

EMILIO BILLI

Storia di lago e *Internet*

Negli ultimi cinque anni noi tutti siamo stati testimoni di una rivoluzione. Attorno a noi è cambiato uno degli aspetti fondamentali della vita umana: la velocità di comunicare le idee, di replicare e salvaguardare i dati che stanno alla base di ogni conoscenza, di moltiplicare e di condividere le informazioni.

Tutto è avvenuto così rapidamente che nessuno in realtà se ne è completamente reso conto; lo stesso *modo di pensare* sta cambiando ad un ritmo impressionante, un ritmo che solo dieci anni fa nessuno sarebbe mai riuscito a prevedere: tutto questo grazie alla diffusione di *internet*.

Non è improbabile che nei prossimi anni si assisterà ad un cambiamento ancor maggiore del modo di lavorare e di pensare, proprio grazie alla continua diffusione dell'uso di queste nuove tecnologie anche in ambiti sino ad oggi ancora "vergini" e che hanno subito poca o nessuna penetrazione consistente.

Ma cosa è in realtà *internet*?

Se dovessimo chiedere a dieci diverse persone che cosa esse intendano per *internet*, con tutta probabilità riceveremmo dieci risposte molto differenti tra di loro; alcune di quelle risposte sarebbero verosimili, altre del tutto errate o addirittura fuorvianti.

In realtà, l'errore che viene fatto comunemente è quello di identificare la rete *internet* con i contenuti che essa propone, ponendo al centro dell'attenzione il contenuto dell'informazione e non il mezzo che permette all'informazione di essere disponibile. Quante volte è capitato di sentire l'affermazione "Io ho internet": nulla di più sbagliato, visto che la corretta acce-

zione è “Io utilizzo internet”, che non appartiene a nessuno, ma è di tutti.

Per capire realmente cosa sia *internet* e di conseguenza quali siano le potenzialità insite in questo mezzo, bisogna comprendere che con *internet* siamo di fronte ad una *tecnologia* che permette di *diffondere* informazioni organizzate. *Internet* è dunque una tecnologia. Ma si deve porre mente ad un fatto importante: *internet* è una tecnologia con cui è possibile condividere informazioni, e a costo di ripetermi ri-sottolineo che *essa non va banalmente identificata con le informazioni che vi si possono reperire*; si deve vedere *internet* allo stesso modo in cui l'alfabeto permette ad uno scrittore di realizzare il contenuto di un libro: nessuno mai sosterrà che l'alfabeto sia identificabile con il contenuto stesso del libro.

La essenza più intima della tecnologia *internet* fa sì che il WWW (*world wide web*) sia naturalmente *dinamico*, cioè posseda una struttura per cui ciò che noi vi troviamo sia sempre e comunque frutto di una nostra scelta, anche quando ciò non appare subito evidente. Se decido di visitare una pagina, quindi, compio una scelta; le mie scelte sono quelle che, usando un termine comune del gergo di *internet*, mi permettono di “navigare”. La sorpresa, l'inatteso stanno sempre dietro l'angolo.

Questa caratteristica di rispondere alle nostre domande restituendoci dei contenuti che in qualche modo soddisfino le nostre richieste permette di intuire le enormi potenzialità offerte dalla tecnologia *internet* nella ricerca scientifica, sia essa di tipo tradizionalmente inteso (medicina, chimica, astronomia, ingegneria, matematica,) oppure concerna discipline più “umanistiche” (archivistica, storiografia, legge, botanica, psicologia, antropologia, lettere ecc.).

Come molto spesso accade, il motore principale di una innovazione è l'*economia*, o meglio la possibilità di sfruttare un'idea o

una tecnologia per produrre ricchezza e muovere denaro. Il principale mezzo per muovere denaro è il commercio; questo spiega perché la principale applicazione delle potenzialità di *internet* sia il cosiddetto *commercio elettronico*. Ma attenzione: si tratta di una scelta, non di un limite.

