

A PROPOSITO DI ALCUNE ESPERIENZE RECENTI DI UMANISTICA DIGITALE

Roberto Palaia

Riassunto

Nel testo è sinteticamente ricostruita l'affermazione dell'Umanistica Digitale: dalle applicazioni ai vari copora testuali di grandi autori per la costituzioni di vocabolari tematici o linguistici, fino alle applicazioni dedicate a settori e ambiti tematici riferibili a tutti i campi delle scienze umane e sociali. Vengono analizzati brevemente i tratti comuni delle metodologie applicate, evidenziati i problemi e le soluzioni proposte, fino a giungere a una illustrazione delle più recenti applicazioni implementate nella rete. Nelle note infine è indicata una prima bibliografia di orientamento utile per eventuali approfondimenti.

Parole chiave: Lessicografia, Memorizzazione, Fonti.

Keywords: Digital Humanities, Lexicography, Big Data, Sources.

Negli ultimi decenni il campo applicativo dell'Umanistica Digitale si è enormemente ampliato, rivoluzionando l'approccio alle forme e ai contenuti dello studio delle discipline umane e sociali (e non solo di quelle ovviamente). All'inizio l'Umanistica Digitale era finalizzata soprattutto allo studio intensivo dei testi e alla loro analisi e codifica che poteva essere svolta, per la prima volta, non solo analizzando i contenuti, ma anche approfondendo gli aspetti formali del testo. Centrali in questo contesto restano le attività svolte, a partire dalla fine degli anni '40, da padre Roberto Busa che pensò di utilizzare i primi grandi calcolatori della IBM per memorizzare e lemmatizzare le opere di San Tommaso. Fu così che nacque l'*Index Thomisticus*, opera che segnò l'introduzione del calcolatore nello studio dei testi letterari e che rappresenta l'inizio delle applicazioni dell'informatica ai testi di cultura¹. Il lavoro svolto sull'opera di S. Tommaso ha rappresentato a lungo un modello di trattamento dei testi da tener presente per impostare un lavoro analitico avvalendosi dell'ausilio delle metodologie informatiche. Sono così seguite le implementazioni di moltissimi testi filosofici e letterari con grandi progetti di catalogazione e lemmatizzazione di termini con la costruzione di grandi tesauri e dizionari, come sono stati per esempio il *Trésor de la langue française*, il *Thesaurus Linguae Latinae*, il *Tesoro della Lingua Italiana delle Origini* oppure il progettato *Lessico Filologico dei Secoli XVII e XVIII*².

La svolta con l'avvento del world wide web

Questa attività ha caratterizzato la prima fase di

digitalizzazione per così dire "interna" dei testi di cultura che ha successivamente conosciuto una svolta grazie all'avvento del *world wide web* agli inizi degli anni novanta, alla disponibilità tecnologica di memorie di massa di enormi capacità a prezzi sempre più bassi³. Questa nuova situazione sembrò potesse realizzare il sogno a lungo inseguito e mai raggiunto di rendere possibile l'accesso ai documenti grazie a linguaggi descrittivi e di interrogazione omogenei.

La possibilità di identificare in modo univoco ogni documento presente nella rete e l'affermazione di standard che permettessero lo scambio delle informazioni ha spostato il focus dagli oggetti digitalizzati sul web allo studio delle loro relazioni e alla possibilità, attraverso tale studio, di ottenere contenuti evoluti. Sembrò quindi si realizzasse l'utopia definita fin dagli inizi dell'età moderna di giungere a realizzare un linguaggio universale per gli oggetti e un calcolo sicuro per analizzarne le relazioni, in modo certo e senza errori. Rendere possibile così «instituire un inventario delle conoscenze umane, nel quale sia riportato tutto quanto vi è di più utile, più certo, più universale e maggiormente adeguato a determinare tutto il rimanente... [affinché] si possano sciogliere i più utili problemi rimasti insoluti, ma tuttavia risolvibili, al fine di migliorare la vita degli uomini e accrescere quanto è possibile, e quanto prima, la nostra felicità»⁴.

