Gli strumenti del nuovo web e l'organizzazione della ricerca in campo umanistico

Gino Roncaglia

Preprint – Il testo è in corso di pubblicazione negli atti del convegno "Le opere filosofiche e scientifiche. Filosofia e scienza tra testo, libro e biblioteca", Lecce, 7-8 febbraio 2008, a cura di Franco A. Meschini.

Premessa: il 'ciclo della ricerca' e la rivoluzione digitale.

Un pregiudizio purtroppo ancora abbastanza diffuso – tanto da continuare ad influenzare alcuni dei meccanismi di finanziamento della ricerca attivi nel nostro paese – individua una delle differenze fra l'attività di ricerca nel campo delle cosiddette 'scienze esatte' e quella svolta in campo umanistico nella (rispettivamente) maggiore o minore dipendenza dall'uso di *strumenti*. Il lavoro in ambito umanistico è così presentato come una forma di attività intellettuale sostanzialmente autosufficiente, che richiede a ben vedere solo qualche libro da leggere, qualche foglio per scrivere e un po' di tempo per pensare (e può dunque ben accontentarsi di investimenti relativamente ridotti), mentre la 'vera' ricerca scientifica ha bisogno di strumentazioni complesse e costose, laboratori, tecnici, e naturalmente di finanziamenti molto maggiori.

La realtà è, come sempre, assai più complessa. Ogni forma di ricerca scientifica, sia essa in campo umanistico o nel campo delle scienze 'hard', ha bisogno dei propri strumenti, e tracciare un confine fra strumenti 'tecnici' e 'intellettuali' è molto più arduo di quanto non potrebbe sembrare a prima vista. La rivoluzione informatica e lo sviluppo delle reti telematiche hanno anzi messo in crisi l'idea stessa dell'esistenza di un confine di questo tipo. Non a caso, la metafora preferita nella fase di sviluppo dei primi strumenti informatici è stata quella – ben nota agli storici della filosofia già a proposito di strumenti prettamente intellettuali come la mnemotecnica, le 'macchine combinatorie' lulliane o i linguaggi artificiali – dell'estensione delle nostre facoltà conoscitive, e in particolare della memoria[1].

Strumenti né puramente tecnici (per via dei fondamenti logici e algoritmici del loro funzionamento, ben riconoscibili nel modello ideale rappresentato dalla macchina di Turing) né puramente intellettuali, i prodotti della rivoluzione digitale – computer, software, reti telematiche – impongono così una radicale revisione di molti paradigmi relativi alle modalità della ricerca in campo umanistico, e dunque anche in campo filosofico. In particolare, la ricerca umanistica mostra con sempre maggiore chiarezza la propria natura di processo di raccolta, elaborazione, distribuzione dell'informazione, con caratteristiche largamente analoghe a quelle che caratterizzano le scienze esatte.

---

[1] Già Vannevar Bush vedeva il Memex – macchina associativa non digitale ma chiara prefigurazione degli odierni strumenti di organizzazione ipertestuale dell'informazione – come "an enlarged, intimate supplement to (…) memory" (Vannevar Bush, "As We May Think", *The Atlantic Monthly* 176, no. 1 (1945), pp. 101-108, http://www.theatlantic.com/doc/194507/bush). L'articolo di Bush ha influenzato molti fra i pionieri dello sviluppo di sistemi informatici (fra gli altri, Douglas Engelbart, che chiama *Augment* la prima rete informatica da lui realizzata considerandola come uno strumento di "human augmentation"), ed è citato da Ted Nelson fra le fonti del suo lavoro sull'idea di ipertesto: cf. Theodor H. Nelson, "As We Will Think", in *From Memex to hypertext : Vannevar Bush and the mind's machine*, a cura di James M. Nyce e James M. Kahn, Academic Press, Boston 1991, pp. 245-260. Su Bush si veda anche G. Pascal Zachary, *Endless frontier: Vannevar Bush: engineer of the American century*, Free Press, New York - London 1997.

A ben vedere, infatti, il lavoro di ricerca scientifica, nel campo delle discipline umanistiche così come in quello delle cosiddette 'scienze esatte', prevede di norma fasi abbastanza costanti, spesso caratterizzate, sia nel loro complesso sia nella struttura interna di ciascuna di esse, da un andamento in parte ciclico. Se cerchiamo di individuarle e distinguerle, possiamo provare a proporre lo schema seguente[2]: 1) formazione di base del ricercatore; 2) scelta e definizione del tema o del problema da affrontare, in genere sulla base di informazioni parziali e lacunose (e con una componente non trascurabile di *serendipity*); 3) ricerca, raccolta e selezione di informazioni rilevanti, 4) analisi, classificazione ed elaborazione creativa delle informazioni raccolte, formulazione e verifica (anche sperimentale) di ipotesi di lavoro, discussione informale, integrazione delle informazioni rilevanti, progressiva definizione delle conclusioni; 5) organizzazione e stesura delle conclusioni raggiunte, in genere – ma non necessariamente – sotto forma di testi: draft, preprint ecc.; 6) validazione di primo livello (procedimenti di *peer review*, selezione editoriale ecc.), pubblicazione e distribuzione dei prodotti della ricerca: libri, articoli, ecc.; 7) raccolta, conservazione, classificazione e fruizione dei prodotti della ricerca, attraverso strumenti quali biblioteche (istituzionali o personali), archivi, repository di vario genere; 8) validazione di secondo livello: analisi e discussione dei prodotti della ricerca all'interno della comunità di riferimento, e – nei casi migliori – ulteriore diffusione dei risultati raggiunti attraverso il circuito della formazione formale e informale (scuola, università, divulgazione, ecc.).