Questo tipo di applicazione venale ha favorito lo sviluppo di una serie di tecnologie collaterali; a loro volta esse hanno permesso di realizzare sistemi in grado di riconoscere i nostri gusti, analizzando in maniera automatica e trasparente (ovvero senza che ce ne rendiamo conto) le scelte compiute durante la “navigazione”, ricavandone – un poco subdolamente, potrebbe concedersi ai critici! – i nostri interessi, al fine di guidarci e consigliarci, esattamente come farebbe un commesso all’interno di un negozio, o un maggiordomo in una casa nobile (e non a caso, *majordomo* è uno degli strumenti informatici in uso in certi siti *internet*...). Queste tecnologie, se applicate all’analisi di dati di tipi e forme diverse (non necessariamente solo di tipo commerciale), costituiscono una serie di strumenti eccezionali per potenza e velocità, che possono essere impiegati per l’extrapolazione di relazioni e collegamenti tra i dati stessi: esattamente come succede a chi, compilando un indice di un volume, si rende conto delle relazioni intercorrenti tra i personaggi citati nel testo e arriva a comprenderne i legami (i *link*, verrebbe da dire).

Qualsiasi tipo di ricerca si basa sull’analisi – ma anche sulla sintesi – di informazioni; diventa quindi evidente quanto sia grande l’importanza di reperire più informazioni possibili, e al contempo la capacità di poterle analizzare nella maniera più completa, strutturata e veloce.

Tale considerazione ci fa comprendere quale valore abbia una tecnologia come *internet*, applicata alla ricerca: la sua totale mancanza di limiti geografici permette di poter raccogliere, organizzare, concentrare e condividere con il prossimo una infi-

nità di dati ed informazioni provenienti da qualsiasi parte del mondo; internet li rende disponibili in qualsiasi altra parte del mondo, in qualsiasi momento, e con una capillarità di ricerca che può giungere al singolo vocabolo.

Ma non finisce qui: internet è una tecnologia *attuale* e soprattutto *semplice*: grazie ad essa si possono creare archivi di dati ben organizzati, compatti ed efficaci, attraverso gli strumenti comunemente chiamati *database*¹, utilizzando cioè un sistema digitale per cui è possibile catalogare sulla base di criteri definibili dall'utente qualsiasi tipo di informazione, compiendo poi su aggregati di informazioni i più svariati e complessi tipi di operazioni di analisi, di sintesi, o di statistica.

Questi strumenti, chiamati *database relazionali*², sono in grado di porre in relazione tra di loro le informazioni che vi si trovano contenute, realizzando operazioni di confronto; le operazioni di analisi su milioni di dati hanno luogo tipicamente in pochi secondi.

Abbiamo quindi a disposizione potenti sistemi per archiviare e lavorare su dati di qualsiasi natura; basta realizzare

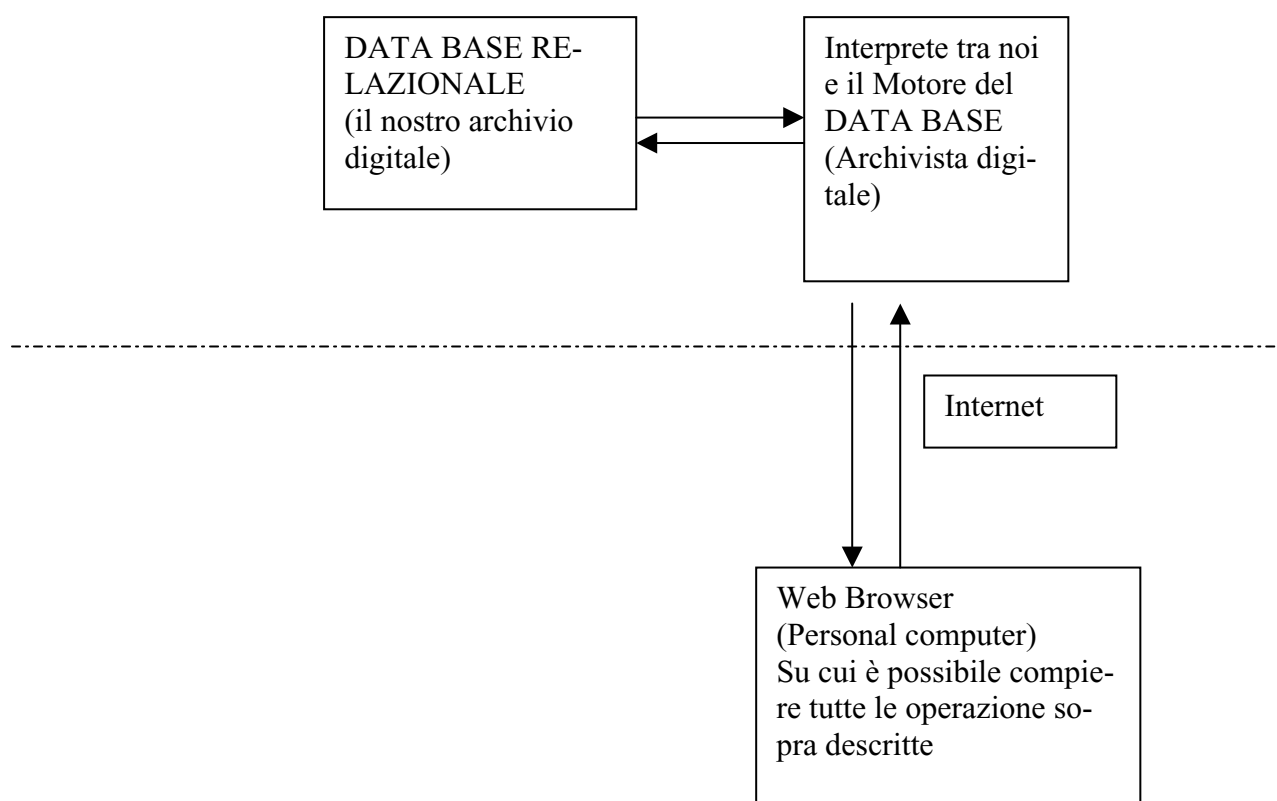
¹ In italiano la parola *database* viene letteralmente tradotta come *base di dati*. La traduzione non rende in maniera ottimale l'idea di cosa sia effettivamente un *database*. *Schedario elettronico* è probabilmente una definizione migliore: un insieme di archivi, correlati tra loro, contenenti dati omogenei che riguardano un certo argomento. La considerazione più importante è che un *database* non è un archivio statico, nel senso di un deposito nel quale immagazzinare dati e dimenticarli; lo si può invece assimilare ad un luogo dove ci si reca per cercare informazioni. Attenzione, però: se si cercano informazioni in luoghi dove esse sono immagazzinate senza un criterio, prima o poi si riuscirà a trovare ciò che si sta cercando, ma si impiegherà molto tempo. Un *database* è invece un luogo progettato per organizzare i dati in maniera tale che questi possano essere resi disponibili a chi ne fa richiesta nel minor tempo possibile.

