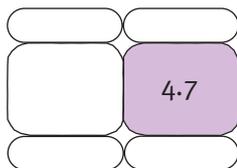




# INFORMATICA UMANISTICA: UNA DISCIPLINA DI CONFINE

Augusto Celentano  
Agostino Cortesi  
Paolo Mastandrea



L'informatica umanistica costituisce il punto di contatto tra scienze umane e scienze esatte: ragionando sui caratteri comuni delle diverse discipline umanistiche e formalizzando le procedure necessarie per condurre la ricerca nei diversi ambiti, propone l'integrazione dei due mondi superando la semplice applicazione di tecnologie avanzate a settori delle scienze umane. L'articolo presenta una rassegna delle molteplici applicazioni in cui tale integrazione è evidente.

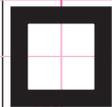
## 1. INTRODUZIONE

La Madre delle Muse, divine ispiratrici delle arti e protettrici del sapere, era chiamata *Mnemosyne* - che in greco significa Memoria; i miti indicavano così nella facoltà umana di accumulare nozioni entro la mente (altra parola corradicale e antichissima)<sup>1</sup> la base di ogni progresso ricreativo della scienza. Delineare un quadro delle novità sopravvenute negli ultimi anni agli studi umanistici grazie alla rivoluzione digitale, che tanto ha contribuito a incrementare le nostre capacità ritenitive, è lo scopo primo delle pagine che seguono; si vorrebbe tuttavia aggiungere almeno un elemento meno prevedibile, riguardo alla opportunità di ottenere vantaggi reciproci dalla collaborazione tra due macroaree culturali sinora lontane, anzi persino

ostili: non è, infatti, raro scorgere atteggiamenti di mutua diffidenza pregiudiziale, per cui gli informatici guardano con sprezzo le "fumoserie" di certo "integralismo umanista"<sup>2</sup>, venendo dalla controparte ricambiati con la simpatica accusa di fare null'altro che "tecnicismo strumentale". Per fortuna, ci pensa poi la realtà esterna a dirimere i nodi, è il mercato delle professioni - nei vari campi della informazione e della comunicazione - a richiedere il sapere degli umanisti. Fuori da ogni logica di steccato accademico, ma anche dall'etica della "paideia disinteressata" (bella ma difficile perché troppo élitaria), chi vuole produrre strumenti di tecnologia avanzata sa di poter trovare nel giovane laureato di estrazione umanistica un discreto esperto di semantica verbale, in grado di padroneg-

<sup>1</sup> Il suo etimo in forma latina era spiegato da Isidoro di Siviglia (*Origines seu etymologiae* 1, 1, 12) con un gioco di assonanze intraducibile: *mens vocata ... quod meminit* (vale a dire, "si chiama mente perché conserva i ricordi").

<sup>2</sup> Definizione tratta da *Contare e raccontare* (Bari, Laterza, 2003, p. 9) di Carlo Bernardini e Tullio De Mauro, una riproposta aggiornata e godibile del solito dibattito scienziati / umanisti.



giare almeno gli elementari sistemi di indicizzazione e classificazione. D'altro canto, entro il mondo della ricerca le figure degli studiosi stanno mutando in silenzio: basti pensare alla critica della letteratura, dove i testi sono ora sottoposti al controllo di strumenti statistici per costruire indici e rilevare concordanze, e la filologia digitale sta rapidamente assumendo un ruolo da protagonista nella ecdotica dei testi. In questi terreni si realizzano strane (perché quasi mai consapevoli) alleanze tra mondo universitario e mondo del lavoro: è giunta l'ora che tali contatti diventino meno sporadici, e ciò si ottiene costruendo *in primis* dei percorsi formativi finalizzati all'acquisizione della più completa competenza professionale nel campo del *Humanities Computing*.

## 2. INFORMATICA PER LE DISCIPLINE UMANISTICHE O CON LE DISCIPLINE UMANISTICHE?

La cosiddetta *Informatica Umanistica* - lo rivela il nome<sup>3</sup> - è disciplina bicefala, nata sulla frontiera del rapporto fra scienze umane e scienze esatte; eppure essa non è così recente in Italia, se il pioniere nell'uso del calcolatore applicato allo studio dei testi, padre Roberto Busa, iniziò le sue ricerche già nel 1949: la tecnologia informatica fu dunque da subito impiegata a risolvere anche problemi non numerici, per cui si può affermare che informatica e informatica umanistica siano pressappoco nate insieme.

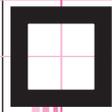
Quale l'identità di questa disciplina di confine? Se si guarda al passato appena trascorso (e la sezione che segue cercherà di delinearne i tratti principali), prevale la visione riduttiva di "Informatica per le discipline umanistiche"; ma ciò presuppone una informatica in vesti ancillari, che mette i propri strumenti al servizio delle scienze umane senza un ritorno in termini di reciproco accrescimento: non è

un caso che questa dicitura sia stata scelta dal MIUR, il Ministero dell'Istruzione, Università e Ricerca, come titolo dell'omologa classe di laurea specialistica. I limiti intrinseci a questo disegno sono essenzialmente due: in primo luogo, si considera dell'Informatica solo il lato applicativo-tecnologico, ignorando la ricchezza di modelli e linguaggi e metodologie che caratterizzano la "scienza dell'informazione" come area scientifica indipendente dalle matematiche e dalle ingegneria industriali; in secondo luogo, non si colgono le potenzialità di influsso delle discipline umanistiche nei confronti dell'Informatica, dimenticando che il padre dei linguaggi formali è un linguista, Noam Chomski; o che dai lavori di un logico, Alonzo Church, è nata la "teoria dei tipi" - una delle aree fondamentali dell'informatica teorica; o ancora che le logiche modali e temporali hanno aperto nuovi orizzonti ai modelli dei linguaggi di programmazione e alla elaborazione della conoscenza. Allargando lo sguardo da una considerazione iper-pragmatica dell'Informatica per le discipline umanistiche, dove ciascun campo di indagine si avvale di mezzi elettronici e metodologie informatiche note, *general purpose*, si avvertono segnali di novità, deboli ma inequivoci: indicano la maggiore ricchezza dell'approccio che vede l'Informatica con le discipline umanistiche delineare un terreno di utile contatto, dove è la conoscenza approfondita dell'oggetto di studio (l'edizione di un'opera di letteratura, l'analisi di dati storico-archivistici, la ricerca localizzata su un sito archeologico) a suggerire nuovi paradigmi di elaborazione adeguati a rappresentarne tutta la complessità, senza forzature e semplificazioni dovute ai limiti degli strumenti esistenti.

