

Alain Jacquesson

Google livres et le futur des bibliothèques numériques

Paris, Éditions du cercle de la librairie, 2010, p. 223, € 36,00

Il 10 marzo 2010 è stato siglato lo “storico” accordo tra il Ministero per i beni culturali e ambientali e Google, grazie al quale sarà digitalizzato e reso accessibile online a titolo gratuito un milione di volumi, oggi conservati nelle biblioteche nazionali di Roma e Firenze.

Dell'accordo, di cui si è molto parlato nelle liste di discussione dell'Associazione italiana biblioteche e nella letteratura professionale, è stata messa in rilievo la reciproca convenienza, sia per le biblioteche italiane, sottofinanziate e con collezioni digitali praticamente inesistenti se confrontate ad altre realtà di paesi avanzati, sia per Google, che prosegue a tappe forzate la sua conquista del mercato mondiale della distribuzione delle opere digitali. Ricordiamo che l'azienda di Mountain View ha già firmato una serie di intese con una quarantina di biblioteche, non solo anglosassoni, diventando in pochi anni la più grande biblioteca digitale al mondo, con circa 12 milioni di testi disponibili sulla sua piattaforma.

La discussione in Italia si è soffermata sui vantaggi economici e sugli aspetti giuridici, in particolare sulla clausola di non esclusività nella distribuzione delle opere digitalizzate. Oltre gli aspetti legali e finanziari, l'aspetto più controverso riguarda la strategia complessiva del colosso tecnologico. Qual è la sua posizione sul mercato globale delle piattaforme digitali? E in che modo, per effetto di tale strategia, sono

rimesse in causa le politiche culturali tradizionali in materia di libro e lettura?

A queste domande cerca di dare una risposta Alain Jacquesson, professore all'École de gestion di Ginevra, già direttore della biblioteca di quella città e autore di importanti opere sull'informaticizzazione bibliotecaria. Vale la pena notare che la riflessione di Jacquesson trae spunto da una realtà, quella francese, che ha lavorato in più occasioni con Google, digitalizzando i fondi di alcune importanti biblioteche, come la Municipale di Lione. È proprio in Francia, però, che si sono levate le voci più critiche sul progetto, anche da parte di moderati come il ministro della cultura Frédéric Mitterrand, il quale ha evocato la necessità di un dibattito sereno e "privo di caricature, da non lasciare ai soli specialisti".¹

L'opera di Jacquesson è praticamente esaustiva e non si limita a tracciare la storia del progetto Google Libri e dei suoi aspetti più o meno controversi, ma approfondisce l'esame delle strategie commerciali delle piattaforme digitali e traccia scenari molto radicali, tali da avere un impatto durevole sulla nozione di pubblica lettura.

Nata nel 1998 e dal 2004 capitalizzata in borsa, Google è un'azienda che ha tratto il

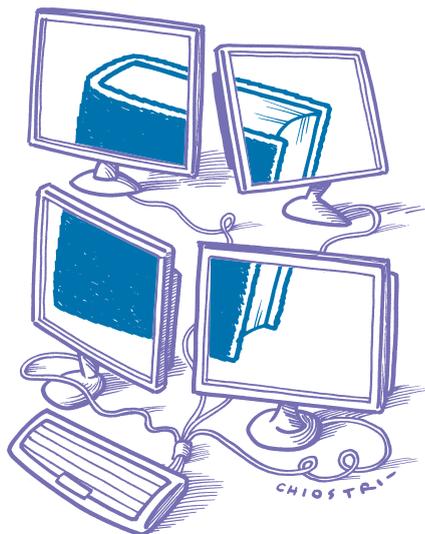
massimo vantaggio dalla sua tecnologia documentale avanzata, che permette ai ricercatori sia il recupero di un'informazione, sia l'accesso ad essa. Di solito si crede che tale ricerca avvenga esclusivamente inserendo parole chiave che sono ricercate dal motore di ricerca in ragione del numero di occorrenze. Questa è certamente la tecnica più nota, ma non è la sola impiegata da Google, che ha lavorato molto alacremente sull'ordinamento dei link ritrovati con algoritmi di ricerca di grande efficacia. Il più famoso di essi è PageRank, fondato sulla popolarità della fonte citata, ma ad esso si affiancano due altri algoritmi: il TrustRank e il FreshRank, basati rispettivamente sui criteri dell'affidabilità e dell'attualità, tesi a migliorare la quantità, la qualità e la tempestività delle citazioni linkate alla risorsa da recuperare.