² In questo caso non si intende l'archivio ma il programma di gestione degli archivi che compongono il *database*. Spesso infatti la parola viene usata per indicare indifferentemente sia l'archivio elettronico che il suo sistema di gestione. Più correttamente bisognerebbe chiamare quest'ultimo DBMS (*Data Base Management System*, sistema di gestione del *database*).

un'infrastruttura che lavori su questi dati, la quale ci permetta innanzitutto di introdurre le informazioni nel nostro archivio e di dialogare con quest'ultimo, informandolo sul tipo di ricerche che siamo intenzionati a compiere. Questa funzione viene svolta da una serie di *linguaggi* che altro non fanno se non tradurre il *nostro* linguaggio in una sintassi comprensibile alla macchina³. Abbiamo quindi tutto quello che ci serve: un potente archivio in cui introdurre tutti i dati che vogliamo e un interprete versatile che ci permette di dialogare con "l'archivista" responsabile del nostro archivio (inteso come l'entità che conosce le chiavi di utilizzo della base dati). Quel che manca, e di cui abbiamo bisogno, è di poter comodamente accedere al nostro archivio, magari per portarvi nuovi documenti, talora per compiere ricerche; vorremmo poi farlo possibilmente nella maniera più comoda possibile, negli orari più impensati e che più ci aggradano, e soprattutto da qualsiasi parte del mondo in cui ci capita di trovarci, curiosando nei depositi di dati di qualsiasi altra parte del mondo ci interessi effettuare un *data-mining* ("escavazione di dati"). Il segreto del successo di *internet* sta proprio qui: in ogni dove, nei cinque continenti (e perché no, pure in Antartica...) le modalità di accesso alla rete sono standard: il punto comune di ingresso (realizzato attraverso

³ Nel 1994 Rasmus Lerdorf sviluppò una serie di strumenti in grado di interpretare e manipolare strutture organizzate dette *macro*. Le operazioni che era possibile svolgere non erano straordinarie, ma da queste semplici operazioni Lerdorf finì per combinare ed organizzare un linguaggio che permetteva di interfacciarsi in maniera naturale alle basi di dati. In seguito i programmatori di tutto il mondo (seguendo un modello chiamato *open source* - "sorgente aperta" dove per *source* si intendono le "righe di codice di programmazione" di libero studio, modifica e utilizzo - secondo il quale la cultura ed il progresso non devono essere utilizzati per scopi di lucro, ma per fornire strumenti utili a disposizione di tutti coloro che ne fanno richiesta) hanno sviluppato, organizzato e reso disponibili una serie di linguaggi, sul modello ideato da Lerdorf, molto versatili: essi hanno originato la rivoluzione *internet* a cui noi stiamo assistendo.

il cosiddetto *client*, ad esempio un Personal Computer) è facilmente condivisibile e utilizzabile.



