

TEMI E ANALISI

Il falso in Internet: autorevolezza del Web, information literacy e futuro della professione

SECONDA PARTE

di Carlo Bianchini

Motore uguale editore

Come utenti che navigano quotidianamente sul Web e accedono alle risorse tramite i motori di ricerca, non siamo abituati a considerare i motori di ricerca ciò che sono realmente:

I motori di ricerca sono i nuovi editori del Web. Se il ruolo di un editore è quello di selezionare e proporre in forma strutturata e organica i contenuti, allora possiamo affermare che in rete sono i motori di ricerca a svolgere questo ruolo, partendo direttamente dalle richieste del lettore¹.

Analogamente, si possono considerare i motori di ricerca come editori anche in quanto imprenditori dell'industria dell'informazione: per esempio Google – un'azienda le cui quotazioni in borsa sono costantemente in crescita – ottiene soldi dalla pubblicità, con tre diversi prodotti editoriali: *Google AdWords*, *Google AdSense* e *Google for Advertisers*².

Nell'affrontare il tema del falso sul Web ci si deve chiedere se un motore di ricerca, per esempio Google, possa dire il falso. Google, per il suo funzionamento tecnico e per le sue politiche editoriali, declina in sé diverse accezioni del termine falso:

CARLO BIANCHINI, Università degli Studi di Pavia, Dipartimento di Musicologia e Beni culturali, corso Giuseppe Garibaldi 178, 26100 Cremona, e-mail carlo.bianchini@unipv.it.

Ultima consultazione siti web: 25 giugno 2014.

Questo testo è la seconda parte del documento con lo stesso titolo pubblicato nel precedente fascicolo della rivista (v. 54, n. 1, p. 243-256).

1 Letizia Sechi, *Editoria digitale: linguaggi, strumenti, produzione e distribuzione dei libri digitali*. Milano: Apogeo, 2010.

2 Stefania Spatti, *Google batte le attese sugli utili, non sui ricavi: brillano i conti Ibm*, «Il Sole 24 Ore», 22 gennaio 2013, <<http://www.ilsole24ore.com/art/finanza-e-mercati/2013-01-22/google-utili-225850.shtml>>.

AIB studi, vol. 54 n. 2/3 (maggio/dicembre 2014), p. 227-240. DOI 10.2426/aibstudi-10130.



- a) È ingannevole, perché dà l'impressione di cercare sull'intero Web (o una sua parte molto consistente)³ e perché non è imparziale;
- b) È infondato, perché il recupero e l'ordinamento dei risultati non si basano sulla semantica (come Google dichiara e l'utente è indotto a pensare) e perché manipola i risultati della ricerca;
- c) È fraudolento, perché raccoglie e vende dati sugli utenti, anche a loro insaputa.

Anche se non è possibile analizzare in modo dettagliato tutti questi aspetti, occorre sottolineare da un lato che, quando usiamo i prodotti messi a disposizione gratuitamente da Google, la politica sulla privacy e altri aspetti evidenziati sono resi espliciti nelle pagine di aiuto di Google; dall'altro che manca ancora la necessaria consapevolezza di questi aspetti critici negli utenti (e in molti bibliotecari) e che diffondere tale consapevolezza è un compito fondamentale del ruolo di istituzioni democratiche per il *life-long learning* che le biblioteche svolgono.

La privacy e il valore potenziale dei nostri dati sono un problema molto sottovalutato: tutti noi che non leggiamo le clausole sulla privacy e clicchiamo su 'accetto' per tutti i servizi Google ma anche per tutte le app del cellulare tendiamo a dimenticare che i dispositivi elettronici sono soprattutto *memorie*. Siamo indotti a ritenere che lasciare libero accesso ai nostri dati personali sia innocuo per noi stessi e che, a maggior ragione, sia inoffensivo l'uso dei nostri dati insieme a quello di milioni di altri utenti. L'analisi di consistenti quantità di dati disponibili in linea consente invece di ricostruire aspetti privati della persona (orientamento sessuale, gruppo etnico, convinzioni religiose, opinioni politiche, grado di intelligenza, livello di realizzazione personale, uso di sostanze che generano dipendenza, età e genere) anche senza che la singola persona li abbia resi espliciti⁴.

3 Se si considera il servizio fornito da *Google Scholar*, da un punto di vista biblioteconomico emerge il problema dell'analisi della qualità di accesso all'informazione garantita. A fronte di un bacino potenziale di informazioni molto ampio (preprint, tesi, libri, abstract e articoli di case editrici accademiche, ordini professionali, database di studi non ancora pubblicati, università e altre organizzazioni accademiche), l'ambito della banca dati di *Google Scholar* non è affatto definito, ma una caratteristica fondamentale delle bibliografie, irrinunciabile sotto il profilo metodologico, è la definizione precisa del loro ambito e scopo. Ciò garantisce al lettore di conoscere con precisione quali fonti può considerare consultate (perché incluse nella bibliografia) e quali altre fonti dovrà ancora consultare per considerare completa la ricerca delle fonti. In *Google Scholar* non è possibile sapere quali banche dati sono già state consultate e quali no; probabilmente non sono in grado di saperlo nemmeno i costruttori di Google (come non sanno indicare con precisione quale percentuale di Web sia indicizzata rispetto al totale). Altrettanto grave è la mancanza di indicazione di eventuali criteri di selezione, per l'inclusione o l'esclusione degli articoli dalla banca dati, anche quando i criteri sono presenti e validi (per esempio, come quando si prevede che l'autore possa eliminare dall'elenco delle proprie pubblicazioni articoli erroneamente attribuiti a lui). Con questi presupposti, non stupisce che gli indici bibliometrici offerti da Google (H-index) possano essere inefficaci e fuorvianti. Cfr., per esempio, Emilio Delgado López-Cózar; Nicolás Robinson-García; Daniel Torres-Salinas, *The Google Scholar experiment: how to index false papers and manipulate bibliometric indicators*, «Digital Libraries», 10 settembre 2013, <<http://arxiv.org/abs/1309.2413>>; Leslie S. Adriaanse; Chris Rensleigh, *Web of Science, Scopus and Google Scholar: a content comprehensiveness comparison*, «Electronic Library», The 31, n. 6, 15 novembre 2013, p. 727-744, doi: 10.1108/EL-12-2011-0174.