Naturalmente il lavoro concreto di ricerca non ha mai una struttura così lineare: le sue varie fasi tendono inevitabilmente almeno in parte a sovrapporsi e talvolta a confondersi, e assumono fisionomie diverse a seconda della particolare natura della ricerca svolta. Inoltre, un modello un po' più raffinato del lavoro di ricerca dovrebbe tenere conto anche di fattori essenziali come il suo finanziamento (e ritorno economico), le aspettative e gli interessi della comunità di riferimento (che possono influenzare sia la scelta del tema affrontato sia l'effettiva diffusione e riutilizzazione dei risultati ottenuti), le forme di tutela della proprietà intellettuale e in generale la cornice normativa e istituzionale in cui tale lavoro si svolge. Ma le fasi di formazione di base, definizione del tema della ricerca, documentazione, elaborazione creativa, scrittura (o realizzazione, nel caso di prodotti non testuali), pubblicazione, fruizione e discussione intersoggettiva, si ritrovano comunque in attività di produzione culturale assai lontane fra loro: dalla scrittura di un romanzo alla produzione di un film, dalla stesura di un articolo o di un saggio alla definizione di un brevetto.

Nel corso del tempo, ciascuna di queste fasi ha visto sviluppare – naturalmente anche in dipendenza dai diversi ambiti e settori di ricerca – i propri metodi e i propri strumenti: metodi e strumenti che costituiscono un po' la nervatura del lavoro di ricerca e di produzione culturale, e contribuiscono a determinarne sia la cornice di possibilità, sia le forme, sia i risultati.

È credo interessante osservare come, una volta 'distribuita' sull'intero processo del lavoro di ricerca e di produzione culturale, la *vexata quaestio* della differenza fra le 'due culture', quella umanistica e quella scientifica, perda almeno parte del suo rilievo, a favore da un lato di una articolazione disciplinare assai più

[2] Esistono numerosi modelli alternativi del 'ciclo della ricerca' o 'ciclo della comunicazione scientifica', elaborati in contesti e con scopi diversi ('*research cycle*', '*publication cycle*', '*cultural production chain*' ecc.; si vedano ad es. K. Subramanyam, *Scientific literature*, in A. Kent, H. Lancour, J. E. Daily (editors), *Encyclopedia of Library and Information Science*, New York, Dekker, v. 26 (1979), pp. 376-548; K. Subramanyam, *Scientific and Technical Information Resources*, New York, Dekker, 1981; B. C. Vickery, *Scientific communication in history*, Lanham - London, Scarecrow Press, 2000; Consortium of University Research Libraries (CURL), *Scholarly Communications Crisis* (briefing paper), 2001; l'articolazione di quello qui presentato – che ha come base il semplice schema *read-research-write-publish* – è una mia proposta; ringrazio Antonella De Robbio e Riccardo Ridi per le preziose osservazioni al riguardo.

complessa e assai meno rigida, dall'altro di problemi e interessi largamente comuni (ad esempio per quel che riguarda la conservazione, la descrizione, l'organizzazione e la diffusione dei risultati della ricerca, o l'efficacia della formazione).

C'è anche un altro aspetto che, dal nostro punto di vista, merita di essere sottolineato: a ben vedere, tutte le fasi del lavoro di ricerca che abbiamo cercato di individuare e differenziare implicano l'uso di strumenti linguistico-testuali. Anche le discipline scientifiche più direttamente legate alla dimensione tecnico-sperimentale hanno le loro riviste (siano esse a stampa o su web), e affidano principalmente ai testi l'articolazione, la descrizione, la diffusione dei propri risultati. L'uso di strumenti multimediali (ad esempio nella formazione o nella divulgazione) allarga indubbiamente le possibilità espressive, e si rivela in alcuni casi particolarmente funzionale, ma – anche nell'era dei nuovi media elettronici – sembra innestarsi in una trama che trova comunque nei testi e nella testualità scritta la propria componente fondamentale.

Nel discutere – in forma comunque necessariamente sommaria – alcuni temi e alcuni strumenti legati alla gestione dell'informazione e al ciclo della ricerca nell'epoca della rivoluzione digitale, non intendo dunque presupporre distinzioni rigide fra scienze fisico-naturali e scienze (o discipline) umanistiche. Intendo piuttosto indirizzare verso una particolare – e comunque non esclusiva – attenzione verso le dimensioni della testualità scritta, dell'uso del linguaggio naturale e dell'attenzione verso la storia della cultura, in particolare nella sua forma di 'cultura del libro'. Credo che, anche al di là dei pur rilevanti riflessi dell'uso degli strumenti informatici in ciascuno dei singoli ambiti che tradizionalmente articolano il campo delle scienze umane (a cominciare dal settore della riflessione filosofica), proprio questa attenzione possa costituire una importante caratteristica dell'informatica umanistica, e contribuire ad affidarle una funzione in qualche misura trasversale e di raccordo rispetto alle tradizionali suddivisioni disciplinari[3].

In questo contesto, mi soffermerò in questa sede, brevemente, su due questioni: la descrizione e classificazione dell'informazione, in particolare nelle prospettive, centrali nella discussione attuale, del Web 2.0 e del web semantico, e l'archiviazione e distribuzione aperta dei contenuti della ricerca, attraverso i cosiddetti 'archivi aperti' (*open archive*).

Descrivere e classificare informazione: 'web 2.0' e web semantico

Come molti termini alla moda, l'espressione 'Web 2.0' ha, più che un significato preciso e ben definito, una connotazione abbastanza generale, e in parte generica. Da un paio d'anni a questa parte, tutti gli strumenti e i siti web che vogliono proporsi come di successo e innovativi si presentano come incarnazione del web 2.0, e a seconda delle situazioni e delle necessità piegano quest'espressione in un verso o nell'altro.

Ma cos'è il web 2.0? E – questione forse più rilevante – quali sono in effetti le tendenze più innovative e promettenti del nuovo web dal punto di vista della descrizione e classificazione dell'informazione?

