

**Informazione,  
conoscenza e Web  
per le scienze umanistiche**

Silvana Castano, Alfio Ferrara, Stefano Montanelli

# **Informazione, conoscenza e Web per le scienze umanistiche**



© 2009 Pearson Paravia Bruno Mondadori S.p.A.

Le informazioni contenute in questo libro sono state verificate e documentate con la massima cura possibile. Nessuna responsabilità derivante dal loro utilizzo potrà venire imputata agli Autori, a Pearson Paravia Bruno Mondadori S.p.A. o a ogni persona e società coinvolta nella creazione, produzione e distribuzione di questo libro.

Per i passi antologici, per le citazioni, per le riproduzioni grafiche, cartografiche e fotografiche appartenenti alla proprietà di terzi, inseriti in quest'opera, l'editore è a disposizione degli aventi diritto non potuti reperire nonché per eventuali non volute omissioni e/o errori di attribuzione nei riferimenti.

I diritti di riproduzione e di memorizzazione elettronica totale e parziale con qualsiasi mezzo, compresi i microfilm e le copie fotostatiche, sono riservati per tutti i paesi.

LA FOTOCOPIATURA DEI LIBRI È UN REATO Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.

Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da AIDRO, corso di Porta Romana n. 108, 20122 Milano, e-mail [segreteria@aidro.org](mailto:segreteria@aidro.org) e sito web [www.aidro.org](http://www.aidro.org).

Copy-editing: Federica Sonzogno

Impaginazione: TOTEM di Andrea Astolfi

Grafica di copertina: Nicolò Cannizzaro

Stampa: Tip.Le.Co. – S. Bonico (PC)

Tutti i marchi citati nel testo sono di proprietà dei loro detentori.

978-88-7192-548-6

Printed in Italy

1ª edizione: febbraio 2009

Ristampa  
00 01 02 03 04

Anno  
09 10 11 12 13

# Sommario

Prefazione	IX
Introduzione Il ruolo dell'umanista	XI
La natura e la storia dell'informatica	XI
Il digital divide	XII
Informatizzare gli umanisti	XIII
<b>Capitolo 1 Dati, informazioni, sistemi informativi</b>	<b>1</b>
Memorizzazione dei dati e basi di dati	2
I sistemi di gestione di basi di dati (DBMS)	2
Dati e livelli di astrazione	4
<b>Capitolo 2 Semantica dei dati: ontologie e rappresentazione della conoscenza</b>	<b>7</b>
Ontologie e informatica: concetti introduttivi	9
Ontologie: definizioni e contenuto	10
<b>Capitolo 3 La progettazione concettuale</b>	<b>15</b>
Entità e attributi	15
Relazioni fra entità	16
Cardinalità delle relazioni	17
Identificatori	19
Gerarchie di generalizzazione	20
Indicazioni metodologiche per la progettazione	21
<b>Capitolo 4 Il modello relazionale</b>	<b>25</b>
Teoria del modello relazionale	26
Valori nulli	28
Relazioni fra tabelle	29
Coerenza dei dati	30
Algebra relazionale	32
Proiezione e selezione	32
Operazioni insiemistiche	35

<b>Capitolo 5 La progettazione logica</b>	39
Eliminazione delle gerarchie di generalizzazione	40
Traduzione delle entità e degli attributi	40
Traduzione delle relazioni	41
<b>Capitolo 6 Creazione e interrogazione di una base di dati: il linguaggio SQL</b>	45
Creazione di una base di dati	46
Definizione di associazioni fra tabelle	50
Popolamento di una base di dati	51
Violazione dei vincoli di integrità referenziale	54
Modifica ed eliminazione di tabelle	55
Interrogazione di una base di dati	56
Condizioni di selezione	58
Join di tabelle	59
Criteri di ordinamento	61
Ulteriori esempi di interrogazione	62
<b>Capitolo 7 Analisi e classificazione automatica di testi</b>	65
Information Retrieval	65
Tecniche di analisi	68
Tecniche di indicizzazione	70
Tecniche di interrogazione e matching	71
Valutazione dei risultati di una ricerca	73
Considerazioni	75
<b>Capitolo 8 Architettura e standard per il Web</b>	77
Breve storia di Internet e del World Wide Web	77
Struttura e funzionamento della rete Internet	78
Indirizzamento degli elaboratori	80
DNS e struttura dei domini	81
Protocolli di comunicazione	83
Il servizio World Wide Web	85
Lo standard URI	87
Richiesta di una pagina web	87
Pagine statiche e pagine dinamiche	89
Altri servizi di livello applicativo: la posta elettronica	90
<b>Capitolo 9 Contenuti sul Web: i linguaggi di marcatura</b>	93
HyperText Markup Language	93
Definire stile e formattazione dei documenti con CSS	97