Ciò che contraddistingue una scienza umana da una scienza esatta è il dualismo intrinseco ai suoi oggetti di indagine: significanti verso significati per la linguistica, testi verso interpretazioni per la critica letteraria, documenti verso fatti storici per la storiografia, immagini verso rappresentazioni per la storia dell'arte, reperti verso ricostruzione delle condizioni materiali per l'archeologia, e così via. Questo confronto bipolare colloca, a prima vista, da una parte elementi certi e quantificabili, dall'altra mere interpretazioni che

<sup>3</sup> Il termine italiano è calco di espressioni quali *Humanities Computing*, *Computing for Humanities*, *Computers and the Humanities* e simili; si dirà invece "Informatica per le discipline umanistiche" (sigla IDU), se si vuole accogliere la titolazione ufficiale attribuita dal Ministero dell'Istruzione alla Classe 24/S delle Lauree specialistiche biennali.





scavo (attraverso il telerilevamento aereo e satellitare, le prospezioni geochimiche e geofisiche), dall'altro, chi conduce lo scavo bisogna che compia descrizioni dettagliate, mai prive di rigore scientifico, non solo dei reperti archeologici ma di tutti gli elementi che possano dare ragione del contesto ambientale del ritrovamento. Così, se lo scavo richiede, per sua stessa natura, una alterazione della situazione di partenza, vengono registrati una quantità di parametri che permettano di risalire con buona approssimazione (benché in misura non totale) a quella situazione di partenza. Si capisce, quindi, come si renda indispensabile ormai una gestione informatizzata dei dati di scavo, data la mole di dati da acquisire ed elaborare, in relazione a un sito archeologico ma soprattutto a più siti connessi fra loro<sup>5</sup>. Oltre a *database* di tipo tradizionale trovano largo impiego sistemi GIS (*Geographical Information System*) per la cartografia dei siti archeologici, che permettono di tracciare carte tematiche digitali e interrogabili. Utile per la ricerca, ma anche per la semplice divulgazione scientifica è l'impiego della modellazione 3D, tramite CAD o applicazioni per la navigazione in ambienti virtuali (VRML ecc.). Per finire, l'attenzione verso i metodi di ricerca quantitativi ha favorito lo sviluppo di *software* per analisi multivariate, per la seriazione, il *clustering* e l'analisi di corrispondenze applicato a dati archeologici<sup>6</sup>.

### 3.2. Biblioteconomia

La biblioteconomia è una delle discipline umanistiche che, grazie all'associazione con l'informatica, ha fornito i contributi più largamente noti al largo pubblico e condivisi soprattutto dall'intera comunità scientifica: la possibilità di condurre ricerche bibliogra-

fiche in rete presso la maggior parte delle biblioteche del mondo<sup>7</sup>. I cataloghi digitali hanno ormai raggiunto la maturità<sup>8</sup>, attraverso interfacce utente adeguate, l'allineamento a standard largamente diffusi e conosciuti (un esempio tra tutti: l'uso della codifica *Unicode* per rappresentare i caratteri di lingue diverse dall'anglo-americano), la possibilità di impostare filtri d'interrogazione anche molto articolati attraverso la combinazione di diverse voci delle schede bibliografiche: autore, titolo, data di edizione, soggetto ecc.. Se le grandi biblioteche hanno ammesso i navigatori di Internet agli archivi dei loro archivi bibliografici, non è ancora giunto il momento in cui si spalanchino le porte delle sale di lettura; solo allora, e sempre in ritardo per i tempi della nostra vita individuale, avverrà il salto di qualità, cioè un *web* che da semplice vetrina di osservazione divenga sistema di *literary machines* in un docuverso<sup>9</sup>, senza più bisogno di rinvii all'universo parallelo dei documenti cartacei. Certo l'opera di digitalizzazione dei materiali posseduti è terribilmente lunga e costosa da portare a termine, ma nel frattempo appare un po' paradossale che da almeno quindici anni le case editrici ricevano dagli autori i libri già in formato digitale, per poi distribuirli nella quasi totalità dei casi solo sotto forma cartacea. Un gradino comunque positivo per l'accesso ai testi, e non soltanto alle notizie bibliografiche, è rappresentato dai servizi di prestito interbibliotecario e dai servizi di *document delivery* (spedizione via posta ordinaria o *e-mail* di un esemplare del documento - articolo o parte di libro - in fotocopia o in formato digitale), da richiedersi contestualmente alla consultazione dei cataloghi digitali via web.

<sup>5</sup> Chi è interessato ad approfondire i temi connessi all'impiego di queste tecniche, può far riferimento alla rivista "Archeologia e Calcolatori", i cui indici sono consultabili al sito: <http://soi.cnr.it/~archcalc/>

<sup>6</sup> Per una bibliografia di base sull'argomento, si veda ad esempio: *Archaeological Statistics: A Brief Bibliography*, <http://archaeology.about.com/blstats.htm>

<sup>7</sup> Un riferimento sintetico ma completo offrono F. Metitieri e R. Ridi, *Biblioteche in Rete - Istruzioni per l'uso*, Bari, Laterza, 2002.

<sup>8</sup> Uno dei cataloghi più consultati è certamente quello della Library of Congress: <http://catalog.loc.gov/>

<sup>9</sup> I termini vengono da Ted H. Nelson, *Literary Machines*, Sausalito (Ca), Mindful Press, 1987; vi era previsto un sistema per il pagamento dei diritti d'autore, ciò che rendeva il progetto Xanadu di Nelson simile più a una libreria ("usa e getta?") che a una biblioteca pubblica di dimensioni planetarie.