Il termine 'Web 2.0' è stato utilizzato in molte occasioni diverse, ma la sua introduzione si deve probabilmente a una intuizione della dinamica casa editrice specializzata O'Reilly, e in particolare del suo vice-presidente Dale Dougherty. È lo stesso Tim O'Reilly a raccontarne la nascita:

[3] Il tema è discusso in forma più approfondita in G. Roncaglia, "Informatica umanistica: le ragioni di una disciplina", *Intersezioni. Rivista di storia delle idee,* n. 3, dicembre 2002, pp. 353-376.

> "Lo scoppio della bolla dot-com nell'autunno del 2001 ha segnato un punto di svolta per la rete. Molte persone sono giunte alla conclusione che la rete fosse assolutamente sopravvalutata, quando, invece, le bolle e le conseguenti crisi sembrano essere una caratteristica comune di tutte le rivoluzioni tecnologiche. Le crisi normalmente segnano il punto in cui una tecnologia in crescita è pronta a prendere il posto che le spetta, al centro del palcoscenico. I simulatori vengono eliminati, le storie di effettivo successo mostrano la loro forza e qui si inizia a comprendere cosa separa le une dalle altre.
>
> Il concetto di "Web 2.0" ebbe inizio con una sessione di brainstorming durante una conferenza tra O'Reilly e MediaLive International. Dale Dougherty, pioniere del web e Vice-Presidente di O'Reilly, fece notare che, tutt'altro che "crollata", la rete era più importante che mai, con nuove interessanti applicazioni e siti nascenti con una sorprendente regolarità. Inoltre, le società che erano sopravvissute al collasso, sembravano avere alcune caratteristiche in comune. Poteva essere che il collasso delle dot-com avesse segnato per la rete un punto di svolta tale che un richiamo all'azione definito come "Web 2.0" potesse avere senso? Concordammo con questa analisi e così nacque la Conferenza Web 2.0."[4]

La prima Web 2.0 Conference (l'evento si ripete con cadenza annuale) ha luogo nell'Ottobre 2004, e ha alla base l'idea che il web si sia ormai trasformato – da superficie sostanzialmente 'piatta' su cui appoggiare informazioni – in una sorta di piattaforma applicativa condivisa, all'interno della quale le informazioni possono essere non solo distribuite ma anche create ed elaborate collettivamente, eventualmente con l'aiuto di apposite 'web applications' utilizzabili direttamente dall'interno del proprio programma di navigazione.

A fare emergere questa idea erano stati probabilmente e in primo luogo due sviluppi del 'panorama di rete': da un lato la diffusione di strumenti come i weblog, che nel luglio 2004 erano già oltre tre milioni e prefiguravano non solo una molteplicità di punti di distribuzione 'individuali' di informazione ma una vera e propria 'blogosfera', rete di contenuti ai quali strumenti come i Fed RSS e il trackback offrivano la possibilità di una circolazione personalizzata e dinamica e di una rielaborazione parzialmente automatica[5]; dall'altro la nascita dei primi siti in cui la condivisione di risorse era esplicitamente orientata alla creazione di relazioni sociali fra gli utenti (il cosiddetto 'social networking', inaugurato da *MySpace* nel 2003). Pochi mesi prima, nel febbraio 2004, era nato anche *Flickr*, strumento per la condivisione di immagini che permetteva agli utenti di 'etichettare' liberamente le fotografie archiviate in rete, secondo il principio del cosiddetto 'social tagging': un'altra delle caratteristiche che, come vedremo, diventeranno tipiche del Web 2.0. Infine, dal punto di vista strettamente tecnico, la creazione di semplici applicazioni utilizzabili direttamente dall'interno del proprio browser, già avviata da strumenti come Javascript e ActiveX, acquistava ulteriori potenzialità interattive con lo sviluppo di Ajax, una tecnica di sviluppo web che a sua volta – come vedremo – associa strettamente il suo nome al web 2.0.

---

[4] Tim O'Reilly, *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*, in rete all'indirizzo http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html, traduzione Italiana all'indirizzo http://www.xyz.reply.it/web20/index.php/1.

[5] Si veda al riguardo G. Roncaglia, "Blogosfera e feed RSS: una palestra per il Semantic Web?", *Networks*, 2, 2003, pp. 47-56, in rete all'indirizzo http://dspace.unitus.it/handle/2067/9.

Il web come piattaforma, dunque. L'articolo del 2005 di Tim O'Reilly, probabilmente il testo che ha avuto maggiore influenza nella discussione sul Web 2.0, elabora e sviluppa questa idea con riferimento ad alcuni strumenti specifici. Si tratta di una intuizione che si è mostrata fruttuosa, e che corrisponde anche all'evoluzione tecnologica e non solo a quella sociale e culturale del web. A due anni di distanza da quell'articolo, una concezione organica della natura del Web 2.0 continua però a mancare. È tuttavia possibile elencare, almeno indicativamente, quelle che sono state finora le aree e le idee più frequentemente collegate al 'nuovo web', e delineare le prospettive e i punti di vista che portano a privilegiare questa o quella concezione del Web 2.0.

Una **prima prospettiva** è legata all'intervento più attivo degli utenti come produttori di contenuti: a determinarla è la crescita impetuosa del cosiddetto '**User Generated Content**' (UGC), il contenuto generato dagli utenti. Una rivoluzione nata tra il 2001 e il 2003 con il fenomeno già ricordato della crescita dei blog, e proseguita negli anni successivi con lo sviluppo di strumenti che facilitano l'immissione in rete e la condivisione anche di contenuti non testuali: il podcast per l'audio, e piattaforme come la già ricordata *Flickr* per le immagini e *YouTube* per il video.