4 Michal Kosinski; David Stillwell; Thore Graepel, *Private traits and attributes are predictable from digital records of human behavior*, «Proceedings of the National Academy of Sciences of the United States of America», 110 (2013), n. 15, p. 5802-5805, doi: 10.1073/pnas.1218772110.

Ma anche rilasciare i nostri dati non è affatto privo di conseguenze o di pericoli: è reale l'esempio delle ditte che richiedono di accedere al profilo Facebook di un candidato all'assunzione o della trasformazione da parte di Facebook dei profili 'segreti' in profili pubblici⁵; ma quale potrebbe essere l'uso dei nostri dati personali o collettivi, in uno scenario politico non più democratico? I dati, come il tempo, non tornano più. È un problema che andrebbe approfondito.

Almeno sulla seconda accezione di falso per Google è necessario un approfondimento. A metà 2013 Google ha rilasciato una nuova versione del proprio algoritmo – denominata *Hummingbird* – che sarebbe in grado di 'conversare' con l'utente, in modo tale da consentirgli di formulare domande in linguaggio naturale come «Dimmi chi sono gli impressionisti» o «Qual è il miglior posto per comprare una TV vicino a dove mi trovo?»⁶.

Se si effettua una prova di interrogazione con entrambe le stringhe suggerite, i risultati non sono quelli dichiarati: nel primo caso si ottengono risultati più pertinenti e più rilevanti con la sola stringa di interrogazione 'Impressionisti'; nel secondo caso si ottengono generici 'consigli per gli acquisti', ma non una risposta pertinente o rilevante.

L'algoritmo di Google si basa ancora semplicemente sull'analisi della stringa e non su un approccio semantico: la stringa di ricerca '*click here*'⁷, che ogni addetto al reference rifiuterebbe come *non sense*, su Google genera oltre 3 miliardi e mezzo di risultati, con tanto di *ranking*!

Secondo uno studio condotto in aprile 2011 dal sito *Searchenginewatch.com*⁸, i primi tre siti elencati nelle risposte di una ricerca su Google ricevono il 58,4% delle visite da parte di chi ha fatto la ricerca. Più precisamente il primo classificato riceve una media di 36,4% delle visite (*click-through rate* – CRT), il secondo il 12,5% e il terzo il 9,5%.

Se interroghiamo Google.it con la stringa 'Barack Obama' otteniamo oltre 261 milioni di risultati. Come professionisti dell'informazione ci dobbiamo porre seriamente la questione di come una macchina (pur complessa, come è un motore di ricerca) proceda per selezionare e ordinare – tra milioni di siti Web – la risorsa che ottiene la prima, la seconda o la terza posizione (*rank*). Anche se per assurdo si suppone che un navigante verifichi i primi venti, cinquanta o cento siti, sono comun-

5 Erica Swallow, *How recruiters use social networks to screen candidates*, «Mashable.com», 2011, <<http://mashable.com/2011/10/23/how-recruiters-use-social-networks-to-screen-candidates-infographic/>>; Francesco Lanza, *Facebook: cosa controllano i datori di lavoro prima di assumerti?*, «Www.downloadblog.it», 2012, <<http://www.downloadblog.it/post/17291/facebook-cosa-controllano-i-datori-di-lavoro-prima-di-assumerti>>; *Addio ai profili 'segreti' su Facebook*, «Www.lastampa.it», 10 ottobre 2013, <<http://www.lastampa.it/2013/10/14/tecnologia/addio-ai-profil-segreti-su-facebook-HdLcdF3CdeRrKjESHSiD3/pagina.html>>.

6 Amit Singhal, *Fifteen years on - and we're just getting started*, «Inside Search. The Official Google Search Blog», 2013, <<http://insidesearch.blogspot.it/2013/09/fifteen-years-onand-were-just-getting.html>>; Nino Grasso, *Google annuncia il nuovo algoritmo di ricerca Hummingbird*, «Quotidiano.net», 27 settembre 2013, <http://www.hwfiles.it/news/google-annuncia-il-nuovo-algoritmo-di-ricerca-hummingbird_48882.html>.

7 I risultati che si ottengono sono indirizzi di pagine che contengono rinvii al software Adobe Reader (come risulta dalle prime risposte) o link a siti per i quali non si è specificato il valore semantico (cfr. <<http://www.w3.org/QA/Tips/noClickHere>>).

8 Cfr. <<http://searchenginewatch.com/article/2049695/Top-Google-Result-Gets-36.4-of-Clicks-Study>>.

que quasi 261 milioni i siti materialmente esclusi dalla vista dell'utente⁹. I sistemi di recupero dell'informazione non si valutano sulla precisione oltre che sul richiamo?

Per ora, al problema del *ranking* dei risultati della ricerca si sono interessate molto di più le aziende commerciali, che offrono alle aziende private servizi di *Search Engine Optimization* (SEO) per i loro siti: la SEO consiste in un insieme di tecniche e di procedure per migliorare la posizione di un sito nei risultati dei motori di ricerca¹⁰. Anche se un motore di ricerca garantisce di recuperare tutte le pagine del Web che ha indicizzato con le parole chiave della ricerca, il criterio della rilevanza è tutt'altro che imparziale o oggettivo: la rilevanza nei risultati dei motori di ricerca può essere modificata e, nella prospettiva del 60% di CTR sui primi tre siti, diventa un fattore di selezione di vitale importanza.

Il problema è più complesso di così: da fine 2009, Google ha introdotto la *profilazione* degli utenti, che consiste in una loro classificazione in base a parametri come età, sesso, interessi, comportamenti in rete ecc. Un'importante conseguenza è che i risultati dell'interrogazione sono diversi a seconda del profilo di chi la compie:

Google usa 57 indicatori – dal luogo in cui siamo al browser che stiamo usando al tipo di ricerche che abbiamo fatto in precedenza – per cercare di capire chi siamo e che genere di siti ci piacerebbe visitare. Anche quando non siamo collegati, continua a personalizzare i risultati e a mostrarci le pagine sulle quali probabilmente cliccheremo. Di solito si pensa che facendo una ricerca su Google tutti ottengano gli stessi risultati: quelli che secondo il famoso algoritmo dell'azienda, PageRank, hanno maggiore rilevanza in relazione ai termini cercati. Ma dal dicembre 2009 non è più così. Oggi vediamo i risultati che secondo PageRank sono più adatti a noi, mentre altre persone vedono cose completamente diverse. In poche parole, Google non è più uguale per tutti¹¹.