---

<b>Capitolo 10 I motori di ricerca</b>	103
Funzionamento di un motore di ricerca	104
Esempi e tipologie di motori di ricerca	105
Calcolo della rilevanza di una pagina web: il caso PageRank di Google	106
Utilizzo di un motore di ricerca: esempi con Google	107
Servizi e strumenti aggiuntivi dei motori di ricerca	111
<b>Capitolo 11 Progettazione di siti web: usabilità e accessibilità</b>	113
Principi generali di usabilità	114
L'approccio di Jakob Nielsen	115
Usabilità e sforzo cognitivo	117
Valutazione di un sito web	120
<b>Capitolo 12 L'evoluzione del Web: Web 2.0 e Semantic Web</b>	121
Web 2.0: da un Web di documenti a un Web di applicazioni	121
Semantic Web: da un Web di documenti a un Web di informazioni	123
<b>Capitolo 13 Gli strumenti essenziali: XML</b>	127
Cos'è e cosa non è XML	127
Struttura di un documento XML	129
Elementi di sintassi	131
XML namespace	133
Validazione e XML schema	136
<b>Capitolo 14 Editoria elettronica e biblioteche digitali</b>	139
Metadati e interoperabilità	140
I problemi dell'editoria elettronica	141
Nuovi fenomeni e nuovi mercati	144
<b>Capitolo 15 Organizzazione dei contenuti: ipertestualità e multimedialità</b>	147
Codifica e organizzazione testuale	147
Testi e strumenti digitali	148
Modelli del testo	152
Ipertesto, multimedialità e principi di progettazione	153
<b>Capitolo 16 Blog, comunità e social networking</b>	157
Autorialità sul Web: blog e spazi di pubblicazione	157
Comunità di utenti e interazione sociale	160
Directory di contenuti e portali collaborativi	163
La vastità dei fenomeni sociali sul Web	165

<b>Capitolo 17 Nuove modalità di fruizione dei contenuti: topic map e ontologie</b>	167
Il linguaggio RDF	168
Il linguaggio RDFS	171
Semantica formale e ontologie	173
<b>Riferimenti bibliografici</b>	175

# Prefazione

*Informazioni, conoscenza e web per le scienze umanistiche* è un testo che si rivolge in primo luogo agli studenti dei corsi di laurea di area umanistica in cui si svolgano insegnamenti introduttivi di informatica.

L'insegnamento dell'informatica è solo uno degli ambiti in cui si propone oggi il tema dell'incontro fra le discipline umanistiche e le scienze dell'informazione. Sempre di più l'informatica si presenta, nella vita professionale e in quella personale, come un sapere generale, necessario alla cultura di ognuno così come lo sono i fondamenti della conoscenza della lingua o dell'aritmetica.

L'informatica è un territorio nuovo su cui si giocano diverse competenze esistenti e non si riduce all'uso del calcolatore ma, al contrario, si presenta come teoria dell'informazione tradotta in oggetto del consumo e in metafora del mondo reale. Non è dunque possibile usare criticamente gli strumenti dell'informatica prescindendo dalla conoscenza della teoria che essa comporta: non basta il saper fare o il saper usare, ma occorre il sapere in sé, sapere cosa sia l'informatica e come operi sulle informazioni che comunichiamo e manipoliamo. Trova dunque conferma il motto baconiano "Ipsa Scientia Potestas Est", poiché nel mondo in cui viviamo cresce la differenza fra chi sa usare e chi conosce e, anche agli umanisti, è sempre più spesso richiesto di comprendere l'informatica e non solo di usare il calcolatore.