La *love story* tra Google e le biblioteche comincia nel 2002, quando Sergey Brin e Larry Page, i due fondatori di Google, creano all'interno dell'azienda una piccola équipe incaricata di realizzare un progetto di digitalizzazione di libri, da svolgere in collaborazione con la Biblioteca dell'Università del Michigan. Secondo i bibliotecari di quell'università, la digitalizzazione del patrimonio bibliote-

cario (7 milioni di volumi) avrebbe richiesto un intero millennio e molte risorse. Davanti a un uditorio sbalordito, Page sostenne invece che la compagnia avrebbe digitalizzato a sue spese quello stesso patrimonio di volumi, il tutto in sei anni!

In seguito, e in modo molto discreto, Google strinse accordi con Harvard, Oxford, con la New York Public Library e la Bodleian Library. Il *big bang* avvenne però nel dicembre 2004, quando Google annunciò la digitalizzazione di sei milioni di volumi, in collaborazione con le biblioteche sopra citate. Negli ultimi cinque anni il progetto ha marciato a ritmi serrati, sia in termini di realizzazioni – 12 milioni di volumi nel 2009 – sia per numero di accordi stipulati con altre biblioteche: nel 2005, la Library of Congress, nel 2007 la Biblioteca di Gand in Belgio, la Bayerische Staatsbibliothek e la Biblioteca cantonale di Losanna, nel 2008 la Biblioteca municipale di Lione e altre 28 biblioteche, tra cui importanti istituzioni in Spagna e in Giappone, nonché un centinaio di biblioteche dell'University of California, nel 2009, un gran numero di biblioteche statunitensi, tra cui quelle delle università di Princeton, di Cornell, di Austin e di Virginia. Nel 2010, infine, c'è stato l'accordo non con una singola biblioteca, ma con uno stato, l'Italia, che prevede la digitalizzazione potenziale del patrimonio delle biblioteche storiche italiane. Quali sono i punti di forza del progetto Google Libri? In che modo l'azienda statunitense riesce a realizzare in modo tanto rapido ciò che altre istituzioni (ad esempio, Europea) realizzano lentamente e faticosamente, con programmi articolati nello spazio di decenni?

La prima innovazione riguarda, com'è ovvio, le tecniche di scannerizzazione. Nonostante i progressi tecnici e per quanto sorprendente possa sembrare, la quasi totalità della digitalizzazione effettuata nell'ambito dei progetti Google è realizzata manualmente. Le pagine del volume sono infatti controllate da un operatore. La tecnica utilizzata è stata secretata e coperta da un brevetto depositato nel 2004 e reso pubblico cinque anni dopo. Da esso si apprende che, rispetto ai metodi già sperimentati nelle biblioteche, l'originalità della procedura risiede nella presenza di un proiettore e di un dispositivo di trasmissione dell'immagine al software di riconoscimento dei caratteri. In altri termini, il procedimento è del tutto tradizionale e la sola novità è rappresentata dal fatto che la curvatura del libro è già parametrata all'interno del dispositivo utilizzato dall'apparecchio fotografico. Altra novità è l'efficienza del processo di lavorazione, dove la successione in lotti della selezione dei documenti e della fornitura a Google per la scannerizzazione è pienamente ottimizzata, escludendo incertezze e tempi morti. Una volta fabbricate, le opere devono essere rese disponibili. La distribuzione avviene attraverso la homepage di Google: l'opera è offerta a testo pieno con tecniche di ricerca booleana in cui, sebbene manchi la gestione della funzione di "Rinvio", si ritrovano tutte le occorrenze della parola scelta. La visualizzazione delle opere non è omogenea e comprende sia il testo intero, sia degli estratti, sia anche, per le opere coperte da diritto d'autore, una porzione limitata di testo o delle "nuvole" di parole che descrivono il contenuto. A vol-



te è anche possibile il rinvio a Google Maps per la visualizzazione dei luoghi citati nell'opera. Se il testo è coperto dal diritto d'autore, è possibile comprare il libro in una libreria elettronica associata a Google o localizzare il volume in biblioteca, attraverso il ricorso al catalogo collettivo OCLC.