Oltre ai possibili effetti su rilevanza, pertinenza, precisione e richiamo – e tralasciando l'enorme problema della censura – si pone anche un problema etico. Come ha osservato Tara Brabazon: bisogna chiedersi quanto gravi possono essere le conseguenze quando, per esempio, uno studente clicca su siti fortemente ideologici che raggiungono una posizione molto alta nelle risposte, sulla base dell'alta popolarità e non sulla base della qualità, della rilevanza o della veridicità delle informazioni che propongono. I siti con una popolarità più bassa, ma che forse contengono le informazioni più critiche e approfondite, sono molto più indietro nella classifica¹².

Google meriterebbe maggiore attenzione da parte del mondo dei bibliotecari anche per altre ragioni: nel rapporto conclusivo di una ricerca condotta dall'Università Tecnologica di Graz vengono messi in evidenza i possibili rischi per la comunità scientifica che derivano da un motore di ricerca come Google:

– la posizione di monopolio *de facto* impedisce la verifica – con strumenti alternativi ed equivalenti – dell'esistenza di altre informazioni di qualità equivalente o superiore;

⁹ Anita Paz, *In search of meaning: the written word in the age of Google*, «JLIS.it», 4, n. 2, 1 luglio 2013, p. 260, doi: 10.4403/jlis.it-8798.

¹⁰ La SEO dei siti commerciali sta diventando un vero e proprio business. Si veda per esempio il sito <<http://www.seoutilily.com>>.

¹¹ Eli Pariser, *Quello che Internet ci nasconde*, «Internazionale», (2011), <<http://www.internazionale.it/news/internet/2011/07/06/quello-che-internet-ci-nasconde-2>>.

¹² Tara Brabazon, *The university of Google: education in a (post) information age*. Aldershot: Ashgate, 2007.

– la funzione di personalizzazione delle ricerche, servizio obbligatorio fornito da Google, viola il principio di uguale accesso all'informazione per tutti sancito, per esempio, dal Manifesto Unesco sulle biblioteche pubbliche;

– la personalizzazione delle ricerche si basa sulla violazione – sistematica ed estensiva – della privacy¹³. Tutte le attività e i comportamenti di qualsiasi utente vengono monitorati e registrati (e venduti), per offrire pubblicità mirate;

– i risultati delle ricerche sono influenzati dall'analisi dei dati degli utenti (*data mining*).

Sono quindi diversi i modi con i quali Google – in quanto motore di ricerca – può privilegiare, omettere, censurare, modificare le informazioni, ovvero manipolare i risultati di una ricerca. Si tratta di falso ideologico, sul quale i professionisti dell'informazione devono vigilare con la massima attenzione, perché ogni manipolazione delle informazioni è riconducibile alla violazione di un principio che, nel mondo delle biblioteche, è sempre stato considerato intoccabile: il principio di imparzialità.

Google è una società multinazionale e non basa necessariamente le proprie scelte su valori fondamentali per il servizio bibliotecario come la libertà intellettuale, la privacy, l'equità nell'accesso alla conoscenza e all'informazione, il diritto a ricercare l'informazione che si preferisce, la democrazia¹⁴.

Per chiudere questa necessariamente breve panoramica su Google, occorre ricordare anche il fenomeno del *copyfraud*, ovvero «le manovre finalizzate a defraudare il pubblico della possibilità di utilizzare liberamente il materiale di dominio pubblico», per il quale si rinvia ai lavori di Alberto Petrucciani¹⁵.

Per comprendere quanto questo fenomeno richieda estrema attenzione, non solo dal punto di vista bibliotecario ma anche editoriale, è utile richiamare quanto dichiarato da Carlo De Benedetti al 66° World Newspaper Congress, il forum mondiale degli editori di giornali del globo, tenutosi a Torino dal 9 all'11 giugno 2014: «Google fa paura. Non a me, ma a tutti. Stiamo attenti, perché con le oligarchie digitali è a rischio la democrazia»¹⁶.

Abbiamo abbastanza paura di Google?

13 Le informazioni per costruire i profili personali dei naviganti vengono raccolte con molti sistemi diversi: analisi delle ricerche effettuate sui motori di ricerca, cookies, traffico email, acquisti online, informazioni dai social network (Facebook prima di tutto), sistemi di informazione geografica (GIS), aggiornamenti dei software, software appositi (*troians* o cavalli di Troia), opzioni del browser, strumenti di analisi dei siti (come *Google Analytics*), *data mining* sui server. Apparentemente ciascuna di queste tecniche appare relativamente innocua. Ma si deve pensare quanto possa essere potente l'incrocio di dati come il tipo di ricerche che si fanno sui motori di ricerca, i tipi di acquisti (per esempio, la registrazione sistematica e duratura di quali libri si acquistano), con i profili delle persone con le quali abbiamo scambi di informazioni via e-mail o Facebook. Non rimane spazio al dubbio che ci sia qualche aspetto della nostra privacy che rimane tutelato.

14 I riferimenti sono molti; siano sufficienti qui i seguenti tre: AIB, *Codice deontologico dei bibliotecari*, 1997, <<http://www.aib.it/aib/cen/deocod.htm>>; Michael Gorman, *I nostri valori: la Biblioteconomia nel 21. secolo*, a cura di Mauro Guerrini, Agnese Galeffi, Carlo Ghilli. Udine: Forum, 2002; Unesco, *Manifesto UNESCO per Le Biblioteche Pubbliche*, a cura di Maria Teresa Natale, «AIB Notizie», 7 (1995), n. 5, p. 1–2, <<http://www.aib.it/aib/commiss/cnbp/unesco.htm>>.