In realtà, in un certo senso il web è nato fin dall'inizio all'insegna dello User Generated Content: uno degli aspetti fondamentali del web era infatti proprio la costruzione di uno spazio informativo integrato, al quale chiunque potesse contribuire in maniera relativamente facile. HTML, il linguaggio usato per costruire pagine web, si proponeva come lo strumento per eccellenza di questa rivoluzione: un linguaggio di marcatura facile, immediato, che chiunque poteva imparare a usare con poca fatica per inserire in rete le proprie pagine e collegarle a quelle degli altri.

E tuttavia, per quanto semplice, un linguaggio di marcatura è comunque qualcosa da studiare, ed era spesso percepito (a torto) dagli utenti come un parente molto vicino dei linguaggi di programmazione: qualcosa che, se possibile, era senz'altro meglio lasciare agli esperti. Inoltre, per inserire informazione in rete non bastava padroneggiare HTML: occorreva disporre di uno spazio web sul quale caricare le pagine, e di strumenti (in genere, client FTP) per trasferirvi i propri file.

Insomma, niente di impossibile per un utente dotato di buona volontà e di una base minima di competenze tecniche, ma neanche una passeggiata.

Il progressivo sviluppo di strumenti che semplificano l'inserimento in rete di informazione strutturata, eliminando prima (attraverso editor capaci di produrre codice HTML utilizzando un'interfaccia simile a quella di un normale programma di videoscrittura) la necessità di conoscere il funzionamento di un linguaggio di marcatura; poi (attraverso l'uso di procedure guidate di upload via web delle pagine) la necessità di sapere come funzionano i programmi per il trasferimento di file; infine (attraverso procedure di creazione automatica e guidata del 'template grafico' di un sito) la necessità di sapere come gestire graficamente il layout di un sito, ha preparato fra il 1997 e il 2001 la strada alla realizzazione di programmi capaci di gestire integralmente e in maniera assai facile un sito web: i cosiddetti Content Management System, o CMS, di cui i programmi per la gestione di un weblog – che cominciano a diffondersi alla fine del 2001 – non sono altro che una versione particolarmente semplice.

Progressivamente, strumenti che semplificano l'immissione di informazione in rete sono stati sviluppati anche per l'informazione sonora, con l'introduzione dei podcast e di piattaforme per la condivisione di musica in rete. La diffusione dei lettori portatili MP3 ha contribuito anch'essa a aumentare la domanda di contenuti audio, e anche se la maggior parte dei contenuti audio in circolazione è inizialmente rappresentata da musica 'pirata' in formato MP3, i contenuti (musicali, ma anche parlati…) generati dagli

utenti cominciano anch'essi a diffondersi rapidamente. E il processo riguarda anche le immagini, con i siti (si è già ricordato *Flickr*) che ne consentono la condivisione via rete, e i video (anche qui si è già ricordato *YouTube*, nato nel febbraio 2005).

Con il moltiplicarsi dei produttori di contenuti in grado di inserire facilmente in rete informazione strutturata, cresce naturalmente anche l'esigenza di descrivere e organizzare questa informazione, per garantirne la reperibilità. E siamo qui a una **seconda prospettiva** di particolare rilievo nel parlare di web 2.0, quella legata alla **classificazione dell'informazione**.

In questa prospettiva, il Web 2.0 nasce, se non in opposizione, almeno in una situazione di qualche tensione rispetto al progetto del cosiddetto 'semantic web' (web semantico) proposto da Tim Berners Lee[6]. Anche il semantic web nasce dalla necessità di trovare strumenti per migliorare la classificazione e la reperibilità dell'informazione; ma il progetto di semantic web è strettamente legato all'idea che il lavoro di organizzazione e gestione dell'informazione debba essere in gran parte automatico e basato su descrizioni fortemente standardizzate e formalizzate. I sistemi di classificazione dell'informazione alla base del semantic web sono – o dovrebbero essere – *ontologie formali*: griglie di classificazione elaborate da esperti del settore, espresse in maniera uniforme e rigorosa e associate all'informazione primaria attraverso l'uso di linguaggi e formalismi a loro volta rigidamente strutturali e ben definiti.

Al contrario, il web 2.0 mette gli utenti al centro anche del processo di classificazione. Al posto di sistemi classificatori predefiniti, gli utenti che immettono informazione in rete e quelli che la usano sono invitati ad aggiungere all'informazione primaria delle etichette descrittive ('tag') totalmente libere, sulla base della loro percezione della natura dell'informazione stessa e dei contesti di sua potenziale utilità: il cosiddetto *social tagging*.

Quale strada è davvero percorribile? E, se lo è, è anche preferibile rispetto ai meccanismi di classificazione più affidabili e 'professionali' che rappresentavano un po' il presupposto del semantic web? Per un verso, dovremmo rispondere sicuramente che è preferibile la prima strada, più rigorosa e controllata… ma è davvero sempre così? La seconda strada è più rapida, più facile… e per certi scopi sembra comunque funzionare abbastanza bene.

Sistemi di social tagging (si usa spesso a questo riguardo anche il termine 'folksonomy'), etichettatura libera e diffusa dei contenuti informativi, sono alla base di moltissimi siti 'Web 2.0'. Così, quasi tutti i sistemi weblog (ad esempio *WordPress*, probabilmente il più diffuso) mettono a disposizione la possibilità di etichettare e classificare liberamente i singoli post; lo stesso fanno, per i contenuti audio, i sistemi podcast più diffusi, per le immagini sistemi come *Flickr*, per i video sistemi come *YouTube*. Ed è possibile classificare in questo modo anche i siti web, creando le proprie liste di siti preferiti e associando a ciascuno di essi una o più etichette scelte in maniera completamente libera, utilizzando strumenti di 'social bookmarking' come *Del.icio.us*. Non a caso, proprio questi siti sono fra i più frequentemente associati all'idea di Web 2.0.