### 3.3. Filologia

L'informatica testuale si può suddividere in documentaria, editoriale ed ermeneutica<sup>10</sup>. La prima si occupa della creazione di banche dati e sistemi di interrogazione, la seconda della rappresentazione dei testi a stampa, a video o su altri *media*, la terza, invece, si dedica alla relazione fra il testo e le sue interpretazioni. La filologia computazionale<sup>11</sup> è la disciplina che coordina questi tre aspetti, integrando in una visione coerente il processo di archiviazione di testi e varianti, la marcatura delle informazioni (lessicali, grammaticali, semantiche ecc.) inerenti ai testi archiviati, producendo, infine, edizioni critiche a stampa o a video. Benché il termine "ipertesto" sia nell'accezione comune associato a una rete di documenti in relazione fra di loro ma privi di prospettiva storica, occorre dire che la filologia tradizionale si è servita sin dalle origini della lettura non lineare propria dell'ipertestualità. In primo luogo, il confronto fra i diversi manoscritti (collazione) richiede il passaggio continuo e sistematico da un documento all'altro, per individuare le varianti; in secondo luogo, l'interpretazione di passi oscuri del testo è illuminata dalla ricerca di luoghi paralleli, nella stessa opera o in altre; in terzo luogo, la consultazione di commentari, lessici, repertori iconografici ecc., è per sua stessa natura una forma di lettura non lineare. Per ciascuna di queste tre forme di ipertestualità si sono sviluppati degli strumenti di ausilio all'attività del filologo, che vengono velocemente passati in rassegna in un riquadro separato.

### 3.4. Musicologia

È questo un altro campo in cui le applicazioni hanno avuto una larghissima diffusione in Italia. Oltre a varie iniziative per mettere in rete banche dati più o meno aperte e partecipate (incipitari di madrigali, cantate o altro; schedatura interrogabile di libretti d'opera; lessici specifici ecc.), esistono

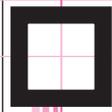
programmi di videoscrittura analoghi a quelli per caratteri alfabetici, dove cliccando sul valore della nota (minima, semiminima, croma ecc.) e portandola sul rigo che interessa, se ne può ascoltare il suono, quindi verificare immediatamente nell'eventualità gli errori che appaiono come "stecche"; gli stessi programmi impaginano e mettono in partitura i brani automaticamente, emettono le diverse parti (primi violini, secondi, fiati ecc.) e applicano il testo alle parti vocali altrettanto automaticamente. In laboratori specializzati (come il centro di sonologia computazionale della Facoltà di Ingegneria di Padova) sono state sintetizzate opere intere, ed è allo studio un algoritmo che riproduce suoni e timbri, con una tecnica diversa da quella del campionatore che ovviamente prevede suoni emessi da strumenti tradizionali, mentre questo sistema li crea da sé. Per far sembrare più vera l'esecuzione, si sta studiando come ottenere esitazioni e imprecisioni simili a quelle umane.

### 3.5. Linguistica

La Linguistica è stata una delle prime discipline umanistiche a comprendere l'importanza del connubio con l'informatica e a figliare una disciplina autonoma, la Linguistica Computazionale. Essa è inoltre riuscita a far crescere i propri progetti, nati in ambito accademico, e trovar loro lavoro nel commercio e nell'industria, diventando prodotti di largo uso: dai correttori ortografici e sintattici, ai sistemi di riconoscimento vocale, ai programmi di redazione di riassunti automatici. L'approccio tradizionale dell'informatica al testo consiste nel trattamento di stringhe di caratteri all'interno dei quali individuare particolari *pattern* che permettano, ad esempio, la suddivisione in parole, il raggruppamento delle forme, l'ordinamento alfabetico ecc.. Compito della Linguistica Computazionale è appunto quello di elaborare modelli in grado

<sup>10</sup> La comoda tripartizione è stata proposta anche di recente da Padre Roberto Busa: si veda per esempio A. Massarenti, *L'umanesimo viaggia sul computer*, ne "Il Sole - 24 Ore" del 7 agosto 2002, oppure in <http://193.205.145.117/docenti/informatica/appello/massarenti.htm>

<sup>11</sup> Risorse online sull'argomento si trovano nel sito: <http://www.griseldaonline.it/informatica/>. Un buon manuale introduttivo è R. Mordenti, *Informatica e critica dei testi*, Roma, Bulzoni, 2001.



di andare oltre questa forma elementare di trattamento e di individuare, in modo automatico (allo stato attuale delle ricerche si dovrebbe dire: semiautomatico), le relazioni linguistiche presenti<sup>12</sup>; procedere in particolare alla lemmatizzazione delle forme (per esempio, ricondurre la forma verbale italiana “lesse” al paradigma di “leggere”); individuare le strutture sintattiche (cioè le funzioni di soggetto, oggetto, predicato ecc. con i relativi elementi subordinati, nonché le funzioni semantiche di agente, tema affetto, strumento ecc.); attribuire tratti semantici pertinenti (associare a “sentire” e “udire” il tratto di percezione acustica); individuare le espressioni relative a un medesimo referente (“Roma” e “la capitale” in un contesto che lasci sottintendere: “capitale d’Italia”); e si potrebbe continuare. Se nei confronti di testi fortemente codificati, com’è il caso di opere tecniche, scientifiche o giuridiche, i risultati sono già apprezzabili (si pensi ai sistemi di traduzione automatica attivati presso i centri amministrativi della comunità europea), nei confronti di testi dal contenuto letterario più libero la ricerca è tuttora aperta<sup>13</sup>. Segni confortanti di sviluppo stanno venendo, tuttavia, dall’interazione della linguistica computazionale con le scienze cognitive<sup>14</sup>. Un ambito complementare (e per certi versi preparatorio) rispetto a quello del riconoscimento automatico di strutture linguistiche è il campo della marcatura semiautomatica delle strutture morfosintattiche e lo sviluppo di sistemi di interrogazione delle strutture che rappresentano l’interpretazione sintattica precedentemente marcata.