La portata di queste nuove tecniche e di uno strumento, come quello che abbiamo descritto, nel campo della ricerca storica ed archivistica è enorme: basti pensare ad esempio come sia possibile, avendo inserito opportunamente i dati relativi alle caratteristiche di alcune tecniche per la produzione di manu-

fatti in una data epoca, ed avendo inoltre classificato anche tutti i dati a disposizione sui ritrovamenti fatti (indipendentemente dai confini geografici, magari ad opera di più ricercatori) andare a scoprire le relazioni commerciali, le interazioni socio-economiche, i flussi temporali di scambio delle merci, talvolta tra popolazioni distantissime tra loro; da tali valutazioni si possono dedurre i movimenti migratori di intere etnie. Si badi bene: i tempi di “reazione” sono di gran lunga inferiori a quelli che si avrebbero conducendo una ricerca con metodi tradizionali. Ma si curi altrettanto bene a non sopravvalutare *internet*, che rimane uno strumento potente solo se chi lo usa è generoso con *internet* stesso. Ad esempio, *internet* non sostituisce il ruolo della persona coinvolta nella ricerca; *internet* non supplisce con alcun apporto di fantasia alla eventuale limitata immaginazione del ricercatore; il risultato ottenuto, semplicemente, è frutto di un campionamento che statisticamente gode di una maggior ricchezza di dati su cui operare (la *base dati* sopra descritta) e quindi ottiene un’analisi più dettagliata (e quindi statisticamente più attendibile) di quella possibile con mezzi tradizionali. È come dire che un sondaggio relativo al gradimento di un prodotto (un programma televisivo, un alimento, un libro, un convincimento politico) è maggiormente plausibile se il campione significativo della popolazione intervistata è costituito da diecimila persone, piuttosto che da cento individui. Per ricorrere ad un altro paragone, altrettanto calzante, a mio avviso, lo strumento di indagine basato sull’utilizzo di *data base relazionali* strutturati attraverso una rete in condivisione (come è appunto quella di *internet*) rappresenta l’equivalente archivistico del microscopio elettronico in microbiologia: si tratta, a tutti gli effetti, di uno strumento che non va evidentemente a sostituire il biologo, ma certamente permette di superare i limiti fisici della microscopia ottica, aprendo nuove possibilità: ricorrendo, cioè, a *nuove tecnologie*. Mi si scuserà, dunque, se

approfitto ancora una volta di un paragone che potrebbe sembrare dissacratorio e fin troppo incendiario: avversare (o peggio *ignorare*) lo strumento *internet* nella nostra era sta al pari del combattere, nel Seicento, la rivoluzione cognitiva copernicana, fidando solo dell'autorità indiscussa (ma ormai fossilizzata) di Tolomeo e Aristotele. E pur *internet*, come il mondo, si muove: Galileo *docet*.

V'è di più: la costruzione di banche dati organizzate secondo un modello come descritto sopra, può rivelarsi vincente non solo come strumento di analisi ma anche come fonte di nuove scoperte e di nuove teorie. L'importanza del muoversi in questa direzione è sottolineato da un aspetto che spesso sfugge al senso comune: esso è uno dei concetti più moderni e affascinanti che siano stati scoperti negli ultimi tempi: in ardita contraddizione (almeno in natura) con l'assioma che *Natura non facit saltus*, si ritiene oggi che le interazioni che portano il corso della storia in una direzione piuttosto che in un'altra, o ancora i movimenti economici, o la scomparsa (talora repentina e apparentemente inspiegabile) di alcune specie animali o botaniche non siano frutto di evoluzione secondo criteri lineari, ma siano piuttosto il risultato di una *non linearità*. In particolare, pare di poter teorizzare che le conseguenze sono fortemente dipendenti dagli eventi appena occorsi, per cui una piccola deviazione dalle condizioni iniziali (e quindi – si potrebbe dire – dalle *ipotesi*) porta il risultato finale a divergere, con una grande deviazione dall'atteso (se fosse dedotta applicando le condizioni di linearità); si origina quindi uno scostamento delle deduzioni dalla realtà dei fatti, non a causa di ipotesi errate, ma piuttosto per un modello deduttivo che non corrisponde con efficacia alla realtà.