Certo, se le 'folksonomies' restano del tutto prive di qualunque meccanismo di validazione, che assicuri un minimo di uniformità e affidabilità alle classificazioni proposte, il rischio della confusione e dell'arbitrio è fortissimo. Chi assicura che le etichette che ho usato io per classificare una certa informazione

[6] T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", *Scientific American*, May 2001, pp. 34–43; per gli sviluppi più recenti delle posizioni di Berners Lee al riguardo cf. N. Shadbolt, W. Hall, T. Berners-Lee, "The Semantic Web Revisited", *IEEE Intelligent Systems* 21(3) pp. 96-101, May/June 2006, in rete all'indirizzo http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf.

corrispondano a quelle che avrebbe utilizzato qualcun altro, e soprattutto che corrispondano a quelle che userebbe un altro utente al momento di ricercare i miei contenuti, o contenuti simili a quelli immessi da me? Il grosso vantaggio delle ontologie controllate è proprio la standardizzazione e l'affidabilità dei sistemi di classificazione, di norma elaborati da esperti.

Anche il social tagging, però, ha una freccia importante al suo arco: e si tratta proprio della sua natura *sociale*. Le etichette con le quali classifico i contenuti informativi sono sì libere, ma non vengono scelte nel vuoto: sono attribuite utilizzando piattaforme usate da centinaia di migliaia di altri utenti, dunque in un contesto sociale assai ricco. E i sistemi di gestione delle informazioni possono *confrontare i miei comportamenti classificatori con quelli degli altri, segnalare automaticamente le etichette più frequenti per certe tipologie di contenuti (così come quelle più rare e anomale), contribuire insomma al filtraggio collaborativo dei metadati*.

Gli strumenti per farlo sono di vario genere: ad esempio le cosiddette 'nuvole di etichette', o 'tag cloud', rappresentazioni delle etichette più usate per descrivere un certo insieme di risorse, in cui le etichette scritte in corpo maggiore sono quelle più usate e frequenti.

Esistono anche altri sistemi per migliorare il social tagging, e in generale per cercare di conciliare ontologie controllate e social tagging. Ma il filtraggio collaborativo è sicuramente il più importante. E in effetti la **terza prospettiva** che vorremmo proporre nel cercare di individuare le caratteristiche specifiche del Web 2.0 è rappresentata proprio dal **filtraggio collaborativo**. In questo caso, utilizzato per selezionare e far emergere comportamenti classificatori, e dunque applicato ai metadati. Ma il filtraggio collaborativo è usato dal nuovo web in maniera generalizzata, anche con riferimento alle informazioni primarie. Il suo uso più frequente è legato al confronto fra i gusti degli utenti. Così, ad esempio, sistemi come *iLike*, *Last.fm* e *Pandora* (per motivi legati alla gestione dei diritti Pandora ha recentemente limitato l'accesso ai soli utenti nordamericani) permettono ai propri utenti di indicare la musica e i musicisti che preferiscono, e segnalano automaticamente brani ed autori che 'dovrebbero piacerci', dato che piacciono ad utenti con gusti assai simili ai nostri. *Amazon* e molti altri siti fanno lo stesso per quanto riguarda i libri.

Come si è detto, il filtraggio collaborativo dell'informazione è basato sulla dimensione sociale della condivisione di risorse via web. Proprio l'accentuazione del rilievo di tale dimensione costituisce la **quarta prospettiva** sotto cui considerare il web 2.0: quella legata ai siti e agli strumenti di **social networking**. Caratteristica tipica degli strumenti di social networking è il tentativo di creare relazioni fra gli utenti, permettendo a ciascuno di evidenziare i propri interessi, i propri gusti, e di 'collegare' il proprio profilo a quello dei propri amici e conoscenti (ma anche di eventuali sconosciuti che risultassero avere profilo e interessi tanto simili al nostro da renderli per noi 'interessanti'). In realtà, le 'reti sociali' che questi siti vogliono aiutare a creare hanno una duplice dimensione: reti di individui, e reti di contenuti informativi legati agli individui che ne sono produttori o consumatori. Molto spesso esiste poi una terza dimensione, rappresentata non dai singoli utenti ma da gruppi o comunità di utenti legati da interessi comuni (questa dimensione è particolarmente rilevante negli strumenti di social networking orientati alla creazione di comunità web, come *Ning*).

Tutti i siti di social networking hanno fra i propri strumenti principali la 'pagina di profilo' del singolo utente; spesso, oltre a una fotografia e ad alcuni dati su di sé e sui propri interessi, e all'elenco dei propri 'amici' all'interno del sistema, queste pagine possono essere fortemente personalizzate attraverso l'aggiunta di immagini, musica, video, e attraverso 'widget', programmini utilizzabili via web che permettono, ad esempio, di far sapere agli altri utenti quale brano musicale si sta ascoltando in quel momento, o che libro

si sta leggendo. Le funzionalità dei widgets di questo genere sono moltissime, e lo sviluppo di widget utili e interessanti (spesso demandato agli stessi utenti attraverso la distribuzione di appositi strumenti di programmazione semplificata) rappresenta uno dei campi di maggior interesse del mondo del social networking.

Il sito pioniere è in questo settore il già ricordato *MySpace*, ma negli ultimi anni la concorrenza si è fatta agguerritissima e ha visto la partecipazione di tutti i colossi della rete (e di molte start-up di belle speranze). Al momento, il sito di social networking probabilmente più interessante è *Facebook*, che ha proprio nella ricchezza di widgets e applicazioni web utilizzabili uno dei principali punti di forza.

Per valutare il rilievo che il social networking sta assumendo nel nuovo web, basti considerare che i due siti più popolari in questa categoria, MySpace e Facebook, hanno complessivamente quasi 200 milioni di utenti. Va osservato, inoltre, che strumenti orientati al social networking si stanno diffondendo anche al di fuori dei siti 'specializzati'. Così, ad esempio, siti come i già ricordati Flickr, YouTube, SlideShare ecc. permettono spesso di creare pagine personali di autopresentazione degli utenti, di costruire reti di amici e gruppi di interesse, e così via.

