



BIG AMBITION: A Franciscan monk cleans the dust from an ancient book in the library of the St Anto monastery in Bosnia. While Google's present digitisation project is of limited scope, the company makes no secret of its ultimate aim: To scan every book ever published. - AP

DIVIDED OPINIONS ON DIGITISING BOOKS

■ Google has a plan to put millions of books online — but authors argue that such an act contravenes copyright laws.

Sfar pg 16 20.3.07

THE future, generally speaking, isn't the sort of thing that happens in the hushed corridors of Oxford University's Bodleian Library. And superficially, there's something low-tech about the curious performance that has been taking place there every day for some time now.

Each morning, a team of technicians carts piles of books from Oxford's collection of 11 million titles to a nearby building. There, behind closed doors, they are placed on scanning machines.

It is laborious work: Each page must be manually turned. After that, however, the technology kicks in. The data is sent to the United States, to the laboratories of the search engine company Google, where it is converted into virtual words and made searchable.

The quiet diligence of the operation belies the scale of Google's ambitions. The library digitisation project involves universities in the United States and Europe. It may be limited to out-of-copyright books in the Britain at the moment, unlike in the United States, but the company makes no secret of its ultimate aim: To scan every book ever published.

"We think we can do it all inside 10 years," Marissa Mayer, a senior Google executive, told the *New Yorker* magazine

»I hope this world evolves so there exists a time where somebody sitting at a terminal can access all the world's information«

DAN CLANCY,
GOOGLE SCIENTIST

recently. "It's mind-boggling to me, how close it is."

An indication of the significance of Google's undertaking came last week, when it provoked a rare outburst from Microsoft. The library project was "cavalier", systematically violated copyright and was "the wrong path", Microsoft counsel Thomas Rubin said in a speech to the publishing industry in Manhattan.

"Companies that create no content of their own, and make money solely on the backs of other people's content, are raking in billions through advertising revenue," he said.

In the United States, he added, Google had "bestowed upon itself the unilateral right" to copy protected work.

Rubin's remarks dripped with irony, given Microsoft's reputation, and the aura of hippy nonconformism that still clings to Google

three years after it floated on the US stock market with an initial valuation of US\$23bil (RM80.5bil).

But however cynical the motivations, the speech was a clear sign of the ferocity we can expect in the next phase of the war for control of the world's information.

Still offline

The truth is that for all Google's virtuosity, most information isn't online. It's a difficult concept to measure, but some estimates put the proportion not on the Internet at 85%, much of it in books.

This is an obviously intoxicating temptation for a company seeking to organise the world's knowledge — as it has also proved for Amazon, and for Microsoft, which is involved in a similar project with the British Library, focusing on out-of-copyright books. Rubin's speech included an appeal to publishers to team up with his firm instead.

It is already possible, at books.google.com, to search thousands of works, both in and out of copyright, and in many cases to access scanned images of a few pages. On some of them — as critics of the project have delighted in noting — you can see the fingers of the person who scanned them.

But the digitisation initiatives have provoked an angry clash with authors. Google's founders are steeped in the ethos first articulated by hippy futurist Stewart Brand, that "information wants to be free".

Their genius, of course, was to make millions of dollars from it anyway. For many authors, by contrast, copyright is everything.

"It's their only freehold," says historian Antony Beevor, a former chair of the Society of Authors. "As soon as they start giving it away, they'll never get it back."

COPYRIGHT ISSUES

S/91 p8 17

> FROM P16

20.3.07

Legal threats

Google is facing two lawsuits in the United States, one from authors and another from publishers. Google defends itself against charges of copyright theft by arguing that, although it is scanning entire books, it is only making a "snippet" available to each searcher, something it says is permitted under law.

Search engines, the company notes, already rely on this principle. The digitisation project, from this perspective, is just a way of indexing libraries — and a development that promises a massive new marketing opportunity for writers.

"The biggest threat here is ... the threat of not being found," Jens Redmer, European head of the Google books project, said. "99% of all authors will not be best-sellers. Imagine you've written a book on something obscure, like Peruvian orchids. How do you convince a local bookseller to stock it? Because it's only once every 45 years that somebody's going to come into the shop and ask for it."

Add it to *Google Book Search*, instead, and every time somebody Googles "Peruvian orchids", they'll be alerted to your book.

spective, that Google is making money from advertising displayed beside book-search results, even as it spends an estimated US\$800mil (RM2.8bil) on a system that authors fear may starve them of income.

"They say they're doing it for altruistic reasons," Chevalier says, "but why on earth would they put all this money into it if it was pure altruism?"

Publisher HarperCollins, meanwhile, has spent the last 15 months "quite frenetically" digitising its own catalogue, says Jim Green, digital development director. They hope it will allow Internet users to search Google for references in HarperCollins books without Google actually owning a digital copy.

Borderless environment

In principle, Google would face a much tougher climate if it tried to scan copyrighted books in Britain, where the law is far more fierce than that in the United States. For a start, anyone copying a work must be able to prove a legitimate purpose for doing so, such as "reporting current events" or "private study". Google could claim neither.

Not that it will necessarily matter. "It's barmy to have these territorial debates when we're living in this

Piracy concerns

It's easy for authors opposing Google's plan to seem Luddite. But for people to start believing books are a product they don't have to pay for could be disastrous, argues Tracy Chevalier, author and current chair of the Society of Authors.

"It's been so frustrating to watch Google just go ahead without taking copyright into account," she says. "Google are saying that at the moment they'll only reveal a small percentage of any book that's in copyright."

But once the company possesses a full electronic copy, Chevalier and others fear, they may go further — never mind the risk of piracy.

The issue can only become more fraught as technology emerges to make reading an entire book on a computer less painful than at present.

Already, says Chevalier, using Amazon's "search inside" feature, "I was able to read a full Annie Proulx short story without having to pay a cent."

Beevor cites the poet Wendy Cope, who pointed out to him that she was more vulnerable than he was: A "snippet" of one of her books might be a whole poem.

Similar perils lie in wait for the authors of cookery books and other reference works.

Side income

It doesn't help, from this per-

when we're living in this borderless environment," says Uma Suthersanen, a copyright expert at the University of London. It's unclear, for example, what would happen if a British library decided to ship copyrighted works to the United States, to have them scanned there.

Ultimate library

There's little doubt about Google's goal, though. "We are talking about a universal digital library," Dan Clancy, the former Nasa scientist behind Google's book-scanning technology, told the New Yorker. "I hope this world evolves so there exists a time where somebody sitting at a terminal can access all the world's information."

Sir Thomas Bodley, after whom the Bodleian is named, had a related ambition. He wanted to make the library's knowledge "available to the whole republic of the learned". Richard Ovenden, its keeper of special collections, argues that he is only remaining true to that maxim.

"We haven't felt that our reputation has been blackened in any way" by working with Google, he says. "A generation of undergraduates, graduate students and young faculty have grown up with the Internet. It's natural for them to assume that the information is online — and that it's on Google." —

Guardian Newspapers Ltd