In modo analogo, la ricostruzione storica deve per forza di cose mediare tra una ricerca infinita e una giusta approssimazione, dettata dalla ovvia necessità di concludere le indagini in

tempo utile e dalla considerazione di un limitato raggio di analisi, in cui inevitabilmente ci si imbatte allorché si compie una ricerca documentale storica.

Lo storico (che non necessariamente coincide con chi abbraccia con convinzione i metodi *internet*) si trova dunque combattuto tra due esigenze: se da un lato non gli è evidentemente possibile (e tutto sommato non gli interessa neppure) individuare e classificare analiticamente *tutte* le cause di una determinata realtà storica, dall'altro è altrettanto chiaro che tante più interazioni gli riesce di definire ed individuare (e dunque quante più fonti si riescono a saggiare), maggiormente precisa potrà essere la ricostruzione. Il risultato ultimo è la maggior plausibilità dell'analisi condotta, e di conseguenza l'efficacia del disegno cui si aspira a dar corpo.

L'asserzione vale con forza ancor maggiore per chi usa *internet*: basti indicare, tra i mille e mille possibili esempi, la sterminata bibliografia virtuale offerta al lettore dai siti di www.bnf.fr (biblioteca nazionale francese), www.loc.gov (biblioteca del Congresso americano), <http://opac.sbn.it> (Istituto Centrale del Catalogo Unico, valido sito documentale italiano): per essi quanto sopra sostenuto è oramai un postulato, una verità che non è più un discutere: maggiori sono i dati di partenza (ricchezza in continua espansione dei cataloghi bibliografici) maggiore sarà la possibilità di analisi (facile e accurata redazione di bibliografie) e maggiori quindi le scoperte di referenze bibliografiche altrimenti impensabili.

Ma si sbaglierebbe considerando *internet* solo come un sistema "generoso" nel fornire dati gratuitamente, su ogni argomento dello scibile umano, con assoluta rapidità: non si deve infatti dimenticare che dietro ogni notizia, rinvenuta in *internet*, sta il lavoro di immissione da parte di un operatore (remunerato o meno); è dunque importante ricordare che in questo complesso sistema è fondamentale non solo disporre della facoltà di

trarre dati, ma anche di quella di inserirne di nuovi, senza nessun tipo di confine; in ciò sta una vera “democrazia culturale” che permette di riconoscere i sempre più ridotti ambiti in cui non si dispone di dati, e quindi i vuoti di conoscenza del sistema che vanno colmati. E sono convinto che, differentemente da altre discipline, quelle umanistiche in genere (e storiche e archeologiche in particolare) siano tra le aree dove il lavoro di completamento delle banche dati è ancora ben lungi dall’essere concluso, o almeno esaurito con sufficiente livello di esaustività. Questa doppia connotazione (il prendere da *internet*, il dare ad *internet*) apre sicuramente la strada ad un nuovo *modus operandi*, che deve essere cosciente dell’esistenza di limiti. Essi sono ovviamente contenuti nella capacità di *insegnare* ad un sistema digitale a compiere analisi su dati viziati da imprecisioni o da discrepanze rispetto alla base dati; esempio tipico può essere dato dalla possibilità che una stessa persona venga citata da fonti diverse con nomi leggermente differenti. Ciò non deve però essere un limite, ma un punto di partenza per la realizzazione di strumenti sempre più raffinati, ad esempio quelli che siano basati su nuovi modelli logici aperti anche all’apprendimento e all’autocorrezione. È il caso degli esperimenti, tanto in auge negli ultimi tempi, di *intelligenza artificiale*⁴.

⁴ La ricerca storica può essere paradossalmente anche fonte di problematiche di tipo informatico. Le tecniche derivanti dall’impiego dell’intelligenza artificiale possono realizzare sistemi molto sofisticati per l’indagine archivistica. La progettazione di software derivato in qualche misura dall’intelligenza artificiale o dalla logica chiamata *fuzzy* (e cioè quella ‘logica lasca’ che venne sviluppata soprattutto a Berkeley a partire dagli anni ’60 del Novecento, per poter maneggiare i concetti di ‘verità parziale’ – affermazione logica ‘non completamente vera’, affermazione logica ‘non completamente falsa’ – oltre che i due gradi di logica ‘booleana’ del ‘vero’ assoluto contrapposto al ‘falso’ assoluto) possono far avvicinare il *modus operandi* di una macchina a quello delle persone, cosa che può rendere le macchine più potenti e più semplici da utilizzare. Infatti, la conoscenza *fuzzy* si traduce in regole *fuzzy*; la logica *fuzzy* implica la costruzioni di deduzioni con un grado di verità non assoluto. Una

E vengo alla parte più direttamente legata alle nostre terre verbanesi. Quanto detto sopra in generale trova un immediato e ottimo campo di applicazione particolare nel lavoro che si deve compiere per arricchire le nostre conoscenze organizzate relativamente alla storia verbanese (e non solo alla storia).