### 3.6. Studi letterari

Gli studi letterari hanno tratto ampio giovamento dalla costituzione di grandi *corpora*

di testi antichi e moderni in CD-Rom (si veda il riquadro separato per un’analisi dei più importanti strumenti e progetti sull’argomento). Generalmente, queste iniziative di digitalizzazione dei testi, condotte sotto la responsabilità di centri universitari, accademie e istituti pubblici o privati, fondazioni o case editrici, sono accompagnate dalla creazione di sistemi di *information retrieval* e alcune volte di analisi testuale metrica, grammaticale ecc.) adeguati alla base di dati stessa. Oltre all’indubbio vantaggio di riunire i materiali di interesse biblioteche, questo tipo di strumenti e, naturalmente, i sistemi di *information retrieval* a essi associati, hanno permesso una vera e propria variazione di scala rispetto alle domande che si pone il letterato. Se gli studi filologici tradizionali favorivano l’indagine approfondita fino al minimo dettaglio del senso di un singolo verso, di una giuntura o anche di una singola parola, questi strumenti danno la possibilità di attraversare l’intero corpo estendendo a dismisura la portata dei nostri interrogativi. Qualunque indagine intertestuale su qualunque campo letterario, espresso in qualsiasi lingua antica o moderna, quella che un secolo fa si chiamava “ricerca delle fonti” (*Quellenforschungen*), oggi si effettua con una rapidità e una certezza inimmaginabili mettendo il critico in grado di affrontare problemi di interpretazione su repertori di materiali che è stata la “macchina” a predisporli sul tavolo. Certo, è nota l’obiezione secondo cui dove si guadagna in estensione, si rischia di perdere in profondità, ma lo smarrimento del dettaglio è comunque ricompensato dall’ampiezza dell’angolo visuale. Come si è detto, spesso le iniziative di digitalizzazione dei testi si fanno promotrici anche dello sviluppo di

<sup>12</sup> Si possono trovare risorse relative a questo settore sul sito del Laboratorio di linguistica computazionale di Venezia: <http://project.cgm.unive.it/>

<sup>13</sup> Basta sottoporre una metafora del tipo “quella donna è una quercia” per far collassare qualunque buon sistema di analisi semantica automatica - a meno che il particolare uso non sia stato preventivamente codificato fra le accezioni del termine “quercia”.

<sup>14</sup> I problemi connessi con la categorizzazione, a prescindere dalla formalizzazione informatica, sono affrontati in modo particolarmente brillante da G. Lakoff, *Women, Fire and Dangerous Things*, Chicago, University Press, 1987. Per una introduzione ai problemi del rapporto fra semantica e informatica, si può far riferimento a F. Rastier, M. Cavazza, A. Abeillé, *Sémantique pour l’analyse. De la linguistique à l’informatique*, Paris, Masson, 1994.

software adeguato a interrogare le basi di dati testuali prodotte. Tuttavia, ci si è presto resi conto che gli strumenti creati *ad hoc* per interrogare una specifica base di dati (per esempio, il *corpus* della letteratura italiana) spesso non soddisfano tutte le esigenze degli studiosi: per esempio possono esistere ordinamenti alfabetici ma non cronologici, possono mancare filtri in base alla tipologia metrica delle opere in poesia ecc.. Tanto le iniziative *no profit* quanto le iniziative commerciali più illuminate (il *Thesaurus Linguae Graecae* in testa) hanno ben compreso la necessità di rendere pubbliche le specifiche di codifica dei testi che costituiscono le loro basi di dati, sia quando siano dei formati proprietari, sia quando rispondano a degli standard ampiamente condivisi. Questo permette, in modo molto opportuno, di separare la base di dati dal sistema di interrogazione che insistono su di essa. Proprio a questo obiettivo tende la *Text Encoding Initiative* (TEI)<sup>15</sup>, che parte dallo studio di numerose tipologie testuali (testi in poesia o in prosa, opere drammatiche, scritti scientifici, edizioni critiche ecc.) per proporre dei modelli di marcatura XML del testo e delle informazioni a esso correlate.

### 3.7. Studi storici

L'influsso dell'informatica sulla ricerca storica è stato più volte oggetto di riflessioni di varia natura e orientamento. Esiste una *International Association for History and Computing*, ramificata in varie associazioni nazionali, che pubblica regolarmente una rivista (*History & Computing*) e ogni anno organizza convegni internazionali, dove trovano spazio anche dibattiti di natura metodologica.

Non di rado le riflessioni degli studiosi italiani<sup>16</sup> – tanto più quando lavorano negli archivi – si concentrano sul problema delle fonti, vale a dire sul tema della rappresentazione e registrazione di informazioni trat-

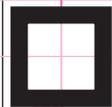
te da documenti storici. Nel periodo tra la fine degli Ottanta e i primi anni Novanta, l'adeguatezza del software cosiddetto commerciale a gestire i dati di natura storica ha offerto il campo per vivaci scontri di posizioni, usati talora come pretesto da quanti – e non eran pochi – si schieravano, di fatto, contro qualsivoglia uso dell'informatica nelle ricerche storiche. In quegli anni, un illustre ma poco illuminato critico dell'Illuminismo poteva ancora sollevare scandalo per il ricorso alla posta elettronica come forma degenerata di comunicazione tra uomini di scienza.

Essendo gli storici persone più di altre sensibili alla questione delle fonti, non stupisce che le basi di dati siano per lungo tempo apparse loro come punto nodale delle applicazioni dell'informatica alla ricerca. In realtà, l'informatica entra oramai in ogni fase del laboratorio storico, anche se il suo impatto non sempre è percepito con chiarezza; da una parte, ciò si deve, in generale, al carattere pervasivo della strumentazione nella vita quotidiana (basti pensare ai mutamenti prodotti dal passaggio ai *word processor* nelle forme della scrittura, anche storiografica), d'altro lato molte potenzialità offerte dai calcolatori sono affatto sconosciute al di fuori di cerchie relativamente esigue di specialisti. Ma per lo più l'elenco delle "facilitazioni" offerte agli storici è abbastanza ovvio, e oltre alla disponibilità di banche dati *on-line* include come dappertutto le liste di discussione specialistica, quindi la forma più immediata ed economica di comunicazione coi propri colleghi in ogni parte del mondo, la possibilità di diffondere risultati anche parziali del proprio lavoro.

Anche in questo ambito è arrivato il momento di chiedersi se l'informatica abbia solo reso più facile il lavoro, o ne abbia mutato la qualità e la natura, o se ciò possa avvenire in futuro. Da un lato, la maggiore disponibilità di risorse comporta senza dubbio un au-

<sup>15</sup> TEI: <http://www.tei-c.org>

<sup>16</sup> Tra i contributi più recenti e approfonditi si segnala il volume di S. Vitali, *Passato digitale. Le fonti dello storico nell'era del computer*, Milano, Bruno Mondadori, 2004; ancora utile la miscellanea *Storia & Computer. Alla ricerca del passato con l'informatica*, a cura di S. Soldani e L. Tomassini (pubblicato dallo stesso editore nel 1996).



mento degli standard qualitativi: molte cose che prima erano impossibili, o troppo dispendiose o troppo faticose, sono oggi a portata di mano, a prezzo scontato. D'altra parte, si riesce a intravedere l'inizio di una lenta ritirata della componente erudita nella ricerca storica, a vantaggio di una accentuazione degli aspetti teorici e concettuali. La comodità di accesso agli archivi elettronici spinge verso una separazione tra le fasi della raccolta dei dati e del loro impiego e, in prospettiva, un distacco tra le persone che si occupano dei processi; non è detto che chi ha costituito una base di dati sia la medesima persona che ne farà uso: sono le stesse incredibili opportunità offerte dalla rete di collaborare strettamente e quotidianamente con studiosi remoti che favoriscono l'abbandono di tradizionali forme di minuta erudizione "territoriale", agevolano l'apertura a temi di vasto respiro e di interesse comparativistico.