Per queste ragioni abbiamo deciso di descrivere un ampio spettro di applicazioni dell'informatica, dalle basi di dati all'information retrieval, dal Web al social networking, dai contenuti testuali a quelli multimediali, insistendo soprattutto sull'analisi critica dei fenomeni e cercando di fornire le conoscenze necessarie alla comprensione degli strumenti e delle tecnologie.

Il volume è idealmente diviso in tre parti: nella prima si affronta il tema, sempre più cruciale, della rappresentazione e memorizzazione dell'informazione, dalle basi di dati relazionali alle ontologie. Nella seconda sezione, si discutono i temi del reperimento e della pubblicazione dell'informazione in rete, con particolare riferimento al Web e alle sue applicazioni. Nella terza, infine, si presentano alcuni dei fenomeni più recenti di applicazione dell'informatica a temi e aree di tradizionale competenza umanistica, quali l'editoria digitale, il social networking e l'organizzazione dei contenuti multimediali.

*Silvana Castano, Alfio Ferrara, Stefano Montanelli*

*Dipartimento di Informatica e Comunicazione  
Università degli Studi di Milano*



Introduzione

## Il ruolo dell'umanista

L'informatica è la disciplina che si occupa del trattamento automatico dell'informazione e più in particolare della sua rappresentazione, conservazione e trasmissione. Se questa definizione coglie i tratti essenziali dell'informatica come scienza, indicandone l'oggetto e gli obiettivi, certamente è insufficiente a esprimere la portata dei fenomeni di informatizzazione della vita lavorativa e personale di ognuno di noi. La sua pervasività in tutti i campi fa infatti dell'informatica un sapere generale, necessario alla cultura personale di ognuno così come lo sono i fondamenti della conoscenza della lingua o dell'aritmetica. Questo tipo specifico di sapere si presenta frequentemente come una conoscenza di natura essenzialmente pratica, come un saper fare o, spesso, come un saper usare: usare strumenti come i *word processor* o il telefono cellulare. Ma, a fronte dell'esigenza di vivere e operare in un mondo informatizzato, bisogna domandarsi se questo sapere di ordine pratico sia sufficiente a garantirci una comprensione dei fenomeni. Per rispondere a questa domanda occorre guardare più da vicino l'informatica come scienza e riflettere sul ruolo che l'umanista è chiamato a svolgere in questo contesto.

## La natura e la storia dell'informatica

---

Il dibattito sullo statuto epistemologico dell'informatica come scienza e persino il tentativo di raccontarne una storia unitaria sono problemi del tutto aperti. La difficoltà nell'inquadrare l'informatica in modo unitario, sia dal punto di vista dei suoi metodi e fondamenti teorici sia dal punto di vista della sua storia, nasce dal fatto che la sua natura e la sua storia sono da declinarsi al plurale. Con questo intendiamo dire che l'informatica è soprattutto un territorio nuovo su cui si giocano diverse competenze esistenti.

La natura profondamente *interdisciplinare* dell'informatica è evidente sin dalle sue origini e continua a esserlo nell'evoluzione delle sue applicazioni. Anche la storia dell'informatica è, infatti, un insieme di storie diverse ma collegate. La storia dei metodi e delle teorie del calcolo, che nasce con la matematica antica e, passando per autori come Pascal o Leibniz, giunge allo sviluppo, a cavallo della seconda guerra mondiale, dei metodi di calcolo e automazione della moderna informatica a opera di autori come Alan Turing o John von Neumann, è collegata alla storia, non meno determinante, delle sco-

perte tecnologiche indispensabili all'informatica moderna, dal *transistor* ai moderni *personal computer*. Insieme a queste storie vi sono poi le vicende affascinanti delle conquiste applicative, soprattutto nel campo del *software*, che hanno portato alla creazione dei sistemi operativi, dei linguaggi di programmazione e delle vere e proprie applicazioni, dalla posta elettronica al Web. In tutte queste storie, le idee e le invenzioni sono il frutto del rapporto costante fra saperi diversi: vi sono le competenze matematiche e fisiche, ma anche, non meno importanti, i contributi delle scienze cognitive, della linguistica, delle discipline della comunicazione.