Questo processo ha determinato una riproposizione nettissima di antichi problemi metodologici insiti nel dualismo proprio degli oggetti della ricerca umanistica; per la necessaria digitalizzazione delle fonti ci si è dovuti confrontare con le relazioni fra significato e significato sul piano dei linguaggi, fra testo e sua

interpretazione sul piano letterario, così come fra documenti e loro rappresentazione storica nella memorizzazione delle fonti, e così via in tutte le discipline filosofiche, archeologiche, artistiche (con i problemi inerenti la memorizzazione delle immagini), musicali, ciascuna con specificità e problemi propri. In generale l'Umanistica Digitale ha contribuito a drammatizzare il dualismo fra oggetto dell'analisi (testo letterario, immagine artistica, ecc.) e sua esegesi, allargando il campo problematico del tema dell'interpretazione al quale si è cercato di dare risposta ricercando nel vasto spazio del *world wide web* un linguaggio descrittivo condiviso e standardizzato⁵.

Ovviamente questa duplicità fra contenuti virtuali e dati reali ha finito per riproporre tutta una gamma problematica fra segno e significato che nel campo delle discipline umanistiche ha una lunga storia, ripresentandosi però in contesti e con strumentazioni completamente nuove rispetto al passato. Uno dei settori nei quali questi aspetti hanno avuto un nuovo significato molto tangibile è stato quello dell'analisi delle grandi masse di informazioni, rappresentazioni esse stesse dei comportamenti degli utenti, ma anche in grado, a loro volta, di condizionarli profondamente.

È molto difficile fornire oggi un quadro che pur vagamente riassume le attività riferite all'Umanistica Digitale, che ormai pervade tutti i campi della ricerca in campo umanistico e che ha cambiato sostanzialmente metodi e pratiche del lavoro scientifico. In termini molto generali è forse possibile individuare alcuni campi di intervento, tenendo presente però che ciascuno di essi è un universo con problematiche specifiche intorno alle quali esiste un dibattito anche teorico di notevole rilevanza⁶.

Un primo campo può essere individuato nelle attività relative alla produzione dei contenuti degli oggetti della ricerca: l'enorme capacità di memorizzazione e la virtuale mancanza di limiti quantitativi nella memorizzazione dei documenti pone alla filologia nuove possibilità e inediti problemi. Nella raccolta delle opere degli autori come si definisce oggi sulle piattaforme informatiche il corpus dei testi considerati? Oggi che diventa possibile documentare per ogni testo considerato tutti i passaggi dal manoscritto fino alla stampa definitiva, il compito della nuova filologia informatica si è trasformato al fine di offrire un oggetto di studio nel quale il testo tradizionale come siamo stati fino ad oggi abituati a considerare (così come l'immagine nel caso della ricerca iconografica, oppure il brano audio in quella musicale e così via) rappresentano soltanto uno dei momenti che daranno luogo alla formazione del documento nella sua completezza⁷.

Un secondo campo di ricerca dell'Umanistica Di-

gitale può essere individuata nella rappresentazione dei contenuti e nelle modalità che presiedono alla loro archiviazione. Uno dei problemi più rilevanti per l'interoperabilità dei dati è stato quello di definire un linguaggio descrittivo dei documenti memorizzati. A partire dagli anni '90 il consorzio TEI (*Text encoding initiative*) ha proposto un linguaggio di marcatura dei testi (XML) per la loro digitalizzazione e per la costruzione di grandi archivi e di banche dati. Correlate a queste procedure è il discorso sull'*open access* e sulla possibilità di scambio di dati che rappresenta un presupposto fondamentale per la costruzione di ambienti digitali aperti e non finalizzati allo sfruttamento commerciale⁸.

Un ulteriore campo di indagine può infine essere individuato nello studio della metodologia dell'organizzazione delle informazioni, anche attraverso la realizzazione di piattaforme specifiche e nella ricerca di ulteriori possibilità applicative delle esperienze realizzate in altri campi. Tale tematica ha assunto progressivamente una sempre maggiore importanza perché la strutturazione delle informazioni determina in gran parte anche la loro accessibilità. Lo studio delle metodologie di funzionamento dei motori di ricerca e le implicazioni, derivanti dalla loro capacità di fornire risultati attendibili e neutri da un punto di vista socio-politico, sono uno degli aspetti più difficili da gestire correttamente come testimoniato da molti avvenimenti recenti.