Com'è noto, OCLC è un gigante che raggruppa 71.000 biblioteche fornitrici di registrazioni bibliografiche da 112 paesi: il catalogo Worldcat conta ormai 170 milioni di registrazioni. Nel 2008 OCLC e Google hanno firmato un accordo secondo cui il primo offre al secondo le registrazioni bibliografiche e riceve in cambio la possibilità di collegare le registrazioni del suo catalogo alle opere scannerizzate. È proprio questo accordo il segno di una ulteriore rivoluzione, questa volta nel campo della lettura pubblica, come Jacquesson non manca di sottolineare. Siamo infatti molto vicini all'utopia del controllo bibliografico e della disponibilità universale dei documenti, come la elaborarono nel 1910 Paul Otlet e Henri La Fontaine. Siamo insomma entrati nel Mundaneum contemporaneo, la città contemporanea del sapere che ospita, come il suo antenato, dodici milioni di opere e permette al pubblico di ricevere risposta a ogni domanda che si ritiene di formulare. L'utopia si è insomma realizzata e ha il volto di un motore di ricerca.

Quale sarà allora il futuro riservato alle biblioteche? Non è proprio la forza di tale accordo a mettere in ombra il loro ruolo? Le controversie sollevate dall'iniziativa della biblioteca digitale più importante al mondo – sostiene Jacquesson – sono importanti perché segnano non so-

lo la cornice materiale di un progetto, ma anche il quadro giuridico, amministrativo e finanziario entro cui rischiano di dovere operare le biblioteche del futuro, se oltre ad essere istituzioni della memoria, aspirano ad essere agenti attivi nella circolazione del sapere.

Vediamo ora quali sono le tendenze controverse nel progetto Google Libri. Prima di tutto: il diritto d'autore. Disponendo di grandi capitali, Google ha fatto una scelta corsara, ma azzeccata dal punto di vista commerciale: prima si scannerizza, poi si regolano le questioni di fondo sul diritto d'autore. La scelta era fondata sul tenue argomento del *fair use*, la possibilità, prevista nel diritto anglosassone, di riprodurre in quantità "ragionevoli" e a fini non commerciali le opere tutelate, salvo poi ricorrere all'*opting out*, con l'esclusione dal catalogo, nel caso si fossero fatti vivi gli aventi diritto. Una scelta corsara, perché Google Libri è tutt'altro che un'iniziativa non commerciale, anche se le sue entrate sono indirette e legate alla pubblicità. Il vantaggio di tale scelta ha comportato la disponibilità di ben due milioni (su dodici) di opere "orfane", opere cioè di cui non è più possibile rintracciare gli aventi diritto, e queste si sono aggiunte ai 2,5 milioni di opere libere da diritto. Sulle opere coperte dal diritto d'autore l'atteggiamento di Google prometteva tempesta e la pioggia di ricorsi non si è fatta attendere.

Il primo processo ha avuto luogo nel 2005, con la class action promossa contro Google dalle associazioni americane degli autori da un lato, e degli editori dall'altro (in collaborazione con alcuni grandi gruppi editoriali indipendenti). Parallelamente, e con lie-

ve ritardo, prendevano corpo i ricorsi in tribunale delle associazioni degli editori di Francia e Belgio.

Dopo tre anni di costosi processi si addiuvano a un accordo, per mezzo del quale Google versa circa 125 milioni di dollari per finanziare un Books Rights Registry, dove vengono iscritte le opere per le quali gli editori vantano crediti derivati dal copyright. Nel frattempo, si susseguono i contratti con editori di spicco: ben 25.000, secondo un comunicato stampa del 2009, tutti favorevoli ad utilizzare Google come propria piattaforma digitale. Nella proposta di accordo le opere tutelate sono consultabili su Google solo per il 20% del loro contenuto; la ripartizione dei proventi pubblicitari è per due terzi dovuta agli aventi diritto e per un terzo a Google. Le biblioteche possono avere accesso a una sola copia per ogni 4000 iscritti con la possibilità di effettuare repliche a tariffa equa. Si tratta solo di una proposta di accordo, che non trova però tutti gli attori consenzienti: le biblioteche temono il monopolio di Google e le limitazioni all'accesso; uguali reticenze sono da ravvisare nel campo dei fornitori sul mercato di contenuti digitali (le piattaforme concorrenti Amazon, Yahoo e Microsoft). Dopo che anche il Dipartimento di Stato americano ha respinto l'accordo, si attendono le conclusioni del giudice.