15 Alberto Petrucciani, *Con Robert Darnton nella giungla del digitale*, «Bollettino AIB», 51 (2011), n. 1–2, p. 101–102, <<http://bollettino.aib.it/article/view/4978>>; *Id.*, *Ancora su Google e la giungla digitale: altri misteri, novità e tendenze*, «AIB Studi», 52 (2012), n. 2, p. 197, doi: 10.2426/aibstudi-6326.

16 *Google fa paura a tutti*, «Il Messaggero Veneto», 12 giugno 2014.

Valutare le informazioni

Nell'analisi delle trasformazioni in atto è fondamentale la distinzione fra gli interessi culturali (propri delle biblioteche e sempre più dei lettori), che includono questioni relativi alla bibliografia come «scienza degli indici»¹⁷ e gli interessi commerciali (propri delle banche dati e dei motori di ricerca), che includono questioni relative all'informazione come valore economico, quindi manipolabile e contrattabile.

Nell'ideale filo rosso che collega l'autore al lettore, l'autorevolezza – intesa come garanzia del valore culturale complessivo delle informazioni trasmesse – si è spostata dall'autore al lettore. Nell'antichità era l'autore a garantire per il valore delle conoscenze che trasmetteva direttamente ai suoi discepoli. Nell'era della stampa manuale e industriale, un ruolo essenziale di autorevolezza è assunto dall'editore, che può decidere di fare dell'autorità un valore aggiunto dei propri prodotti. Con il passaggio progressivo alla circolazione delle informazioni in rete e con la disintermediazione, l'onere del controllo di qualità delle informazioni, la verifica dell'autenticità e dell'autorevolezza si è spostata dal momento della creazione (autore) o della circolazione (editore, comitato editoriale, *peer reviewing*) al momento della fruizione (lettore). Ciò significa che sono cambiate profondamente le modalità del controllo che, nella comunicazione in rete, è a valle della pubblicazione ed è molto di più a carico del lettore.

Per utilizzare in modo corretto l'informazione disponibile su Internet i lettori devono essere preparati, devono acquisire specifiche competenze: nel nuovo contesto è più che mai necessario educare all'*information literacy*, ovvero alla competenza nella gestione corretta delle informazioni. Secondo il CILIP (*Chartered Institute of Library and Information Professionals*), per un lettore, possedere questa competenza significa essere in grado di identificare i propri bisogni informativi, conoscere quali sono le risorse disponibili e le modalità con le quali sono state prodotte, essere capaci di trovare le informazioni, essere in grado di valutare i risultati della raccolta delle informazioni, sapere come usare e come ottenere il massimo dalle informazioni che trova, usare le informazioni in modo etico e responsabile, oltre che comunicare e condividere le proprie scoperte¹⁸.

Occorre intervenire sul lettore, promuovendo in particolare alcuni aspetti della sua *information literacy*: 1) la capacità di valutare le informazioni raccolte; b) la consapevolezza della necessità di usare le informazioni in modo corretto.

È essenziale porre alla base di questa azione di educazione permanente l'idea che la lettura è un'attività creativa, nella quale il lettore interagisce in modo personale e sul piano intellettuale, emotivo e spirituale, con uno o più autori; come scrive Seneca a Lucilio (XI, 84):

Credo che la lettura mi sia necessaria, anzitutto perché mi impedisce di essere pago delle mie meditazioni, poi perché mi fa conoscere le indagini altrui, permettendomi di valutare i risultati e di integrarli con l'apporto del mio pensiero.

Le risorse bibliografiche sono strumenti per *costruire un proprio pensiero*, la propria visione del mondo, e la lettura è il processo che consente di realizzare questa azione creativa.

17 Alfredo Serrai, *Storia e critica della catalogazione bibliografica*. In: *Storia Della Bibliografia*. Roma: Bulzoni, 1997, p. 17.

18 CILIP, *Information literacy: definition*, «www.cilip.org.uk», 2011, <<http://www.cilip.org.uk/get-involved/advocacy/information-literacy/Pages/definition.aspx>>.

Interagire con l'autore significa prima di tutto capirne le intenzioni e valutarne le argomentazioni; nella comunicazione in rete occorre analizzare sia il supporto che il contenuto delle informazioni disponibili.

Analizzare il supporto significa comprendere come l'informazione è stata prodotta, utilizzando gli strumenti messi a disposizione dalla rete. Per esempio, chi ha prodotto o registrato il sito che pubblicizza l'acqua dalle eccezionali proprietà terapeutiche?

Su questo sito è possibile fare due rapide verifiche. La prima consiste nell'inserire in Google la stringa 'link:www.uniquewater.com.au/'; il risultato è il numero di link – presenti nel Web indicizzato da Google – che hanno come destinazione finale il sito con indirizzo <www.uniquewater.com.au>. Si può considerare come un indice di fiducia o di interesse che le persone che creano il Web assegnano al sito. Nel nostro caso il valore restituito da Google (.it o .com) è o (a fronte, per esempio, di 562 link 'in entrata' del sito <http://www.aib.it>).

La seconda verifica è analizzare chi ha registrato il sito, per esempio usando il servizio gratuito <http://whois.domaintools.com> e digitando l'indirizzo <http://whois.domaintools.com/uniquewater.com.au>. Con qualche passaggio, è possibile stabilire che il sito è registrato a nome del proprietario di una ditta di distribuzione di bevande (e non, per esempio, a nome di un istituto di ricerca medica).

Nel valutare il contenuto, si mira a stabilire la qualità e l'affidabilità delle informazioni in rete; i lettori possono avvalersi di metodi in parte già sviluppati per le risorse tradizionali, analizzando i diversi elementi a disposizione¹⁹:

- a) Autore: è indicato? Ci sono informazioni su di lui? È esperto dell'argomento? È contattabile?
- b) Editore o ente finanziatore: è identificato chiaramente? C'è un «chi siamo»? Indirizzo (URL) e layout del sito sono coerenti?
- c) URL: verificare il server, il dominio (.edu, .net, .gov, .org, .com) e il path (per esempio la tilde ~ indica una pagina personale);
- d) Punto di vista/orientamento: è esplicito? Si cerca di orientare l'opinione del lettore? È un sito commerciale? Il sito appartiene a una azienda con una *mission*? L'argomento è oggetto di un dibattito in corso?
- e) Aggiornamento: c'è una data? Si può stabilire?
- f) Accuratezza e affidabilità: c'è possibilità di verifica delle informazioni offerte? Ci sono errori di ortografia o grammatica? Vengono citate fonti? Nel caso si tratti di un lavoro scientifico vengono forniti dati e metodo? Ci sono link ad altre fonti? La pagina è 'citata' da altre pagine?
- g) Copyright e proprietà intellettuale: ci sono informazioni disponibili?