L'evoluzione è più evidente nei settori della storia economica, sociale, demografica, e in tutti i campi dove l'analisi statistica riveste un ruolo fondamentale. I metodi di analisi hanno conosciuto qui progressi significativi, aprendo il campo ad analisi sofisticate e innovative, impensabili fino a pochi anni fa; diviene più facile sottoporre a verifica empirica teorie e modelli sviluppati da discipline attigue, appunto come l'economia, la sociologia, la demografia, l'epidemiologia: dalla descrizione dei fenomeni socioeconomici si passa alla loro interpretazione su solide basi statistiche, dove la specificità dei casi locali perde rilievo rispetto all'interesse del complessivo quadro concettuale, proiettando l'analisi su piani comparativi: dove non basta ovviamente disporre di archivi e risorse di calcolo adeguate, ma sempre serviranno competenze multidisciplinari approfondite.

#### **4. L'INFORMATICA UMANISTICA NELL'OFFERTA FORMATIVA UNIVERSITARIA**

Questa natura dell'Informatica Umanistica come disciplina di confine mostra le sue contraddittorietà, e manifesta le diverse visioni che su di essa le comunità accademici

che hanno maturato, quando si osservano le diverse offerte formative che negli ultimi anni alcuni Atenei italiani hanno incluso nei loro manifesti degli studi. Secondo le opinioni di molti (forse la maggioranza della comunità scientifica del settore: se ne fanno i più vivaci "messaggeri" alcuni allievi romani del compianto Giuseppe Gigliozzi), occorre favorire la creazione di nuovi insegnamenti universitari specifici di Informatica Umanistica, i cui docenti abbiano competenze specifiche nelle discipline umanistiche e generiche in informatica, in modo tale da procedere all'alfabetizzazione, all'insegnamento della storia della disciplina, dei suoi aspetti teorici e metodologici, delle basi della programmazione. A Pisa e a Firenze si sono, invece, aperti cicli, rispettivamente triennali e biennali (cioè di primo o secondo livello, in osservanza ai nuovi assetti organizzativi degli atenei) di un corso di studi che preveda l'approfondimento di specifiche discipline umanistiche (soprattutto letterarie e linguistiche) accompagnate dalla frequenza di un consistente montatore di informatica. L'esperienza veneziana, infine, destina all'Informatica Umanistica un corso di secondo livello universitario (biennio della laurea specialistica) in base ai seguenti principi:

1. accesso da parte di laureati di primo livello che abbiano acquisito competenze specifiche in una delle diverse discipline umanistiche (storia, letteratura, lingua straniera, conservazione dei beni culturali, filosofia) o, al contrario, in informatica;
2. approfondimento di discipline metodologiche quali la filosofia della scienza, la logica, la linguistica, che consentono il ragionamento critico sui metodi peculiari delle scienze umane e delle scienze esatte;
3. frequenza obbligatoria di corsi di Informatica presso la Facoltà di Scienze, per chi proviene da trienni umanistici, e all'opposto di corsi di materie fondamentali umanistiche presso le Facoltà di Lingue e di Lettere per chi proviene dal triennio di Informatica;
4. tesi di laurea sistematicamente concepite come lavoro di sintesi, di assimilazione di metodi pertinenti alle scienze esatte e alle scienze umane.

#### 4.1. L'esperienza dell'Università Ca' Foscari di Venezia

Si spenderà qualche parola in più sul corso di laurea specialistica aperto a Ca' Foscari <<http://lettere2.unive.it/infouman/>> per sottolineare la novità di approccio pienamente paritetico, sin dai primi passi di ideazione e progettazione, tra diverse matrici culturali. È interessante osservare che questa iniziativa coinvolge docenti di area umanistica e informatica, con una presenza paritaria che trova riscontro sia nella composizione del collegio didattico che nello *status* di corso interfacoltà a ogni effetto. Il corso tende, per quanto possibile, a integrare le doti di perizia comunicativa e persuasiva, sensibilità linguistica e letteraria, gusto estetico e artistico, insomma i caratteri culturali propri di chi sceglie gli studi umanistici, con le competenze tecniche e il costume scientifico acquisibili nell'ambiente dell'informatica; lo scopo è di creare figure professionali desiderate non solo dal mondo della produzione e del commercio, cioè dal mercato del lavoro a breve e a medio termine, ma anche dalle istituzioni dove si fa ricerca avanzata, in campo nazionale e internazionale. Chi ha pensato e disegnato la struttura del corso si pone l'obiettivo di preparare soggetti capaci di accoppiare nella stessa persona competenze ormai dovunque indispensabili (ma oggi ripartite tra almeno due differenti figure professionali), in possesso di sicura autonomia operativa e decisionale, quindi presto riconvertibile in presenza di continue innovazioni delle tecnologie e trasformazioni del mercato. La composizione del piano di studi prevede, fra gli altri: insegnamenti di informatica (programmazione, *web design*, linguaggi per la rete come HTML e XML, gestione di basi di dati); insegnamenti di teoria della scienza (epistemologia, logica, filosofia del linguaggio, storia della scienza); insegnamenti di linguistica (fonologia e fonetica, linguistica computazionale e linguistica informatica); insegnamenti di letteratura (antica e moderna) e filologia (classica, romanza, italiana); insegnamenti metodologici in settori umanistici specifici (ricerca archeologica, teoria della letteratura, metodi quantitativi per le scienze sociali); nozioni di economia e di diritto (geografia economica, diritto commerciale); corsi di mi-

crolingua mirati al lessico informatico e ai prodotti multimediali. I tirocini avvengono, ove possibile, presso aziende pubbliche e private che operano nell'ambito della multimedialità, della creazione di software di supporto alle attività culturali, di traduzione automatica e semiautomatica; poi, nel campo dell'editoria e del giornalismo, e presso le istituzioni culturali quali archivi e biblioteche che facciano ampio uso delle nuove tecnologie. L'elaborato finale, al quale lo studente dedica l'ultimo semestre del biennio, consiste nella stesura di una dissertazione originale, di natura teorica o applicativa.