Se da un lato questa varietà, in una disciplina giovane e ancora poco definita, non consente di tracciare confini definiti e formulare una teoria unitaria sulla sua natura, dall'altro non può che indurci a vedere l'informatica come una grande opportunità di incontro e sviluppo delle diverse competenze.

La varietà e complessità dell'informatica ci pone di fronte anche a un'altra importante evidenza: l'informatica può essere ridotta al mero trattamento dell'informazione e del calcolo solo in senso strettamente tecnico ma, se guardiamo alla varietà dei prodotti e alla fecondità delle idee, ci rendiamo conto che ciò che è calcolo in teoria diviene *mondo* in pratica. Un moderno film di animazione o un videogioco, osservato in sé e tecnicamente è certamente solo calcolo e manipolazione dell'informazione ma, se osservato per sé e per i fruitori, costituisce un'esperienza cognitiva completa e più ampia del *codice* che lo genera. Il programma informatico, ridotto a algoritmi e dati, è di per sé un mondo e un'esperienza di realtà altra. La natura propriamente *virtuale* dei prodotti informatici, a cavallo fra il potenziale e il reale, li differenzia da ogni altro prodotto della tecnica e fa dell'informatica qualcosa di più di uno strumento. Nel motore a scoppio infatti o nella tecnologia della macchina per il caffè non vi è nulla di virtuale in questo senso, poiché queste tecnologie non mirano a tradurre in prodotto una visione del mondo e una teoria. L'informatica, al contrario, è teoria dell'informazione tradotta in oggetto del consumo e in metafora del mondo reale. Non è dunque possibile usare criticamente alcuno di questi oggetti prescindendo dalla conoscenza di tale teoria: non basta il saper fare ma occorre il *sapere* in sé cosa sia l'informatica e come operi sulle informazioni che comunichiamo e manipoliamo.

## Il digital divide

---

Con il termine *digital divide* si indica tradizionalmente un fenomeno sociale e culturale caratterizzato dal divario esistente fra chi ha accesso alle tecnologie dell'informazione e della comunicazione e chi ne è escluso. In genere, parliamo di *digital divide in grande* quando pensiamo all'esistenza di paesi tecnologicamente ed economicamente più avanzati di altri e *in piccolo* se ci riferiamo alle diverse condizioni economiche e culturali all'interno di una stessa popolazione. Il divario descritto dal *digital divide* è dovuto, in entrambi i casi, al fatto che le condizioni di competizione e di accesso ai beni, non ultimo il lavoro, nelle società tecnologicamente avanzate è condizionato dalla possibilità di accedere alle tecnologie. Insomma, la disponibilità della tecnologia è una condizione necessaria a non perdere il passo della società che ci circonda e a garantire un soddisfacente ingresso nel mondo del lavoro. Tuttavia, la semplice disponibilità materiale della tecnologia non è sufficiente. L'altro fattore determinante è infatti di ordine non materiale e con-

siste nella conoscenza degli strumenti tecnologici. La cosiddetta *alfabetizzazione* informatica si riferisce appunto all'esigenza di formare le persone all'uso degli strumenti tecnologici, come condizione fondamentale per ridurre il divario. Per usare una similitudine, potremmo dire che per spostarci con l'automobile non basta possederne una, ma occorre saperla guidare.