I contenuti scientifici e, in particolare, quelli di natura medica, richiedono naturalmente un livello di analisi più approfondito e specifico, e maggiori competenze da parte del lettore²⁰.

¹⁹ Ramsey Library, *Evaluating Web information*, «Www.lib.unca.edu», 2013, <http://www.lib.unca.edu/library/lr/evalweb.html>.

²⁰ Cfr. Avi Roy; Anders Sandberg, *The seven deadly sins of health and science reporting*, «The Conversation», 2013, <http://theconversation.com/the-seven-deadly-sins-of-health-and-science-reporting-21130>; William J. Sutherland; David Spiegelhalter; Mark Burgman, *Policy: twenty tips for interpreting scientific claims*, «Nature», 503 (2013), n. 7476, p. 335-337, <http://www.nature.com/news/policy-twenty-tips-for-interpreting-scientific-claims-1.14183>.

Usare correttamente le informazioni

È indispensabile educare all'uso corretto delle informazioni disponibili; anche se non esistono un 'proprietario' del Web o un'organizzazione centrale che detiene i diritti sul Web, non significa che le informazioni pubblicate in rete siano prive di diritti. La normativa sul diritto d'autore (e sul copyright) stabilisce come deve essere utilizzata l'opera intellettuale di altri. Nella maggior parte dei casi è necessario chiedere un permesso per riutilizzare il materiale.

Il lettore, quando si cimenta come autore, deve conoscere le tecniche di uso corretto delle informazioni, per evitare di incorrere nel plagio. È considerato plagio ciascuna di queste azioni²¹:

- presentare come propria l'opera di qualcun altro;
- copiare parole o idee di qualcun altro senza citare la fonte;
- omettere di inserire le virgolette quando si cita;
- fornire informazioni scorrette sulla fonte di una citazione;
- sostituire le parole mantenendo la struttura del periodo senza citare la fonte;
- copiare parole e idee da una fonte in quantità tale da farle diventare la parte più consistente della propria opera, senza citare la fonte.

Quando si produce un lavoro è indispensabile citare tutte le fonti utilizzate nel Web, come avviene per le fonti a stampa; esistono a questo scopo programmi – anche gratuiti – che facilitano la citazione delle fonti e la produzione delle bibliografie (Mendeley, Zotero, Endnote ecc.).

Il danno che può derivare alla ricerca scientifica da questa prassi illegale è molto grave: per questo motivo sono stati creati specifici software, già molto diffusi in mondo anglosassone ma anche in alcune università italiane²², messi a disposizione della comunità accademica e degli editori, per l'individuazione dei plagi.

Quando è il lettore che riceve le informazioni, può attuare diverse soluzioni tecniche che lo aiutano a proteggersi dal plagio. I metodi per individuare un plagio letterario, o più in generale per identificare l'autore originale di un concetto espresso in una data forma, sono essenzialmente tre:

1. ricerca di paragrafi 'tipici';
2. uso di programmi specifici;
3. analisi stilistica.

Se non si ha a disposizione un programma antiplagio, talvolta è sufficiente effettuare ricerche mirate su specifiche parti del testo tramite un motore di ricerca. Le parti da analizzare si possono individuare scegliendo brani che si discostano dallo stile del rimanente testo. Questa tecnica, molto semplice dal punto di vista tecnologico, è stata utilizzata da Bohannon²³ per smascherare la rivista «American Journal of Polymer Science», un periodico 'scientifico' di un editore pirata (SAP, Scientific & Academic Publishing) che ha copiato letteralmente la propria linea editoriale da quella del periodico «Journal of Polymer Science», pubblicato da Wiley a partire dal 1947.

²¹ Tratto da: <<http://www.plagiarism.org>>.

²² Per gli effetti, si veda: <<http://www.universita.it/software-anti-plagio-tesi-universita-venezia>>. Per un esempio dei risultati che si ottengono dall'analisi con il software antiplagio si veda: <<http://www.tesilink.it/docenti.jsp>>.

²³ John Bohannon, *Who's afraid of peer review?*, «Science», 342 (2013), n. 6154, p. 60–65, doi:10.1126/science.342.6154.60.

Un approccio più tecnologicamente avanzato è offerto dai programmi antiplagio; l'offerta è molto varia, in relazione al tipo di uso che se ne deve fare.

Si può partire da siti online sui quali effettuare la ricerca di parti di testo; per esempio *Duplichecker* (<<http://www.duplichecker.com>>) consente senza registrazione la verifica di un testo di lunghezza massima di 1500 parole. Per effettuare più di una ricerca nello stesso giorno, è necessario registrarsi (gratuitamente).

Esistono programmi che hanno una versione client da scaricare, come *Viper* (<<http://www.scanmyessay.com>>); anche questo programma è gratuito nella versione base; richiede la creazione di un account e offre una interfaccia semplice e intuitiva (Figura 2), la possibilità di selezionare il paese di provenienza e di scegliere l'ambito disciplinare all'interno del quale il programma deve verificare l'eventuale presenza di parti plagate. La versione gratuita ha un limite d'uso giornaliero che corrisponde circa a una quindicina di pagine.