Il prezioso lavoro sin qui condotto (e ben lungi dall'aver mai fine, tanto è vasto) da gruppi e associazioni verbanesi che agiscono in modo "tradizionale" ha assolutamente bisogno, in un'era in cui si conoscono più facilmente dati e fatti generali relativi a luoghi lontani, che non particolari minuti riguardanti la storia locale, di essere completato dalla creazione di uno strumento di interpretazione incrociato, che consenta di effettuare nuove deduzioni e creare legami tra fatti e nomi difficilmente conoscibili, o tracciabili: in questo senso ha senso cercare di capire chi fossero persone talora oscure, o assolutamente

regola *fuzzy* correla i concetti nella forma di enunciato condizionale: *Se X è A allora Y è B*. Ad esempio se il traffico è ELEVATO allora mantenere più A LUNGO il verde. È solo questione di buon senso. La regola strutturata che stiamo cercando è quella che insegna alla macchina ad associare il concetto di traffico ELEVATO alla soluzione di mantenere il verde PIÙ A LUNGO. Sono le combinazioni possibili tra la densità del traffico e la durata dell'illuminazione il problema più complesso da far analizzare ad una macchina, poiché queste combinazioni sono un numero infinito. Ciò nonostante, esiste un'unica regola *fuzzy* che le correla tutte.

Allo stesso modo è possibile realizzare *database fuzzy* che vengano pilotati da regole di buon senso. Un buon modello potrebbe essere quello di far attivare una analisi *fuzzy* ad ogni immissione di nuovi dati. In un secondo tempo il sistema si incaricherebbe di registrare i risultati ottenuti, in modo da poter compiere se necessario dei successivi "ragionamenti" su questi dati e fornire delle risposte finali molto complesse, basate appunto sul buon senso; e addirittura si potrebbe insegnare al sistema a sviluppare tattiche e strategie complesse, a seconda del grado di probabile verità, falsità o buon senso insito in una catena di deduzioni logiche. A seguito dello sviluppo di un modello *fuzzy*, ad esempio, un computer potrebbe stimare le mosse di un giocatore di scacchi, e per ciascuna apertura strategica immaginare molti e diversificati scenari di gioco; nell'evoluzione della partita, ad ogni successiva mossa verrebbero inoltre scartate le mosse ritenute inadeguate, mantenuti gli scenari valutati tuttora probabili e ipotizzati altri nuovi, scaturiti dalle mutate condizioni del gioco.

normali, senza le quali, però, l'opera di gente famosa, di personaggi importanti (artisti, uomini politici, storici, letterati, scienziati) sarebbe stata forse meno efficace.

Si potrebbe obiettare che le nostre riviste di storia e cultura locale stanno facendo ciò da anni, se non da decenni e decenni: vero. Ma è altrettanto vero che quelle stesse riviste, oltre ad essere spesso appannaggio quasi esclusivo di specialisti⁵, o appassionati, finiscono per lo più per non essere di facile accesso ed uso: in certe mancano ad esempio gli strumenti degli indici analitici, di altre le collezioni complete trovano a fatica, solo in certe biblioteche, e sono talvolta di non agevole consultazione, sottostando ai criteri di prestito e di apertura delle pubbliche sale di lettura. E anche ammettendone la possibilità e comodità di consultazione, le opere cartacee hanno alcuni limiti assai pesanti: il primo consiste nella serializzazione delle informazioni, a cui si può accedere solo disponendo di strumenti (essenzialmente costituiti dagli indici analitici e sommari) che – pur ammessane l'esistenza – sono comunque limitati in potenzialità (ad esempio non si può cercare se non tra nomi di luogo e di persona, mancando per evidenti motivi di costi nelle riviste un indice dei termini lessicali) e predefiniti, oltre che immutabili per il futuro (la stampa non può essere corretta, se non a prez-

⁵ Sono assolutamente convinto (e non sono il solo) che un nuovo modo di fare cultura sia alle porte, e ritengo che nei prossimi anni si assisterà ad un fenomeno analogo a quello vissuto dalla *scienza dell'informazione* negli ultimi cinque anni, ovvero la nascita di movimenti *open* il cui principale scopo sia quello di andare ad arricchire e diffondere la conoscenza senza creare circoli elitari.