A questo punto, solo un cenno alla **quinta prospettiva** rilevante per il web 2.0: si tratta, come già accennato, proprio della possibilità di utilizzare il web come piattaforma e di creare siti web in grado di interagire con gli utenti attraverso procedure semplici ma potenti. Alla base vi è una tecnologia, **Ajax**, sulla quale sarebbe troppo complesso soffermarsi in questa sede. Ma è importante sapere che, per i programmatori esperti, Ajax è senz'altro una delle dimensioni più importanti del web 2.0.

Infine, una **sesta prospettiva** forse meno 'pesante' dal punto di vista degli strumenti e delle tecnologie, ma non meno rilevante per comprendere la complessa connotazione del termine 'web 2.0' è legata al **design**. L'evoluzione dei nuovi strumenti web ha infatti coinciso con lo sviluppo di siti dal design semplificato e ripensato in funzione dell'immediatezza dell'interazione, caratterizzati da una forte riduzione del numero di elementi grafici e di contenuti informativi – in particolare testuali – compresenti in una stessa pagina (ad esempio, sparisce spesso la distribuzione in colonne) e da un largo uso di colori pastellati, riflessi, ed icone di grandi dimensioni e di forte impatto.

Ma che collegamento hanno gli strumenti fin qui elencati con il tema affrontato in partenza, quello del ciclo della ricerca? E perché, in particolare, dovrebbero interessare chi si occupa del campo umanistico?

La risposta a questi interrogativi è negli innumerevoli, spesso sorprendenti esempi di uso 'accademico' o comunque validato di questi strumenti, esempi che – non a caso – riguardano (e in maniera largamente analoga) tanto le discipline umanistiche quanto le scienze 'hard'.

Così, ad esempio, praticamente tutti i siti delle maggiori organizzazioni di ricerca a livello internazionale sono gestiti attraverso CMS (e – come abbiamo visto – non si tratta solo di una scelta 'tecnica' ma dell'adozione di precisi modelli di organizzazione dell'informazione), i weblog sono usati sempre più spesso per tener traccia dello sviluppo di progetti di ricerca – individuali e collaborativi – e per raccogliere opinioni e commenti della comunità scientifica di riferimento, le piattaforme wiki (ne parleremo fra breve) sono utilizzate come strumento di redazione collaborativa di testi, podcast e videocast sono utilizzati da importanti istituzioni accademiche e di ricerca per diffondere lezioni e conferenze, siti di social network come i già ricordati *Ning* e *Facebook* ospitano gruppi di lavoro, ne facilitano la collaborazione e aiutano nell'organizzazione di incontri e iniziative. Proprio come la nascita delle riviste erudite e delle società scientifiche non ha rappresentato solo un cambiamento esteriore ma un'autentica rivoluzione nelle

modalità di organizzazione del lavoro intellettuale dell'Europa del '600, e pur con tutte le (notevoli) differenze del caso, anche la nascita e la diffusione di questi strumenti non rappresenta solo una curiosità ma l'occasione di una profonda riconfigurazione delle forme di organizzazione e distribuzione del sapere, che non può non interessare direttamente l'umanista.

Il 'caso' Wikipedia

Può essere opportuno, a questo riguardo, soffermarci brevemente su uno degli strumenti più discussi, e per certi versi più significativi, del nuovo web: Wikipedia.

Come è noto, l'idea alla base di questa iniziativa è la realizzazione di una enciclopedia universale costruita interamente in rete, in maniera aperta e collaborativa ma anche indubbiamente e quasi programmaticamente priva di garanzie di validità e qualità editoriale. L'enorme successo (almeno se si guarda al numero degli utenti e dei collaboratori) e la rapidissima espansione di Wikipedia si devono soprattutto a due caratteristiche: la natura aperta della stesura delle voci (chiunque può collaborare scrivendone o modificandone i contenuti), e lo strumento tecnico utilizzato per la sua realizzazione: un particolare tipo di sito Web, denominato 'wiki', il cui contenuto può essere facilmente integrato o modificato dai suoi visitatori. Il software di gestione tiene traccia di queste modifiche, permettendo di discuterle e, volendo, di annullarle.

Queste caratteristiche permettono la crescita collaborativa di Wikipedia, ma – almeno a prima vista – non sembrano in grado di garantirne serietà e affidabilità. Mentre una enciclopedia tradizionale prevede infatti precise responsabilità editoriali e autoriali, nel caso di Wikipedia queste responsabilità sono diffuse e difficilmente individuabili. Chi garantisce l'effettiva competenza degli autori delle voci che leggiamo? Come sono giustificati (se lo sono) inserimenti e omissioni? Come evitare che malintenzionati, esaltati, fanatici, o semplicemente incompetenti, creino o trasformino – magari in buona fede – i contenuti di Wikipedia facendone un guazzabuglio inaffidabile?

In effetti, la mancanza di precise responsabilità autoriali ed editoriali e di strumenti di controllo qualificato dei contenuti (la mancanza cioè di strumenti di *validazione*), e la conseguente arbitrarietà, inaffidabilità e scarsa qualità scientifica delle voci, rappresentano una delle critiche sollevate più frequentemente contro Wikipedia. Una critica della quale si è fatto autorevole interprete in Italia uno dei nostri massimi storici della filosofia ed 'enciclopedisti', Tullio Gregory. In un articolo comparso sul Sole 24 ore del 18 febbraio 2007, Gregory mette in fila una serie di esempi particolarmente eclatanti di errori, omissioni, inesattezze presenti nelle voci italiane di Wikipedia, e ne deriva una critica radicale all'idea stessa della costruzione totalmente aperta di uno strumento di questo genere. L'autorialità aperta e incontrollata di Wikipedia non è compatibile, a suo avviso, con serietà e affidabilità dei contenuti.