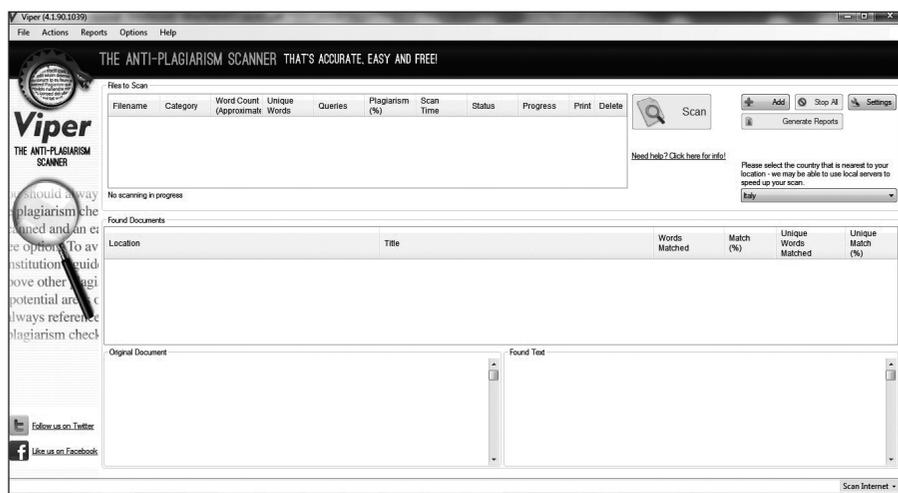


Figura 3 – Schermata di avvio del programma Viper

Esistono infine programmi a pagamento, come *Compilatio* (<<http://www.compilatio.net/it>>) e *Turnitin* (<<http://turnitin.com/it>>), che richiedono investimenti possibili solo a istituzioni come le Università.

L'analisi stilistica è il terzo approccio possibile alla scoperta di un falso: dall'analisi della Donazione di Costantino effettuata da Lorenzo Valla, agli studi di Jean Mabillon, il padre della Diplomatica, il principio è quello espresso da Antonio Augustin: «se non si stabilisce a priori ciò che è autentico e ciò che è falso, non si può compiere nessuno studio fondato»²⁴. Il riconoscimento di variazioni di stile all'interno dello stesso testo è il metodo più antico per scoprire un falso; per un docente è facile, per un lettore è necessario maggiore allenamento.

²⁴ Antonio Augustin, *Dialogos de medallas, inscripciones y otras antiguedades*. Tarragona: Felipe Mey, 1587, p. 443.

Esistono programmi per elaborare anche per l'analisi stilistica: l'uso di Signature (<<http://www.philocomp.net/humanities/signature.htm>>) e di Java Graphical Authorship Attribution Program (o JGAAP; <http://evllabs.com/jgaap/w/index.php/Main_Page>) ha consentito di attribuire a J.K. Rowling, creatrice di Harry Potter, un romanzo uscito con lo pseudonimo di Robert Galbraith²⁵.

Il ruolo del servizio bibliotecario

La comunicazione in rete ha profonde conseguenze sul cambiamento dei modelli di comunicazione: se ci si allontana dai tre sistemi consolidati (editoriale tradizionale, open access e sociale cooperativo) vengono meno gli strumenti di controllo – in fase di produzione – dell'autorevolezza dell'informazione disponibile. Una maggiore conoscenza delle proporzioni tra le informazioni rese disponibili con i modelli consolidati e quelle prodotte in modo disintermediato potrebbe aiutare a definire le proporzioni del problema del falso in rete. Sarebbe così possibile verificare se davvero il cinquanta per cento delle notizie che si trovano sul Web sono false²⁶.

Questa perdita di controllo sulla qualità e sull'effettiva disponibilità delle informazioni ha due conseguenze molto rilevanti, che riguardano in modo indiretto e diretto il futuro della nostra professione.

La risposta del servizio bibliografico alla diffusione di false informazioni in Internet non potrà mai essere la selezione preventiva delle informazioni; si tratterebbe di una forma nascosta di censura, inaccettabile dal punto di vista deontologico.

L'unica risposta percorribile consiste nell'educazione e nella diffusione tra i lettori di un atteggiamento critico e costruttivo verso ogni informazione disponibile in rete: «La migliore protezione che possiamo avere contro la disinformazione è adottare un atteggiamento critico verso tutte le informazioni sul Web»²⁷.

Anche l'AIB, nel documento programmatico *Rilanciare le biblioteche pubbliche italiane*, ricorda che

la biblioteca di ente locale [...] intes[a] come sistema diretto a promuovere lo sviluppo socio-economico degli individui e la qualità della vita, può affiancare alle funzioni tradizionali di promozione della lettura e supporto allo studio altre attività: [...] promuovere l'alfabetizzazione informativa e informatica per permettere ai propri utenti di accedere all'informazione di rete, sviluppare le competenze nella ricerca dell'informazione e stimolare la capacità di valutazione critica dei risultati delle ricerche²⁸.

Rispetto alle basi per l'*information literacy* individuate dal CILIP (cfr. paragrafo *Valutare le informazioni*), la maggior parte delle biblioteche sembra essersi concentrata

25 Federico Guerrini, *La paternità del romanzo sotto pseudonimo di J.K. Rowling 'suggerita' da un software*, «La Stampa», 20 luglio 2013, <<http://www.lastampa.it/2013/07/20/tecnologia/la-paternita-del-romanzo-sotto-pseudonimo-di-jk-rowling-suggerita-da-un-software-XfNkDNFYn67UDColEnYqDM/pagina.html>>.

26 Rai 2 alla caccia delle web-bufale con Roberto Giacobbo, 2013, <<http://www.digital-sat.it/ds-news.php?id=34191>>.

27 Paul S. Piper, *Better read that again: Web hoaxes and misinformation*, «Searcher», 8 (2000), n. 8, <<http://www.infotoday.com/searcher/sepoo/piper.htm>>.

28 AIB, *Rilanciare le biblioteche pubbliche italiane: documento programmatico*, «Www.aib.it», 2011, <<http://www.aib.it/attivita/campagne/2012/12818-rilanciare-le-biblioteche-pubbliche-italiane-documento-programmatico>>.

sulla risposta ad alcuni obiettivi, ovvero educare il lettore a stabilire quando e perché ha bisogno di informazioni e dove trovarle.

Le biblioteche sembrano avere fatto proprio solo in parte l'obiettivo di educare a valutare le informazioni; spesso a questo compito si adempie creando e mettendo a disposizione elenchi di aspetti da esaminare (autore, editore ecc.)²⁹ più che con attività progettate su misura (come avviene per esempio per l'educazione all'uso dei cataloghi e delle banche dati ecc.).

L'offerta di servizi che mirano a educare i lettori a dove e come usare le informazioni e a come comunicarle in modo corretto è ancora troppo ridotta, mentre nel contesto digitale, dove il confine tra autore e lettore sfuma e tende a svanire, queste due competenze sono sempre più essenziali alla formazione di cittadini consapevoli.