Agli inizi degli anni Novanta del secolo scorso movimenti culturali americani ben presto approdati anche in Europa hanno sconvolto i modelli tradizionali di intendere l'ingegneria del software mandando in crisi istituzioni universitarie prestigiose (ad es. Università Berkeley) e potenti aziende impegnate nel campo della produzione di software. Tutto questo è stato permesso grazie alla rete internet; analogamente a quanto sto constatando di persona nel campo umanistico, anche questi movimenti furono all'inizio *snobbati*, mentre ora sono l'unica fonte certa di software di qualità e il punto focale di lavoro di molte università anche italiane (Milano, Pisa, Trieste, tanto per citare qualche *Alma Mater*...)

zo di una riedizione). Il secondo limite, poi, consiste nell'impossibilità di creare criteri di ricerca sofisticati, che prevedano la combinazione logica di più condizioni. Si pensi ad esempio la ricerca dei termini “*agrumi* [e/o *limoni*, *aranci amari*, *bergamotti*, ecc.], *lago Maggiore* [e/o *Verbano*], *commercio* [e/o *traffico* e/o *merci*], *Svizzera* [e/o *Ticino*], *secolo XVIII* [e/o *Settecento*”]: questa cospicua serie di condizioni (che non si possono formulare assieme in una ricerca sul cartaceo, se non laboriosamente e a prezzo di alti costi temporali) risponde alla curiosità di chi sia interessato a conoscere le modalità di sviluppo dei commerci degli agrumi nel Settecento tra Lago Maggiore e Svizzera. Lo strumento *internet*, invece, supera di slancio i vincoli strumentali le tradizionali banche dati cartacee, e consente in pochi secondi di conoscere la risposta al quesito. In tutto ciò sono fondamentali sia la disponibilità di ampie basi dati, sia la capacità dell'utente *internet* di esercitare il proprio acume (e perché no? la fantasia...), suggerendo al sistema una valida e non banale chiave di ricerca (e suoi successivi adattamenti e riconfigurazioni): pena la risposta negativa (“Non sono stati trovati dati relativamente all'argomento richiesto”) o la risposta generica (e scoraggiante!) “Sono state trovate *x migliaia* di occorrenze del termine richiesto”.

Nasce in questo clima di sfida teorica un progetto pratico: Il Magazzino Storico Verbanese, un laboratorio di ricerca che da un lato offrirà strumenti innovativi al ricercatore sia esso uno storico, un economo, un botanico, da un altro costituisce un terreno di studi per le *basi di dati*: terreno che se opportunamente coltivato porterà alla realizzazione di un sistema complesso.

La struttura di base del *Magazzino Storico Verbanese* è quella del *portale*: in esso è possibile innanzitutto trovare informazioni di interesse generale relative al materiale contenuto (come

quelle ad esempio presenti in un giornale, dove l'occhio è attratto prima di tutto dai titoli, opportunamente posti in risalto), o approfondire la ricerca, far emergere spunti per nuove ricerche, o semplicemente ricevere informazioni sullo stato dei lavori in un determinato settore. Questa struttura viene edificata su di un *database*, che permette di implementare tutte le opzioni e le funzionalità descritte sopra.

Su questo oggetto complesso va ad inserirsi il *Magazzino* vero e proprio, inteso come applicazione software e contenitore intelligente, strutturato secondo precisi criteri che ricalcano la struttura di un archivio tradizionale: superandone, ovviamente, i limiti fisici già evidenziati in precedenza nei casi generali descritti nella prima parte di questa relazione.

Ad una analisi strutturale il *Magazzino* appare costruito su una serie di livelli. Il primo livello è il *livello macchina*, ovvero lo strato logico dove viene svolto quasi tutto il lavoro 'meccanico' organizzativo delle applicazioni che vanno a costituire il *Magazzino*; sopra, come strati di maggior 'intelligenza' poi vi sono il sistema operativo e la *base di dati*.

La struttura è tale da permettere l'inserimento di dati e la loro consultazione da più postazioni, da luoghi differenti ed in maniera indipendente e coordinata.