#### 5. LE DISCIPLINE UMANISTICHE PER L'INFORMATICA?

Si conclude questo intervento con una provocazione, rivolta in particolare a chi fa ricerca in Informatica; l'invito è ad aprire gli orizzonti dei propri interessi culturali a mondi "affini", pensando che la miscela di saperi diversi, di linguaggi e metodologie lontane, non può che arricchire ed essere fonte di nuova creatività. Ecco a puro titolo di esempio alcune aree di ricerca che potrebbero beneficiare di questa "contaminazione umanistica".

##### 5.1. Linguaggi di Programmazione

La torre di Babele dei linguaggi di programmazione si arricchisce di anno in anno di nuove proposte, di nuovi dialetti. Ma qualora si volesse procedere a un confronto sistematico di tali proposte, ovvero a una catalogazione sistematica dei linguaggi in base a criteri quali semplicità, espressività, complessità, lessico, perché non guardare a quanto è stato prodotto dai colleghi linguisti nella catalogazione dei linguaggi naturali, e nello studio della loro evoluzione storica?

##### 5.2. Human-Computer Interaction

La progettazione di interfacce utenti, la valutazione di usabilità dei sistemi, la realizzazione di sistemi di *help*, già guardano con interesse a quanto è stato prodotto in Psicologia Cognitiva. Così come nella progettazione di applicazioni didattiche non si può più ignorare quanto la Psicologia dell'età evolutiva ha elaborato nell'ultimo decennio relativamente alle modalità (e ai limiti) dell'apprendimento.

### 5.3. Ingegneria del Software

È riconosciuto da tutti che la qualità di un “progetto” informatico non si riduce alle caratteristiche del solo codice ma dipende fortemente da fattori, quali la qualità dell’analisi, la documentazione del codice sorgente, la leggibilità della manualistica: tutti elementi che possono beneficiare degli interessanti studi sull’italiano per la scrittura tecnica sviluppati recentemente in alcuni dipartimenti di italianistica.

### 5.4. Rappresentazione della conoscenza

Nella formulazione di nuovi modelli di calcolo nella comunità informatica, l’interazione con le Scienze biologiche e naturali si è rivelata quanto mai efficace. Ma il coinvolgimento di studiosi con *background* filosofico potrebbe aiutare ad andare oltre alle difficoltà tecniche insite in questi approcci per ipotizzare scenari più completi e complessi di rappresentazione ed elaborazione della conoscenza.

Un grande maestro, il massimo studioso dei veicoli cui gli uomini affidano da duemilacinquecento anni la trasmissione della loro memoria alle generazioni successive, ha scritto di recente che “nella cultura scritta informatica i modi e i processi della composizione, registrazione, trasmissione e conservazione dei testi sono automaticamente provocati e determinati dalla legge del profitto dell’industria multimediale [...]. I processi di trasmissione dello scritto stanno passando in mano a mediatori naturalmente estranei ai prodotti di cui si occupano, di fronte ai quali essi sono tecnicamente analfabeti. [...]. Sembra proprio che si sia giunti molto vicini alla rottura di una catena di trasmissione testuale durata, sia pure con cadute e faticose riprese, alcuni millenni”<sup>17</sup>. Si ritiene, al contrario, che esista un antidoto alle idee pessimistiche che motivano tali profezie, e che questo stia nella nostra abilità di condurre una gestione accorta ma altruistica, lungimirante, generosa delle risorse informatiche e telematiche, così da rendere l’innovazione scientifica e la consapevolezza umanistica “compagne di strada” in questo viaggio, avendo entrambe come fine ultimo l’inte-

resse a realizzare le più felici aspirazioni di ogni uomo e dell’umanità intera.

### Bibliografia

- [1] Aloni G.: *Scienze dell’antichità per via informatica. Banche dati, internet e risorse elettroniche nello studio dell’antichità classica*. Bologna, Clueb, 2002.
- [2] Ciotti F., Roncaglia G.: *Il mondo digitale*. Bari, Laterza, 2000.
- [3] Cristofori A., Salvaterra C., Schmitzer U. (cur.): *La rete di Arachne. Nuove tecnologie, didattica e antichità classiche*. Stuttgart, Steiner 2000.
- [4] Fiorimonte D.: *Scrittura e filologia nell’era digitale*. Torino, Bollati-Boringhieri, 2003.
- [5] Gigliozzi G.: *Introduzione all’uso del computer negli studi letterari*, a cura di F. Ciotti, Milano, Bruno Mondadori, 2003.
- [6] Griseldaonline: <http://www.griseldaonline.it/informatica/> (a cura di Francesca Tomasi).
- [7] Landow G. P.: *L’ipertesto. Tecnologie digitali e critica letteraria*, a cura di P. Ferri, Milano, Bruno Mondadori, 1998.
- [8] Merzweb: <http://www.merzweb.com/> (a cura di Gino Roncaglia).
- [9] Mordenti R.: *Informatica e critica dei testi*. Roma, Bulzoni, 2001.
- [10] Numerico T., Vespignani A.: *Informatica per le scienze umanistiche*. Bologna, Il Mulino, 2003.
- [11] Retimedievali: <http://www.retimedievali.it/> (a cura di Andrea Zorzi).

AUGUSTO CELENTANO è professore ordinario di Sistemi per l’elaborazione delle informazioni all’Università Ca’ Foscari di Venezia. Laureato in Ingegneria Elettronica al Politecnico di Milano nel 1973, ha svolto attività di ricerca nelle aree dei linguaggi di programmazione, dell’ingegneria del software, dei sistemi informativi e dell’e-learning, e si interessa attualmente di sistemi multimediali e di human-computer interaction. Insieme agli altri due autori, è tra i componenti del primo collegio didattico del corso di Laurea specialistica in Informatica per le discipline umanistiche dell’Università Ca’ Foscari. [auce@dsi.unive.it](mailto:auce@dsi.unive.it)

AGOSTINO CORTESI è professore straordinario di Informatica all’Università Ca’ Foscari di Venezia. Laureato in Matematica all’Università di Padova nel 1986, ha conseguito il dottorato di ricerca in Matematica Computazionale ed Informatica Matematica nel 1991. Svolge attività di ricerca nell’area dei Linguaggi di Programmazione, in particolare su tecniche di interpretazione astratta per l’analisi statica e la verifica di programmi. [cortesi@dsi.unive.it](mailto:cortesi@dsi.unive.it)

<sup>17</sup> A. Petrucci, *Prima lezione di paleografia*, Bari, Laterza, 2002, p. 125.

