
• U.S. joined in 1988
• Requires copyright term of at least life + 50 years
• Concept of “national treatment”: every treaty country must give non-nationals the same rights as it accords its own nationals
Public Domain Enhancement Act

- Creates concept of “abandoned work”
- Imposes “maintenance fee” of $1 for each copyrighted work, payable 50 years after first publication (or Dec. 31, 2004, whichever is later) and every 10 years thereafter
- Failure to pay causes copyright to expire
- Similar to patent maintenance fees
- Has chance of passage
Digital Era Copyright Enhancement Act

• H.R. 3048, 105th Congress, introduced November 13, 1997
• Distance learning
  – Allows free distribution of works in digital form to students enrolled in a course
• OK to copy if copying “does not conflict with the normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.”
• No chance of passage
Outline

- **Objective of UDL**
  - universal access to information
- **Where does information reside?**
  - Books, WWW, databases, images, other works, people’s brains
- **Barriers to obtaining information?**
  - Digitization (technical). We can solve technical problems
  - Copyright (legal, social, economic) We can’t solve these
- **Show what a big problem copyright is**
- **Reduce it to a technical problem so we can solve it**
- **Lawsuits**
  - Google, Internet Archive
Outline

• Review classes of works
• International treaties
• Fair use
  – Profit-making uses are likely to be unfair
• Paying for use
  – How to measure?
  – Plans: PLR
  – Where does the money come from?
• Proposal: use the money
Methods of Payment

- Government-paid (e.g. public lending right)
- Subscription model (e.g. HBO)
- Metered use (e.g. page views)
- Free-to-read, pay-to-print
- Compulsory license
Classes of Works

- OUT OF PRINT
- PUBLIC DOMAIN
- FREE PERMISSION
- PAID PERMISSION
- IN COPYRIGHT
Classes of Works

POTENTIAL UDL GENERATED CONTENT: EVERYTHING!

PUBLIC DOMAIN

OUT OF PRINT

FREE PERMISSION

PAID PERMISSION

IN COPYRIGHT