Google Libri non è solo biblioteca digitale, ma anche editoria, in due settori di attività. Da un lato, vi è infatti la pubblicazione di opere fuori diritto, di classici scannerizzati con riconoscimento ottico dei caratteri e messi a disposizione dei lettori grazie a tecniche di print-on-demand. Le case editrici

che rieditano i classici hanno così perso il loro mercato naturale. Ma è soprattutto nella seconda attività, la produzione di e-book, che Google ha sfondato, riuscendo a offrire contenuti a Amazon per Kindle, a Sony per Reader e ad Apple per iPad. Per quanto riguarda il libro elettronico, non c'è dubbio che buona parte della partita del futuro si giocherà sugli standard, siano essi comuni – PDF, RTF, ePub – o proprietari, come quelli di Amazon o Barnes & Noble.

Per questo il punto maggiormente controverso delle strategie industriali di Google Libri è il suo peso nella diffusione dei contenuti e l'impatto esercitato sulle politiche culturali nazionali. Secondo il minuzioso calcolo di Jacquesson, se è vero che il costo della digitalizzazione di 10 milioni di opere è stato di 0,74 euro a pagina, pari cioè a 2,01 miliardi di dollari, l'investimento globale rappresenta quindi né più né meno che i benefici di un solo semestre della compagnia americana. Non sono dunque né le imperfezioni tecniche, né le questioni legali i punti controversi del progetto. Come sostengono sia Jean-Noël Jeannevey, sia Robert Darnton, puntualmente citati da Jacquesson, la vera posta in gioco è l'accentramento delle risorse e dei contenuti in una sola grande compagnia e la possibilità di detenere il monopolio sull'accesso all'informazione.

Va ricordato che Google Libri non è l'unico attore fornitore di opere digitali; molti altri editori, compagnie, e persino enti pubblici forniscono lo stesso servizio. Si può però parlare di posizione dominante e la reazione del Dipartimento di Giustizia americano marcherà un

punto essenziale nella controversia in corso.

Questo enorme concentrato di risorse e di contenuti influenza fortemente il futuro delle biblioteche digitali. È vero che, anche in presenza di un concorrente-cooperante così ingombrante, le funzioni bibliotecarie rimangono quelle di selezionare, acquisire, descrivere ed offrire degli spazi di tranquillità ai lettori perché possano consultare le collezioni sia fisiche, sia digitali. È vero anche che il catalogo di Google è poco performante e i dieci milioni di opere sono intasati in un deposito digitale alla rinfusa con il risultato che molti sono i doppioni e le ridondanze con altre biblioteche digitali. Tuttavia, lo sviluppo delle biblioteche digitali è una strada troppo importante per essere lasciata nelle mani di un solo attore privato.

Sono così nate iniziative parallele: il ramo “testo” dell’Internet Archive, lo Hathi Trust, associazione delle più grandi università del Midwest americano che promette di offrire un deposito di 18 milioni di opere digitali entro il 2012, per non parlare dei progetti in corso in vari paesi (in Francia, Gallica), di Europea, la rete europea di opere digitali e di altre reti di carattere settoriale.

Il grande stacco di Google nei confronti delle biblioteche è quello di avere fatto della digitalizzazione delle opere un’impresa industriale, con metodi di ottimizzazione degni di una grande impresa e soprattutto di avere posto il progetto di digitalizzazione al centro della strategia industriale dell’impresa, tra pubblicità, portale di ricerca, collezione di mappe e i numerosi altri servizi offerti dal gigante di Mountain View.

Che lo si voglia o no – e Jac-

quesson non manca di sottolinearlo – Google è diventato un attore inevitabile del paesaggio bibliotecario contemporaneo.

Giuseppe Vitiello

Nato Defense College, Roma
g.vitiello@ndc.nato.int

¹ Frédéric Mitterrand ha richiesto il 21 ottobre 2009 un rapporto a Marc Tessier, già presidente di France Télévisions, pubblicato nel dicembre 2009.