Alcuni esempi di esperienze e progetti di *information literacy*

Il tentativo perseguito nell'articolo di presentare in un quadro complessivo e unitario e, per quanto possibile coerente, molti temi che afferiscono all'ambito disciplinare e alla professione bibliotecaria seguendo come tema centrale la categoria del falso, offre il vantaggio di consentire una visione d'insieme e le connessioni tra molti temi specialistici. Come rovescio della medaglia, nel panorama che ne risulta è possibile soltanto richiamare i diversi temi che afferiscono a quello principale, senza approfondirne alcuno e limitando i rinvii bibliografici ai dati raccolti e selezionati, alle sole fonti, dando per scontato che il lettore (studioso o bibliotecario) abbia tutti i mezzi, materiali e intellettuali, per sviluppare con successo una ricerca bibliografica su ciascun tema specifico.

Anche nel caso del tema delle esperienze realizzate e dei progetti in corso in ambito di *information literacy*, la prospettiva non può che essere esemplificativa.

Una prima esperienza interessante di *information literacy* si svolge dal 2000 presso la Biblioteca umanistica³⁰ dell'Università di Firenze³¹. La biblioteca ha attivato alcuni corsi stabili di ricerca in biblioteca: ricerca in biblioteca per l'area umanistica, per biblioteconomia, per la storia dell'arte e DAMS, per la psicologia e per la formazione primaria. I corsi si svolgono due volte l'anno, hanno una durata variabile da 16 a 20 ore, sono frequentati in prevalenza da laureandi, possono prevedere una relazione finale o esercitazioni e test finale, includono un test di ingresso e un questionario di gradimento, si svolgono in aula informatica e talvolta prevedono il riconoscimento di CFU.

I corsi applicano metodi didattici basati sul lavoro di gruppo, sulla partecipazione attiva, sull'apprendimento attraverso la pratica e sulla condivisione tra pari. L'obiettivo in generale non è tanto promuovere la conoscenza e la capacità di usare spe-

²⁹ UC Berkeley, *Evaluating web pages. Techniques to apply & questions to ask*, «www.lib.berkeley.edu», 2012, <<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html>>; Johns Hopkins Sheridan Libraries, *Evaluating information found on the Internet*, «[Guides.library.jhu.edu](http://guides.library.jhu.edu)», 2013, <<http://guides.library.jhu.edu/content.php?pid=198142&sid=1657518>>.

³⁰ La Biblioteca umanistica serve gli utenti delle scuole di Psicologia e di Studi umanistici e della formazione.

³¹ Un particolare ringraziamento alla dott.ssa Stefania Puccini, che ha presentato il progetto dell'Università di Firenze durante la tavola rotonda conclusiva del corso "Il falso in internet: *information literacy* e autorevolezza del web", organizzato dall'AlB sezione Emilia Romagna e tenutosi il 20 gennaio 2014 presso la Mediateca di San Lazzaro di Savena (BO) e che ha fornito, con pronta disponibilità e sincero entusiasmo, tutte le informazioni sull'esperienza fiorentina.

cifiche risorse informative (con particolare attenzione a quelle del Sistema Bibliotecario di Ateneo) quanto mettere il corsista in condizione di saper cogliere la complessità dell'ecosistema informativo, di sapere scegliere tra i diversi sistemi informativi in base ai propri bisogni informativi, di sviluppare la propria capacità critica e l'autonomia di giudizio per essere un cittadino più consapevole, in un quadro caratterizzato dalla formazione continua e dalla cultura partecipativa della rete.

La seconda esperienza di *information literacy* proposta a titolo di esempio è stata sviluppata presso la Biblioteca delle Scienze dell'Università di Pavia, in collaborazione con le altre biblioteche del polo scientifico, Biblioteca di Area medica e Biblioteca della Scienza e della Tecnica³².

La biblioteca ha elaborato e realizzato un ampio progetto che ha visto lo svolgimento di corsi di formazione a diversi livelli:

- a) corsi di livello base diretti all'accoglienza di studenti delle scuole superiori (in collaborazione con l'Ufficio Orientamento), agli studenti Erasmus (in collaborazione con l'Area affari internazionali) e alle matricole, con corsi periodici di *Biblioaccoglienza* durante tutto l'anno accademico;
- b) corsi di livello intermedio, che propongono un'introduzione alla ricerca bibliografica con strumenti cartacei e digitali;
- c) corsi di livello avanzato, rivolti a studenti del 3° e 5° anno (metodologia della tesi, con particolare riguardo alle fasi di ricerca delle fonti e di stesura del testo), ai dottorandi (banche dati specializzate, diritto d'autore, indicatori bibliometrici), e corsi *on demand* per docenti e ricercatori³³.

La gradualità e la differenziazione dell'offerta in base all'utenza è certamente un elemento di forza di questa esperienza, da esportare in altri progetti. Altri due fattori sono risultati determinanti per la buona riuscita del progetto: l'esperienza maturata negli anni precedenti nel gruppo di lavoro sul Reference e il coinvolgimento e il supporto da parte dei docenti.

Il successo raccolto dalla Biblioteca delle Scienze ha portato alla costituzione di un gruppo di lavoro allargato a tutte le aree del sistema bibliotecario d'ateneo (Gruppo di coordinamento attività di Assistenza Bibliografica), che si occuperà di *quick reference*, di *reference specialistico* e di estendere le attività d'*information literacy* anche nell'area giuridico-economico-umanistica dell'Università di Pavia.

L'ultimo caso riguarda il progetto in corso "Educare a documentarsi: la competenza informativa per lo sviluppo di nuove idee", elaborato presso la Biblioteca umanistica e della formazione dell'Università di Udine³⁴. Il progetto si propone di rispondere alla necessità dell'utenza di conoscere e utilizzare al meglio il patrimonio bibliografico dopo il recente riassetto organizzativo della Biblioteca e delle sue raccolte. L'obiettivo degli interventi, organizzati su più livelli diretti rispettivamente a studenti, laureandi e dottorandi, docenti e ricercatori, studiosi e utenti esterni, è «sviluppare la capacità dell'utente di cercare informazioni per risolvere problemi, di rendersi autonomo nel valutare positività e criticità e di creare nuova conoscenza e

32 Un particolare ringraziamento alla dott.ssa Maria Carla Uberti, Responsabile del Servizio Biblioteche dell'Università di Pavia, per la disponibilità dimostrata e per il materiale informativo fornito sull'esperienza in corso e sui progetti futuri.