Questa caratteristica rende possibile una rapida crescita delle informazioni contenute nella base dati ed una fruibilità senza confronti, andando ad annullare i *tempi macchina* necessari per eventuali aggiornamenti richiesti in sistemi strutturati in maniera tradizionale. La *base dati* viene realizzata utilizzando una applicazione relazionale strutturata a tabella in grado di gestire velocemente e facilmente più richieste parallele. La struttura deve poi rispondere ad esigenze di robustezza tali da garantire l'integrità dei dati in essa contenuti.

Una struttura adeguata al *Magazzino* deve essere pesata e modellata per supportare l'archiviazione e la consultazione di dati

da parte di una intera comunità e non solo di un singolo utente.

Un punto molto complesso e delicatissimo è poi la *gestione degli errori*: poiché è possibile che uno o più dati possano essere inseriti *contaminati all'origine*, bisogna pensare ad una struttura che non sia facilmente compromissibile da questi ultimi. Il metodo migliore per gestire una serie di dati complessi e correlati è quello di creare una base dati distribuita con archivi posto in precisa relazione tra di loro. In più questo sistema offre inoltre numerosi vantaggi, in quanto tende ad organizzare il materiale in strutture gerarchiche. Va da sé, poi, che una simile struttura deve essere protetta da anomalie di cancellazione, che possono in qualche modo compromettere i contenuti della *base dati*.

Diventa fondamentale, a fronte di quanto detto prima, implementare un processo di *normalizzazione*⁶ efficiente ed efficace in fase di progetto. Un fattore importante nella normalizzazione dei dati nel Magazzino è *l'analisi delle dipendenze transitive*⁷. Tutto questo lavoro di analisi richiede un cospicuo dispendio di energie in fase di progetto della *base dati* e un gran-

⁶ Il processo di normalizzazione è vitale per poter lavorare con una base dati relazionale. Quest'ultimo è composto da diverse forme "normali" (Forma normale *Boyce-Codd*, *Quarta forma normale*, *Quinta forma normale*, forma normale *Dominio/Chiave*) Essenzialmente queste forme consistono nell'organizzare i dati in una struttura a *tabella* assicurandosi che essi rispondano a delle caratteristiche ben precise:

1. Ogni colonna deve contenere un valore "atomico" (Questo significa che ogni cella deve contenere un solo valore).
2. Ogni colonna deve avere un nome unico.
3. La tabella deve avere un insieme di valori che identifichino univocamente le righe (questa è la cosiddetta *chiave primaria* di identificazione).
4. Due righe non possono essere identiche.
5. Non sono ammessi gruppi di dati ripetuti.

⁷ Una *dipendenza transitiva* è una situazione dove esiste una colonna che non è direttamente dipendente dalla chiave primaria, mentre il suo contenuto dipende per qualche motivo da un altro campo, che a sua volta dipende dalla chiave primaria. Questi tipi di dipendenze possono creare dei gruppi ripetuti di dati.

de sforzo nel controllo di congruenza della struttura definita in teoria; ma una volta realizzato quanto descritto sopra ciò che otterrà sarà uno strumento semplicissimo, ma al contempo estremamente potente di analisi e ricerca; esso sarà soprattutto dotato di una flessibilità non raggiungibile né immaginabile con metodi tradizionali basati su supporti cartacei.

Ecco dunque detto in breve, e spero con chiarezza, lo scopo del Magazzino Storico Verbanese, che altro non ambisce ad essere il “braccio informatico” dei diversi gruppi di ricerca e studio che hanno il Verbano e la sua ricca cultura in oggetto. Lo scopo del Magazzino, oltre che nell’emulare le certose redazioni di “selve di dati” e “magazzini storici”, compiute da uno dei precursori dei moderni storici di lago, Giovanni Andrea Binda (1803-1874) sta nelle altrettanto lontane (1905) e lungimiranti parole di Attilio Bricchi (1875-1968), mai dimenticato maestro di cose verbanesi, da cui speriamo di mutuare grinta e passione infinite per il nostro Verbano: « Far meglio conoscere il lago Maggiore ai figli suoi, innamorarli delle sue bellezze naturali ed artistiche, evocare alle anime loro le memorie dei padri; illustrare a loro esempio le nobili figure degli uomini superiori che passarono tra noi, studiando, lavorando, beneficiando...».

Ecco lo scopo del Bricchi; ecco – affiancato dalle nuove metodiche di ricerca e studio in *internet* – lo scopo del Magazzino Storico Verbanese.