Tutto considerato comunque la rete e le implementazioni che la riguardano dimostrano comunque delle attività intense e degne di nota⁹. Tra gli esempi recenti di originali implementazioni di piattaforme informatiche complesse si possono osservare ad esempio *Clavius on the Web* (<http://claviusontheweb.it/>) e *Pelagios Commons* (<http://pelagios.org/>). Il primo, che vede coinvolti l'Archivio Storico della Pontificia Università Gregoriana, in collaborazione con l'IIT-CNR e l'ILC-CNR, ha reso possibile la costruzione di una piattaforma dedicata all'epistolario del gesuita Christophorus Clavius. I testi delle lettere sono stati trascritti e codificati (in XML-TEI), per poi essere sottoposti ad un trattamento di lemmatizzazione. Ciascuna lettera è stata annotata, sia linguisticamente che semanticamente, sulla base di un lessico e di un'ontologia di dominio. Le cui entità sono connesse a database semantici. La piattaforma è costruita per permettere un'ampia interoperabilità con gli studenti delle scuole superiori e rappresenta un importante e originale esempio di incontro tra ricerca, umanistica digitale, didattica e processi di apprendimento. Tra i vari aspetti di innovazione e di interesse ci sono gli strumenti e le strategie di visualizzazione interattiva dei documenti, che fanno della piattaforma un ottimo

esempio di integrazione tra *Public History*, *Cultural Heritage* e *Digital Humanities*.

Pelagios Commons, un'iniziativa supportata e finanziata dalla Andrew W. Mellon Foundation, ha come proprio obiettivo lo sviluppo di strumenti e servizi finalizzati all'interconnessione della enorme mole di risorse storiche presenti nel Web. L'annotazione dei luoghi di interesse storico citati nelle risorse digitali stesse e il loro inserimento nella rete dei *Linked Open Data* è stata realizzata partendo dalla convinzione che questo approccio alla geografia antica possa offrire agli studiosi, ma anche al pubblico interessato, nuovi metodi di studio e visualizzazione dei dati annotati. Tra le ultime attività sviluppate, sono stati presentati strumenti che permettono l'integrazione delle annotazioni geografiche con quelle relative a cose e a persone nonché la realizzazione di grafi per la visualizzazione dei dati.

Altri recenti progetti dedicati alla ricerca umanistica sono *DanteSources* (<http://perunaenciclopedia-dantescadigitale.eu>), che nasce dalla collaborazione tra l'ISTI-CNR e il Dipartimento di Filologia, Letteratura e Linguistica dell'Università di Pisa. Oltre a consentire il libero accesso ad alcune opere dantesche (al momento *Convivio*, *De vulgari eloquentia*, *Monarchia* e *Vita Nova*), l'applicazione implementata permette di visualizzare, attraverso grafi e tabelle, i testi utilizzati da Dante nel processo di scrittura delle sue opere e di condurre ricerche in relazione ai luoghi testuali in cui egli cita una specifica opera o un autore, o nei quali viene discusso un tema rilevante. Uno strumento importante e utile, di sicuro interesse per la comunità scientifica, ma anche utilizzabile nei contesti di formazione scolastica.

Granai della Memoria (<http://www.granaidellamemoria.it/>), è invece promosso e realizzato dall'Università degli Studi di Scienze Gastronomiche e da Slow Food, con la collaborazione dell'Università Suor Orsola Benincasa di Napoli, dell'Università degli Studi del Piemonte Orientale e della Fondazione Ignazio Buttitta di Palermo. Obiettivo di questo lavoro è la salvaguardia delle tradizioni, delle esperienze e delle storie di vita che rischiano di scomparire per la sempre maggiore incuria e mancanza di riflessione sul passato. Per contrastare quella che definiscono una vera e propria 'emergenza antropologica', i ricercatori coinvolti si sono posti l'obiettivo di raccogliere, conservare e rendere accessibili le testimonianze, registrate in Italia e all'estero, di contadini, artigiani, operai, partigiani, cioè di persone che probabilmente non avrebbero mai avuto spazio né voce. L'archivio multimediale costituisce dunque una preziosa raccolta di memorie, per l'approfondimento di studi storico-antropologici e per la preservazione del patrimonio culturale orale.

Già solo considerando il solo panorama italiano, le specifiche attività e, più in generale, le linee di ricerca collocabili all'interno del variegato universo dell'umanistica digitale, appaiono vivaci ed estremamente feconde. I processi di standardizzazione delle pratiche e delle metodologie per la costruzione degli archivi, per la concettualizzazione e organizzazione dei dati raccolti e per la visualizzazione dei contenuti modellizzati attraverso i nuovi approcci alla *knowledge organization*, sono in costante incremento e rappresentano un oggetto di un variegato dibattito.