33 Lorenza Morisi; Anna Bendiscioli, *Il Progetto Della Biblioteca Delle Scienze* (Pavia, 2014), 1-3.

34 Un particolare ringraziamento alla dott.ssa Rosanna D'Andrea, Direttore tecnico, e alla dott.ssa Antonella Passone, Responsabile del Settore Servizi al pubblico e reference, per la disponibilità e per le informazioni fornite sul progetto della Biblioteca Umanistica e della formazione.

nuove idee». Il progetto si avvarrà – nella fase iniziale – del supporto di volontari del Servizio Civile Volontario, appositamente selezionati e successivamente preparati, per favorire l’attuazione dei principi della *peer education*.

Ritornare ai valori fondamentali della professione

In questo contributo, l’approccio è partito dal Web, ma il problema della qualità delle informazioni riguarda, come si è visto, qualsiasi supporto e qualsiasi canale di comunicazione; pensare che il falso dilaghi sul Web o, peggio ancora, solo sul Web significa perpetrare un luogo comune: «Science» e «Nature», prestigiose roccaforti del sapere scientifico, non sono nuove alla pubblicazione di studi completamente falsi³⁵ e le ritrattazioni delle riviste scientifiche o degli autori sono diventate molto più frequenti negli ultimi dieci anni³⁶, al punto che esistono siti specializzati per segnalarle³⁷. Anche l’*Encyclopaedia Britannica* talvolta è superata dalla rivale Wikipedia³⁸.

C’è una seconda emergenza che ha importanti ricadute sulla nostra professione: forse non si è riflettuto abbastanza che sulla rete, al di fuori dei servizi bibliotecari offerti con strumenti tradizionali e digitali (ovvero all’esterno della biblioteca ibrida contemporanea), non è garantito alcun controllo sulla effettiva accessibilità delle risorse e sull’accesso libero, equo e democratico alle informazioni.

Il rischio reale che la profilazione degli utenti e altri strumenti di manipolazione delle informazioni sul Web possano impedire a qualsiasi lettore un accesso alle informazioni su base libera, equa e democratica riguarda direttamente i valori della nostra professione di mediatori e la impegna per il futuro. È un periodo di profondo cambiamento e di profonda crisi per la professione; in passaggi cruciali come questi è necessario distinguere tra ciò che abbaglia soltanto e ciò che può costituire un solido presupposto per rinnovare il servizio bibliotecario e la sua funzione sociale, per segnare la direzione da prendere per superare questa crisi, mantenendo salde le radici nei valori della nostra professione.

35 Cfr. per esempio <<http://www.ilfattoquotidiano.it/2011/11/14/sono-inventato-tutto-fallito-confessione-luminare-imbrogliore/170688>>.

36 Ivan Oransky, *The Good, the Bad and the Ugly: what retractions tell us about science transparency*, 2011, <<http://www.slideshare.net/CrossRef/the-good-the-bad-and-the-ugly-what-retractions-tell-us-about-scientific-transparency-2011-annual-meeting>>; Elena Dusi, *Così gli scienziati truccano le ricerche*, «*Www.repubblica.it*», 3 ottobre 2012, <http://www.repubblica.it/scienze/2012/10/03/news/cos_gli_scientiati_truccano_le_ricerche-43746932>.

37 Cfr. <<http://retractionwatch.wordpress.com>>.

38 Così sostiene l’articolo Jim Giles, *Internet encyclopaedias go head to head*,” «*Nature*», (2005), n. 438, p. 900–901, doi: 10.1038/438900a.; il comitato editoriale dell’enciclopedia ha aspramente criticato il metodo adottato nella ricerca e ha fornito la propria versione in: *Encyclopaedia Britannica Inc., Fatally flawed. Refuting the recent study on encyclopedic accuracy by the journal Nature*, 2006, *Encyclopaedia Britannica*, <http://corporate.britannica.com/britannica_nature_response.pdf>; le controdeduzioni della rivista «*Nature*» si trovano all’indirizzo: <<http://www.nature.com/nature/britannica/index.html>>.

Articolo proposto il 19 marzo 2014 e accettato il 28 aprile 2014.

ABSTRACT AIB studi, vol. 54 n. 2/3 (maggio/dicembre 2014), p. 227-240. DOI 10.2426/aibstudi-10130.

CARLO BIANCHINI, Università degli Studi di Pavia, Dipartimento di Musicologia e Beni culturali, corso Giuseppe Garibaldi 178, 26100 Cremona, e-mail carlo.bianchini@unipv.it.

Il falso in Internet: autorevolezza del Web, information literacy e futuro della professione (SECONDA PARTE)

Il ricorso sempre più diffuso alla rete come strumento di ricerca da parte di studenti e ricercatori pone problemi di valutazione della qualità e dell'autorevolezza delle informazioni disponibili sul Web. L'articolo presenta un approccio sistematico ai meccanismi di produzione, trasmissione e ricezione di informazioni false nel Web, analizza il problema in relazione al futuro del servizio bibliotecario e dimostra che è indispensabile comprendere bene questi meccanismi ed esserne consapevoli – sia come lettori che come bibliotecari – per utilizzare le informazioni in modo critico, corretto e creativo.

(Questo articolo è la seconda parte di un contributo più esteso. La prima parte è stata pubblicata sul numero 1/2014 della rivista)

The Internet and the spread of false information: reliability of the websites, information literacy and the future of the profession (PART 2)

Today the majority of students and scholars (83%) begin their searches starting from a search engine. Thus, the evaluation of the quality and reliability of the information on the Internet becomes an essential task for every librarian. The paper examines how false information is created, disseminated and used on the Internet, considering the issue in relation with the future of library service, and remarking how readers – as well as librarians – must improve their comprehension of how the Internet works for a critical, correct and creative use of information.

(NB: this is the second part of a more extensive study. The first part has been published in AIB Studi 1/2014)