PAOLO MASTANDREA è ordinario di Lingua e Letteratura Latina a Venezia; cura l'organizzazione di *corpora* letterari in formato elettronico, interrogabili in rete ([www.unive.it/poetitalia](http://www.unive.it/poetitalia)) e su disco (ALB, Poesis2, PoetriaNova); si occupa di storia della tradizione e critica dei testi antichi mediante analisi 'intertestuale'

assistita dal computer (saggio di applicazione: *La continuità del sistema letterario: mezzi di ricerca e fini didattici nello studio dei classici. Un'indagine su Dante Tasso e Lucrezio, condotta grazie a strumenti elettronici*, Genova, Latina Didaxis 17, 2002).  
mast@unive.it

## Strumenti informatici per le discipline umanistiche

### FILOGIA

Per la collazione automatica e la redazione di apparati critici, si segnala fra tutti il TUSTEP creato presso l'università di Tübingen <<http://www.uni-tuebingen.de/zdv/zrlinfo/tustep-des.html>> e adottato da prestigiose case editrici per la ricchezza delle sue funzioni.

Il progetto Kleio, diretto da Manfred Thaller del Max Planck Institute for History di Goettingen <<http://wwwuser.gwdg.de/~mthaller2/>>, gestisce invece un sistema di mappatura fra il testo digitalizzato e le immagini digitali dei manoscritti.

Fin dall'Ottocento, strumenti tradizionali per la ricerca di luoghi paralleli erano *Indices e Concordantiae* a stampa, che permettevano di individuare le occorrenze di una parola, in ordine alfabetico. Il progetto di redarre gli indici dell'*Opera omnia* di Tommaso d'Aquino, iniziato da Roberto Busa nel 1949 e conclusosi negli anni Novanta, segna l'atto di nascita dell'informatica testuale. Nel corso dei decenni, le iniziative di questo tipo si sono moltiplicate, fornendo strumenti capaci di redigere indici e concordanze dinamici, consultabili anche via web, su *corpora* molto vasti di autori.

Alla terza tipologia di strumenti, utili non solo al word-retrieval e ricerca di luoghi paralleli, ma anche finalizzati ad agevolare la consultazione della letteratura secondaria (commentari storici, linguistici, prosopografici ecc.), è dedicato tra gli altri il Perseus Project <<http://www.perseus.tufts.edu/>>, dove si sviluppa l'idea dell'ipertesto come enciclopedia specialistica.

### Studi letterari

Fra i più autorevoli archivi elettronici di testi letterari in CD-ROM si possono menzionare: il *Thesaurus linguae Graecae* (TLG), l'archivio del *Packard Humanities Institute* (PHI), il *Cetedoc Library of Christian Latin Texts*, la *Bibliotheca Teubneriana* e i *Monumenta Germaniae Historica* on CD-Rom, il *Patrologia Latina Data Base*, ALB e Poesis di Zanichelli. A questi si affiancano collezioni più o meno ampie disponibili in rete, fra cui ad esempio: la sezione dei classici greci e latini del Perseus Project, l'Oxford Text Archive, The Latin Library, il progetto Gutenberg, il progetto Manuzio. Tenendo conto dell'assenza di apparato critico nella quasi totalità delle edizioni elettroniche, i criteri per valutare un archivio di testi digitalizzati sono principalmente: la completezza del corpus, la scelta delle edizioni a stampa di riferimento, la conformità alle edizioni a stampa scelte. Anche rispetto ai migliori repertori, rimane sempre aperta l'esigenza di aggiungere, aggiornare e correggere i testi.

Lo studio dei testi letterari richiede la possibilità di gestire informazioni riguardanti la disposizione e la visualizzazione del testo (suddivisione in capitoli, numerazione di versi; caratteri greci e latini, corsivi, segni diacritici uniformi ecc.) e di individuare strutture appartenenti a diversi livelli di analisi: metrica, morfologica, sintattica, retorica ecc.. Per introdurre nei documenti elettronici questa mole di informazioni aggiuntive, sono stati elaborati degli standard di marcatura, in particolare dalla *Text Encoding Initiative* (TEI) e dal *World Wide Web Consortium* (W3C). Il Perseus Project si distingue fra le iniziative volte a offrire, oltre a un numero considerevole di testi classici consultabili in rete, anche la possibilità di effettuare l'analisi morfologica di ogni singola parola, di visualizzarne la traduzione in inglese, di individuare il numero di occorrenze della parola nell'intero corpus ecc..

La codifica dei testi secondo uno standard (SGML/XML) o secondo un protocollo proprietario messo a disposizione di terze parti (per esempio, il protocollo di codifica del PHI) assicura l'indipendenza dell'archivio digitale dai programmi di interrogazione. Fra i programmi di *text retrieval general purpose* si può menzionare TreSy (*Text Retrieval System for SGML/XML*), implementato dal CRI-BeCu (Centro Ricerche Informatiche per i Beni Culturali), il quale è in grado di estrarre informazioni, a partire da un'interrogazione strutturata, da qualunque documento che rispetti le specifiche di marcatura SGML/XML.

Per il TLG ed il PHI si possono, invece, menzionare le applicazioni di text retrieval Musaios, Lector, Pandora o Workplace Pack. Si distingue dai precedenti SNS Greek & Latin, implementato dal Laboratorio Informatico per le Lingue Antiche della Scuola Normale Superiore di Pisa, che, grazie alla costituzione di nuovi indici, alle funzionalità aggiuntive e all'ottimizzazione del motore di ricerca, offre prestazioni molto superiori ai precedenti prodotti. I dischi Poesis2 e PoetriaNova permettono la ricerca di parole sulla produzione poetica in latino, dalle origini alla metà del XIII secolo.