Gli strumenti informatici però, lo abbiamo ribadito in precedenza, non sono come le automobili, nel senso che non sono strumenti neutri rispetto al loro scopo, ma incidono in maniera determinante sui contenuti e sui processi comunicativi, cambiandone la natura e la cognizione. Possiamo dunque prendere a prestito la nozione di digital divide per introdurre un'ulteriore causa di divario relativa alle tecnologie: la distanza fra chi *saper usare* e chi *conosce*. Immaginiamo ad esempio di dover diffondere un messaggio sul Web e di avere per obiettivo l'efficacia della comunicazione e della sua diffusione. Certamente, per poter in primo luogo realizzare il nostro obiettivo occorre avere accesso alla pubblicazione sul Web e possedere le conoscenze pratiche necessarie a comporre una pagina di un sito. Questo *saper usare* ci distingue da chi non ha gli stessi strumenti e ci dà accesso alla rete e al suo pubblico. Ma tutto ciò è sufficiente perché il nostro messaggio competa in un contesto nel quale milioni di informazioni sono contemporaneamente presenti e fruibili? Pensiamo a come si accede alla rete: in primo luogo è ragionevole immaginare che i nostri destinatari arrivino al messaggio passando da qualche punto di accesso, come ad esempio un motore di ricerca o un blog. Arrivati al messaggio lo interpreteranno alla luce di processi cognitivi a cui sono abituati dalla consultazione costante di pagine web: giudicheranno il messaggio per come è scritto, per come appare sul nostro sito, per il suo corredo iconografico o ipertestuale. Adotteranno insomma un criterio interpretativo abituato alla fruizione di contenuti manipolati e trasformati dall'informatica. Perché dunque il nostro messaggio sia efficace dobbiamo garantirgli una buona visibilità sui motori di ricerca e comporlo alla luce delle trasformazioni rese possibili dalla tecnologia. Ottenere questi risultati richiede di sapere come funziona un motore di ricerca, come sono organizzati blog e comunità web, come il testo e le immagini vengono codificati, rappresentati e infine trasmessi. Occorre in definitiva conoscere e non solo saper fare. In altri termini è quanto mai vero il monito baconiano:

*La mano nuda e l'intelletto abbandonato a se stesso servono poco. Per compiere le opere sono necessari strumenti e mezzi d'aiuto, sia per la mano che per l'intelletto; e come gli strumenti meccanici servono ad ampliare o regolare i movimenti delle mani, così gli strumenti mentali estendono o trattengono il movimento dell'intelletto. La scienza e la potenza umana coincidono, perché l'ignoranza della causa preclude l'effetto, e alla natura si comanda solo ubbidendole: quello che nella teoria fa da causa nell'operazione pratica diviene regola. (Bacone, 1620)*

## Informatizzare gli umanisti

Alla luce di quanto detto possiamo ora comprendere meglio cosa significhi per un umanista accedere all'informatica. Per molto tempo, per gli umanisti insegnare e apprendere l'informatica ha significato soprattutto indicarne gli strumenti e suggerirne l'uso. In virtù di questi sforzi e per la diffusione rapidissima degli strumenti informatici nella vita quo-

tidiana, questo tipo di conoscenza è nel migliore dei casi già posseduta dalle giovani generazioni e inutile nel peggiore. In generale l'informatizzazione intesa come guida all'uso è insufficiente. Come abbiamo infatti già detto è necessaria anche una conoscenza della disciplina. Vi sono però due approcci alla conoscenza dell'informatica: da un lato la conoscenza dei metodi e degli strumenti informatici per come essi si applicano e agiscono nel contesto delle discipline umanistiche; dall'altro la conoscenza dell'informatica in sé, come problema culturale e teorico, alla luce però degli strumenti interpretativi e conoscitivi propri della tradizione delle discipline umanistiche. Possiamo dunque dire che c'è un'*informatica umanistica* che studia gli strumenti dell'informatica nella loro applicazione ai contenuti considerati tradizionalmente umanistici, dall'analisi del testo alla gestione dei contenuti editoriali ad esempio e c'è un'*informatica umanistica* che studia i problemi sollevati dall'informatica alla luce delle discussioni filosofiche e culturali precedentemente sviluppate dalle discipline umanistiche, dai problemi di logica e rappresentazione della conoscenza ai problemi della comunicazione. L'obiettivo dell'informatizzazione degli umanisti è dunque in primo luogo un problema culturale e scientifico. Esso ha però un impatto rilevante sulla collocazione professionale delle persone che hanno studiato e frequentato discipline tradizionalmente umanistiche. Da un lato, infatti, molte professioni tradizionali si avvalgono e richiedono oggi competenze di tipo informatico. Nel settore editoriale e della comunicazione, ad esempio, i contenuti sono sempre più spesso veicolati da strumenti informatici. Parallelamente, l'informatica ha dato origine a nuove professioni che richiedono conoscenze umanistiche affiancate da solide conoscenze informatiche. Un primo settore in cui sono fiorite nuove professioni è certamente il Web. Il numero di siti web cresce a ritmi elevati e, allo stesso tempo, la natura dei siti stessi cambia. In origine le competenze necessarie alla creazione di un sito erano quasi esclusivamente di tipo informatico, poiché molte pagine erano statiche e l'attenzione per i contenuti bassa; a un progettista serviva poco più che la conoscenza di HTML. Ovviamente, la trasformazione dei siti web in vere e proprie applicazioni software e il conseguente accrescimento della complessità tecnologica connessa alla pubblicazione sul Web produce una distinzione maggiore fra il ruolo degli informatici, degli amministratori dell'infrastruttura e dei programmatori delle applicazioni e di chi deve invece occuparsi di contenuti e lavorare alla progettazione dei siti. Questi ultimi devono possedere un'ampia conoscenza dei fenomeni legati al Web, dai blog al *social networking* e, al tempo stesso, devono possedere gli strumenti informatici necessari a comprendere come si possono organizzare i contenuti e diffonderli. Queste conoscenze non sono tradizionalmente competenze informatiche e coinvolgono in prima battuta gli strumenti linguistici e interpretativi di molte discipline umanistiche. Un altro settore di rilievo è legato al mondo della formazione. Gli strumenti informatici introducono e rendono possibile un approccio nuovo all'insegnamento, il cosiddetto *e-learning*, in cui non solo la formazione può svolgersi a distanza, ma le modalità di fruizione e composizione dei materiali di apprendimento è del tutto nuova. In questo ambito, dunque, vi è l'esigenza non solo di avere esperti della gestione dei contenuti o di comunicazione ma anche esperti nella formazione e nell'insegnamento che, dotati di competenze tecnologiche, individuino le potenzialità degli strumenti e accompagnino i discenti nella fruizione dei contenuti formativi. Infine, un ambito di recente sviluppo e grandi potenzialità riguarda la gestione della conoscenza o *knowledge management*. Si è scritto e detto più volte di come la conoscenza, intesa come patrimonio di dati organizzati e fruibili, sia una ricchezza di organizzazioni e aziende. In generale i