La critica è legittima e pertinente, e nei casi specifici sicuramente fondata, ma solleva un immediato interrogativo: come mai la versione inglese di Wikipedia – pur se anch'essa non priva di inesattezze – è comunque in generale assai più affidabile e completa di quella italiana, tanto da permettere già nel 2005 alla prestigiosa rivista scientifica internazionale *Nature*, al termine di un dettagliato quanto controverso

studio[7], di paragonarne l'affidabilità addirittura a quella dell'*Encyclopaedia Britannica*? I collaboratori italiani di Wikipedia sono davvero tanto più incompetenti dei loro omologhi anglofoni?

In realtà, in questo caso il problema principale è un problema di numeri: probabilmente i collaboratori della versione inglese di Wikipedia non sono, in media, più bravi o competenti di chi collabora alla versione italiana dell'enciclopedia, *ma sono molti di più*. Avendo a disposizione un numero assai maggiore di collaboratori, aumentano le probabilità che ogni singola voce 'trovi' (almeno) un estensore o revisore dotato di competenze specifiche e capacità autoriali e redazionali sufficienti a costruire un 'buon testo', o a trasformare una voce generica e imprecisa in una voce più precisa e affidabile. Un po' alla volta, collaborativamente, le voci di Wikipedia tendono a crescere e migliorare, ma questo processo richiede una 'massa critica' di collaboratori (possibilmente provenienti anche dal mondo accademico e della ricerca) sicuramente maggiore di quella finora disponibile nel nostro paese. Giustamente, dunque, nello stesso numero del Sole 24 ore Roberto Casati rileva come, paradossalmente, la risposta migliore ai problemi di Wikipedia potrebbe essere quella di… rimboccarsi le maniche e aiutarla a crescere!

Certo, il numero e la qualità dei collaboratori non sono tutto. I rischi collegati a un processo redazionale programmaticamente poco controllato e controllabile non sono mai del tutto eliminabili, e lo stesso Larry Sanger, uno dei cofondatori di Wikipedia, suggerisce oggi l'opportunità di subordinare i risultati di una collaborazione totalmente aperta alla revisione di redattori ed esperti qualificati. Il suo nuovo progetto di enciclopedia in rete, Citizendium, ha proprio queste caratteristiche (ma sembra per ora mancare dell'indispensabile massa critica di collaboratori).

Vinceranno i grandi numeri di Wikipedia, o il modello tradizionale basato su responsabilità redazionali e autoriali precise e controllate? Nonostante gli errori e gli orrori che caratterizzano oggi alcune voci di Wikipedia, in particolare nelle sue versioni non anglofone, la risposta a questo interrogativo è molto meno scontata di quanto potrebbe sembrare. E la contaminazione fra due modelli editoriali apparentemente così diversi potrebbe rivelarsi più interessante e produttiva di una loro semplice contrapposizione.

La distribuzione dei risultati della ricerca: gli archivi aperti

La rete Internet si è dimostrata in questi anni uno strumento prezioso ed estremamente efficiente per la raccolta, l'integrazione e la distribuzione (pubblicazione) di documenti e risorse informative di ogni genere. Come tale, è stata ed è ampiamente utilizzata della comunità scientifica e accademica per la pubblicazione e la diffusione di documenti, articoli, preprint, e in generale come piattaforma per la condivisione e lo scambio dei risultati della propria attività di ricerca, ed è dunque ormai entrata a pieno titolo, come si è già accennato, fra gli strumenti specifici del ciclo della ricerca.

Tale evoluzione ha reso centrale il problema di sviluppare standard e strumenti che consentano di raccogliere in maniera organizzata e affidabile, di descrivere, di conservare, di reperire tali risorse, garantendo almeno un livello minimo di interoperabilità fra le piattaforme di archiviazione e pubblicazione in rete.

[7] I risultati dello studio e una ricostruzione completa del relativo dibattito sono disponibili sul sito di Nature, all'indirizzo http://www.nature.com/nature/journal/v440/n7084/full/440582b.html (richiede registrazione).

Gli standard sviluppati nell'ambito della cosiddetta Open Archives Initiative[8] (OAI) hanno offerto una prima risposta a questi problemi, attraverso la previsione di una duplice tipologia di strumenti: 1) depositi di documenti e risorse informative, detti anche archivi o *repository*, e denominati tecnicamente *data provider*, che pur adottando, se del caso, piattaforme software, formati e politiche operative diverse, rendano disponibili ('espongano') in forma standard i metadati descrittivi relativi alle risorse informative conservate, e 2) centri di 'raccolta' (*harvesting*) dei metadati, denominati tecnicamente *service provider*, in grado di richiedere, raccogliere e integrare i metadati offerti da più data provider e di fornire su di essi servizi centralizzati (in primo luogo – ma non unicamente – di ricerca). In sostanza, anche se con qualche semplificazione, possiamo pensare ai data provider come ai singoli archivi e depositi di contenuti e informazioni in formato digitale (ad esempio articoli, preprint, documenti di varia natura), e ai service provider come a servizi di raccolta centralizzata di indici e cataloghi relativi ai contenuti ospitati da una pluralità di data provider, e di ricerca integrata su tali cataloghi.