Fra i word processor studiati per la redazione di edizioni critiche si segnalano Classical Text Editor, Collate2 e Critical Edition Typesetter, recensiti diffusamente all'indirizzo: <<http://www.economia.unibo.it/dipartim/stoant/rassegna1/software.html>>

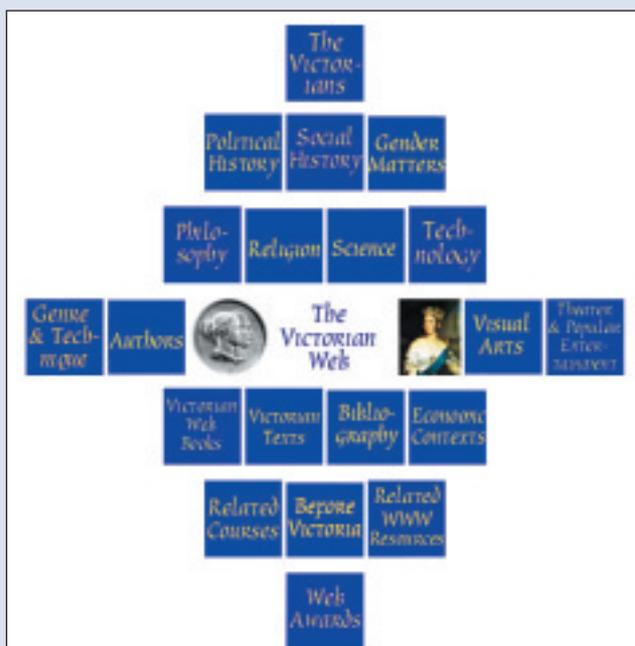
## Apologia della biblioteca digitale

Per caso ci siamo trovati a vivere in un'epoca di rapidi progressi, di straordinari ricambi nelle forme di trasmissione del nostro patrimonio culturale; le circostanze possono sembrare analoghe a quelle che Santo Mazzarino (*La fine del mondo antico*, 1959) scorgeva nell'età tardo-imperiale romana: come allora, è in atto ora una specie di "democratizzazione della cultura", portata dalle dinamiche sociali dell'ultimo mezzo secolo; tali processi si pagano con inesorabili abbassamenti del livello medio, con una produzione editoriale scolastica che sempre più semplifica e compendia, sforna grammatiche elementari, lessici manuali, breviari storici, compendii scientifici: una realtà che può magari non piacere, ma cui vanno contrapposte mosse prudenti, per uscirne senza danni e con dignità. È ovvio che non sempre, e non da tutti, il trasferimento da carta stampata a supporto elettronico sia visto sotto buona luce: e qualche rischio davvero c'è, ora come in ogni fase di passaggio innovativo, anche se la cosiddetta "morte del libro" come veicolo privilegiato di comunicazione delle conoscenze resta una ipotesi lontana e improbabile. Più serio, ma anche circoscritto, sembra il pericolo che venga a formarsi una sorta di "nuova vulgata" diffusa dalle raccolte di opere letterarie affidate a supporto informatico o telematico; ostacolare la montante tendenza alla rinuncia dell'esercizio critico (non limitata all'ambito filologico) è però un dovere morale e intellettuale che in nulla ostacola i vantaggi enormi offerti dalla disponibilità di archivi di testi elettronici collegati a programmi di ricerca verbale. Vantaggi per chi fa indagini scientifiche, e può rimpiazzare quasi completamente la strumentazione lessicologica a stampa con mezzi rapidi, flessibili, sicuri. Vantaggi per chi insegna, se si pensa alla libertà offerta ai docenti di ogni grado scolastico nelle scelte antologiche o nelle letture di brani per esercizi di traduzione. Vantaggi per chi impara, o almeno per i più curiosi tra gli studenti: cui è aperta la strada a un repertorio letterario assai più vasto e diversificato rispetto a quello proposto di solito, ben al di sopra di qualsiasi canone degli *auctores*.

Come ha mostrato una volta per tutte Ernst Robert Curtius (*Europäische Literatur und lateinisches Mittelalter*, 1948), la cultura letteraria comune dell'Occidente presenta una matrice in gran parte greca e latina, ma non ispirata ai soli autori ed opere dell'antichità 'classica'; anche a questo scopo, per aiutarci a considerare ogni oggetto nella giusta dimensione del sistema, che dall'Iliade e dall'Odissea e da Ennio *alter Homerus* ha perduto fino all'età moderna senza subire grosse interruzioni, vanno sfruttati a fondo gli archivi di testi in veste digitale, cui collegare programmi di ricerca di parole; possono venire risultati inattesi dai nuovi modi di fruizione di tesori immensi, ora aperti a tutti, accessibili per vie che nemmeno avrebbero potuto concepirsi senza la disponibilità e la facilità d'uso dei mezzi elettronici.

## George P. Landow e gli ipertesti sulla letteratura inglese e americana

Tra gli argomenti che descrivono i contatti tra l'informatica e le scienze umane non è stato trattato il vasto mondo dell'e-learning, l'applicazione di tecnologie informatiche all'insegnamento, per due semplici ragioni: la prima, di sostanza, è che l'e-learning attraverso potenzialmente tutte le discipline e non solo quelle umanistiche; la seconda, pratica, è che per dire qualche cosa di significativo servirebbe uno spazio pari almeno a quello occupato da questo articolo. Non si può però tralasciare di citare il contributo dato da George P. Landow (Professor of English and Art History alla Brown University di Providence / RI) all'insegnamento della letteratura con strumenti informatici. I suoi materiali didattici centrati su complesse reti di documenti ipertestuali, realizzate con sistemi *ad hoc* alla fine degli anni '80 e con siti Web recentemente, hanno dato un grande contributo allo sviluppo delle tecnologie e metodologie ipertestuali, svelandone le potenzialità come strumento di comunicazione, ma evidenziando anche la necessità di un lungo e profondo lavoro di redazione, di selezione, di trasformazione dei materiali didattici, per raggiungere gli obiettivi di qualità necessari. Il lettore interessato può iniziare la sua navigazione dal sito <http://www.landow.com/>, che funge da indice per i principali siti Web realizzati da Landow. La figura mostra la mappa di accesso al primo livello di argomenti trattati nel Victorian Web, la cui qualità è dimostrata da un numero elevatissimo di premi e menzioni conferiti nel corso degli anni.



La mappa degli argomenti del Victorian Web di George P. Landow