dati e metadati che un'organizzazione raccoglie nel tempo hanno un valore commerciale ed economico che è possibile sfruttare e ulteriormente valorizzare. Affinché però i dati e le informazioni acquistino un reale valore occorre che siano organizzati e gestiti in modo efficiente ed efficace. La gestione della conoscenza non deve occuparsi solo della progettazione di strumenti software efficienti; al contrario, lo sviluppo in ambito commerciale dei prodotti informatici rende sempre meno indispensabili queste competenze tipicamente tecnologiche all'interno di un'organizzazione che può reperire gli strumenti migliori direttamente sul mercato. Diviene quindi più strategico il ruolo di chi gestisce e organizza, per mezzo degli strumenti disponibili, il patrimonio conoscitivo. Quest'ultima attività richiede competenze logiche e conoscenze che sono a disposizione degli umanisti e che vanno applicate al contesto informatico.

Più in generale, possiamo concludere che gli umanisti sono chiamati a svolgere un ruolo di mediazione e di guida. Vi è una mediazione fra l'utenza e le squadre di sviluppo, composte da informatici, e una mediazione, sul piano dei contenuti, fra il messaggio e le modalità di fruizione. Gli umanisti sono sempre più spesso chiamati in ambiti diversi a guidare lo sviluppo informatico orientandolo verso l'obiettivo comunicativo che, in sé, è esterno all'attività specifica dell'informatica. D'altra parte, non tutto ciò che il committente o l'utenza richiede è tecnicamente realizzabile; in questo caso la mediazione viene svolta nella direzione opposta. L'umanista competente di informatica raccoglie i problemi e i limiti tecnici e li converte in direzioni di sviluppo e in confini entro cui sviluppare il progetto comunicativo. Accanto alle competenze tecniche, quindi, l'umanista necessita di una conoscenza generale e ampia di ciò che gli strumenti informatici hanno creato nei diversi ambiti, come il Web, l'editoria, la comunicazione sociale.

L'ultima parte del presente testo affronta dunque alcuni di questi temi ritenuti centrali per il rapporto fra discipline umanistiche e informatica, cercando di introdurre gli strumenti tecnologici accanto agli strumenti interpretativi e di analisi dei fenomeni informatici.