Il 'dialogo' fra data provider e service provider avviene utilizzando un 'linguaggio' specifico e standardizzato, l'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): un protocollo che prevede l'uso di normali richieste HTTP (le stesse usate dal browser per richiedere pagine web) per l'interrogazione dei data provider da parte dei service provider, e l'uso di XML per le risposte da parte dei data provider (e dunque per l'inoltro dei metadati). I metadati offerti da un data provider attraverso il protocollo OAI-PMH possono essere in qualunque formato, anche se il formato standard Dublin Core (largamente utilizzato dalla comunità dei bibliotecari) è raccomandato per una interoperabilità di base

Questo meccanismo garantisce – attraverso la mediazione rappresentata dai service provider – la potenziale visibilità e reperibilità dei contenuti di tutti gli archivi aperti che adottano le specifiche OAI, indipendentemente dalla loro collocazione, dalle loro dimensioni, dalla natura dei loro contenuti. Potendo rispondere in pieno ai principi proposti dalla Open Access Initiative[9] in materia di accessibilità dei contenuti e utilizzazione di standard aperti, gli archivi in rete che adottano gli standard Open Archives offrono dunque uno strumento prezioso per la raccolta, la conservazione, la descrizione e la distribuzione di ogni forma di produzione intellettuale codificabile in formato digitale (e, se del caso, anche di metadati relativi a risorse non digitali). Al momento, i due sistemi OA più diffusi – entrambi distribuiti gratuitamente con la formula dell'open source e con vaste comunità di utenti e sviluppatori – sono *EPrints*, il cui sviluppo è stato avviato dalla University of Southampton[10], e *DSpace*, il cui sviluppo è stato avviato dal MIT[11].

Accanto a queste caratteristiche 'tecniche', però, il movimento che ha portato alla nascita e alla crescita degli archivi aperti è strettamente legato anche a delle scelte che potremmo dire di 'politica della ricerca'. La principale è rappresentata dal principio secondo cui i risultati del lavoro di ricerca finanziato con denaro pubblico dovrebbero essere liberamente e pubblicamente accessibili. Un principio apparentemente ovvio, ma che – per le politiche editoriali e distributive fortemente restrittive adottate da molti fra i principali editori scientifici e accademici internazionali – si rivela in realtà assai meno scontato di quanto non sarebbe desiderabile.

[8] http://www.openachives.org/.

[9] Cf. http://www.soros.org/openaccess/.

[10] Cf. http://eprints.soton.ac.uk/ e, nel panorama bibliotecario italiano, http://eprints.rclis.org/.

[11] Cf. http://www.dspace.org.

In effetti, ad essere in questione è qui uno snodo fondamentale del 'ciclo della ricerca' discusso nella prima parte di questo intervento, quello della distribuzione esterna dei prodotti della ricerca. Tradizionalmente, questa distribuzione era (e in larga parte continua ad essere) a) validata preventivamente attraverso il meccanismo di *peer review* e b) affidata a un mercato editoriale specializzato, che in cambio della gestione della fase di *peer review* e della distribuzione in forma riconoscibile e validata (tipicamente attraverso riviste internazionalmente riconosciute) dei risultati della ricerca, ne richiede uno sfruttamento economico intensivo, basato sul completo controllo dei diritti di copia e distribuzione e sulla vendita (a caro prezzo) dei diritti d'uso ai singoli utenti, privati o istituzionali. Strumenti come gli archivi aperti, e il movimento per l'accesso aperto ai prodotti della ricerca, propongono di fatto un modello di distribuzione completamente diverso: a) autoarchiviazione dei prodotti della ricerca da parte dei singoli studiosi, senza un preventivo processo di validazione se non quello, legato alla persona più che ai contenuti, rappresentato dall'accesso all'archivio istituzionale; b) validazione successiva da parte della comunità scientifica di riferimento attraverso quello che è sostanzialmente un meccanismo di *collaborative filtering* (i contenuti più validi o interessanti emergono progressivamente attraverso la crescita dei riferimenti e delle citazioni, con un processo di validazione collaborativa completamente aperto) c) libera distribuzione e accesso completamente aperto ai contenuti.

La discussione delle differenze fra questi due modelli, e dei rispettivi meriti e problemi, richiederebbe uno spazio molto maggiore di quello qui disponibile. È il caso di ricordare, tuttavia, che la realizzazione di archivi aperti e repository istituzionali è un settore che ha conosciuto un notevole sviluppo degli ultimi cinque anni, così come notevole è stata la crescita del movimento internazionale per l'open access. La *Budapest Open Access Initiative*, del Febbraio 2003, costituisce la prima iniziativa internazionale rilevante in tal senso, seguita dalla *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*[12] dell'Ottobre 2003, e, per l'Italia, dalla *Dichiarazione di Messina*[13] del novembre 2004, cui ha fatto seguito la creazione di una apposita commissione Open Access da parte della CRUI. La vitalità del movimento Open Access, e la particolare attenzione che sta ricevendo anche nel nostro paese, sono confermate anche dalla recente conferenza organizzata nel settembre 2007 dall'Università di Padova per promuovere l'open access: "*Berlin 5 Open Access: From Practice to Impact: Consequences of Knowledge Dissemination"*[14].

Mentre il mercato editoriale specializzato legato alla pubblicazione dei risultati della ricerca scientifica aveva caratteristiche assai diverse nel settore delle scienze 'hard' e in quello delle discipline umanistiche (anche perché i principi di funzionamento dei meccanismi di *peer review* erano abbastanza diversi nei due casi, tanto da far dubitare in molte situazioni della loro effettiva applicabilità all'ambito umanistico), il modello OA sembra funzionare in maniera largamente analoga in entrambi i settori. Anche per questa strada, dunque, gli strumenti della rivoluzione digitale sembrano proporre una integrazione delle 'due culture' molto maggiore di quella tradizionale, e una articolazione molto più uniforme del ciclo della ricerca.

[12] Il testo è disponibile in rete all'indirizzo http://oa.mpg.de/openaccess-berlin/berlindeclaration.html.

[13] Il testo è disponibile in rete all'indirizzo http://www.aepic.it/conf/index.php?cf=1.

[14] Cf. http://www.aepic.it/conf/index.php?cf=10**.**