Note

¹ *Index Thomisticus: Sancti Thomae Aquinatis operum omnium indices et concordantiae in quibus verborum omnium et singulorum formae et lemmata cum suis frequentis et contextibus-variiis modis referuntur quaeque / auspice Paulo 6. Summo Pontifice, consociata plurium opera atque electronico IBM automato usus digessit Robertus Busa*, Stuttgart, Frommann Holzboog, 1974-1980, 56 voll.

² Il *Trésor de la langue française* è un dizionario di lingua francese del XIX e XX secolo, in 16 volumi pubblicato tra il 1971 e il 1994; Il *Thesaurus Linguae Latinae* (TLL) è un monumentale dizionario di latino. Comprende la lingua latina dalle sue origini fino alla latinità altomedievale; Il progetto ha avuto inizio nel 1894, la pubblicazione del primo volume nel 1900 e la fine è in programma per il 2050 circa; l'impresa è finanziata dalla *Bayerische Akademie der Wissenschaften*. Il *Tesoro della Lingua Italiana delle Origini* (TLIO) è la parte antica del *Vocabolario Storico della Lingua Italiana* elaborato presso l'Istituto del CNR OVI (Opera del Vocabolario Italiano). Il *Lessico filosofico dei secoli XVII e XVIII* è stato un progetto parzialmente realizzato presso il Centro di studio per il Lessico Intellettuale Europeo del CNR (poi confluito nell'ILIESI CNR) pubblicato presso l'editore Olschki di Firenze tra il 1993 e il 1999.

³ Per un'approfondita analisi dei vari aspetti legati all'Umanistica Digitale si rimanda a: *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, J. Unsworth, Oxford, Blackwell, 2004, <http://www.digitalhumanities.org/companion/>.

⁴ G. W. Leibniz, *Inizi ed esempi della Scienza Generale*, in *Scritti Filosofici*, Torino, UTET, 2000, vol. 1, p. 229.

⁵ A. Celentano, A. Cortesi, P. Mastandrea, *Informatica Umanistica: Una disciplina di confine*, «Mondo digitale», 4, 2004, pp. 44-55.

⁶ D. Fiormonte, T. Numerico, F. Tomasi, *The Digital Humanist. A Critical Inquiry*, Brooklyn NY, Punctum Books, 2015, p. 95 e sgg.

⁷ Un interessante approccio al tema della filologia digitale è alla base del lavoro svolto dall'*Institut des Textes et manuscrits modernes* del CNRS di Parigi; per approfondire i temi appena accennati nel testo si rinvia al sito: <http://www.item.ens.fr/>.

⁸ Sugli standard e le procedure di codificazione dei testi (Text Encoding Initiative) si rimanda alla documentazione contenuta nel sito: <http://www.tei-c.org/index.xml>.

⁹ Per uno sguardo sinottico sulle recenti attività in Italia rimando alla informata nota di M. Tardella, *Digital Humanities e beni culturali: quale relazione*, in «Lexicon Philosophicum. International Journal for the History of Texts and Ideas», 4, 2016, pp. 227-231: <http://lexicon.cnr.it>.

ROBERTO PALAIA

Storico della filosofia è primo ricercatore presso l'ILIESI-CNR.

Dopo essersi laureato presso l'università 'Sapienza' ha frequentato vari corsi di perfezionamento in Storia della Filosofia e in Informatica per le scienze umanistiche in Italia e in Germania. I suoi interessi sono orientati soprattutto allo studio del pensiero filosofico e scientifico europeo dei secoli XVII e XVIII e in particolare alla figura di G. W. Leibniz, al quale ha dedicato numerose pubblicazioni. Inoltre ha focalizzato i suoi studi sulla formazione e evoluzione dei linguaggi tecnici della filosofia nell'età moderna, approfondendo

soprattutto il tema della formazione della lingua dell'Aufklärung. Ha partecipato e diretto numerosi progetti nazionali e internazionali dedicati a temi di Cultural Heritage e Digital Humanities. È stato docente presso l'Università de L'Aquila, membro di varie società filosofiche nazionali e internazionali e socio fondatore della Sodalitas Leibnitiana.

Contatti: ILIESI-CNR - Via Carlo Fea, 2 - 00161 Roma
E-mail: roberto.palaia@cnr.it