

# Dall'Informatica umanistica alle culture digitali

Atti del convegno di studi (Roma, 27-28 ottobre 2011)

in memoria di Giuseppe Gigliozi

a cura di

Fabio Ciotti e Gianfranco Crupi



SAPIENZA  
UNIVERSITÀ EDITRICE

digilab  
ricerca e servizi

DOI: 10.7357/DigiLab-32



# Che cos'è, oggi, l'informatica umanistica?

## L'impatto della tecnologia

Dino Buzzetti, Università di Bologna

1. Rispondere al quesito che ci è stato proposto, non può non tenere conto degli effetti che lo sviluppo delle tecnologie informatiche ha necessariamente esercitato sulle pratiche di ricerca che, da qualche decennio, hanno tratto ispirazione da alcune precise motivazioni teoriche. A proposito di tali presupposti teorici, una definizione dell'informatica umanistica, che può essere utilmente richiamata, si trova formulata, nella pubblicazione conclusiva di un *Thematic Network* europeo<sup>1</sup> dedicato allo *Advanced Computing in the Humanities* (ACO\*HUM) e precisamente in un capitolo che tratta espressamente dell'applicazione di «metodi formali» alle discipline

---

<sup>1</sup> L'Agenzia Esecutiva per l'Istruzione, gli Audiovisivi e la Cultura (EACEA), istituita dalla Commissione delle Comunità Europee, così definisce gli obiettivi di un *Thematic Network*: «Lo scopo principale dei *Thematic Networks* è quello di elevare la qualità [...] in una determinata disciplina accademica o campo di studi», EACEA, *Erasmus Thematic Networks (Socrates): Objectives*, [http://eacea.ec.europa.eu/static/en/overview/ThemNetwks\\_overview.htm](http://eacea.ec.europa.eu/static/en/overview/ThemNetwks_overview.htm) (consultato il 23 marzo 2012).

umanistiche.

Nel testo, redatto sotto la direzione di Tito Orlandi, si legge:

[...] cercheremo di definire le proprietà essenziali di ogni tipo di informatica applicata sulla base del tradizionale binomio *strutture di dati* e *algoritmi*, riferendolo alle caratteristiche particolari di ciascuna disciplina: a) i metodi che servono a rappresentare l'informazione, in uno specifico campo di conoscenze, in modo tale da renderne possibile l'elaborazione mediante sistemi computazionali, producono le *strutture di dati* richieste da una certa disciplina; b) i metodi che servono per formulare i problemi e i procedimenti di ricerca, in uno specifico campo di conoscenze, in modo tale da permettere l'applicazione di procedure di elaborazione computazionali, producono gli *algoritmi* applicabili a una data disciplina.<sup>2</sup>

Se dunque l'informatica consiste nel rappresentare l'informazione con determinate strutture di dati ed elaborare l'informazione mediante algoritmi, quella particolare specie di *informatica applicata* in cui consiste l'*informatica umanistica* può essere considerata come informatica applicata alle discipline umanistiche e consisterà nei diversi modi di rappresentare l'informazione in forma idonea alle caratteristiche proprie delle varie discipline e nell'elaborare l'informazione con procedure idonee alle domande specifiche che la ricerca si pone nei diversi campi disciplinari.

---

<sup>2</sup> K. de Smedt et al. (ed.), *Computing in Humanities Education: A European Perspective*, Bergen, University of Bergen – HIT Centre, 1999, ch. 2, European studies on formal methods in the humanities, § 2.3, [M. Thaller,] *Defining humanities computing methodology*, <http://gandalf.aksis.uib.no/AcoHum/fm/fm-chapter-final.html> (consultato il 23 marzo 2012).

2. Assumendo questi presupposti teorici come una caratterizzazione effettiva dell'informatica umanistica, così come si è venuta sviluppando fin dal primo apparire dei suoi concreti progetti di ricerca, ci si può allora chiedere se la pratica comune delle applicazioni dell'informatica alle discipline umanistiche, alla quale abbiamo assistito, si possa riconoscere in questa sua definizione, nelle diverse fasi di sviluppo delle tecnologie informatiche, che di volta in volta si sono rese disponibili. In questa prospettiva si possono distinguere tre momenti principali dello sviluppo tecnologico, tali da influenzare in forma diversa l'adozione di metodi computazionali nelle discipline umanistiche.

Già al tempo dei grandi calcolatori, o *mainframe*, si è assistito al varo di grandi progetti di ricerca, soprattutto nel campo della linguistica computazionale, e alla creazione di importanti archivi linguistici e testuali. L'onerosità e la scarsa accessibilità dei dispositivi e delle attrezzature necessarie alla ricerca rendeva possibile solamente l'avvio di grandi progetti sostenuti istituzionalmente da rilevanti risorse finanziarie e condotti a termine con la partecipazione da ampi gruppi di studiosi qualificati. Si sono così costituiti i grandi centri di ricerca e le grandi banche dati per l'elaborazione di risorse soprattutto testuali.

Una fase nuova si è aperta con l'introduzione dei piccoli calcolatori, i cosiddetti *mini* calcolatori o *personal computer*.

L'accessibilità agli strumenti informatici è stata così offerta potenzialmente a tutti i ricercatori e si è assistito ad una grande proliferazione di iniziative individuali, o promosse da piccoli gruppi di ricerca. L'introduzione del *personal computer* ha anche comportato l'adozione di forme standard di codificazione primaria dell'informazione e l'interscambiabilità dei dati e dei programmi. Capacità di elaborazione potenziate potevano essere realizzate con l'impiego di

calcolatori di dimensioni intermedie, o *workstation*, che restavano però escluse dalle prime reti telematiche che collegavano, con protocolli ancora proprietari, solo i grandi calcolatori.

Un ulteriore incremento dell'accessibilità dei dati e della diffusione delle iniziative di ricerca si è prodotto con l'introduzione del *WorldWideWeb*. Con la sua rapida e immediata espansione, la pervasività del mezzo informatico per la comunicazione dei dati e delle informazioni ha praticamente coinvolto, a livelli diversi, ogni forma di attività e di produzione culturale. L'accessibilità e la diffusione di contenuti multimediali, alla cui prima apparizione si era già assistito nella fase precedente, ha reso possibile l'estensione dell'applicazione di procedure computazionali a materiali non esclusivamente testuali anche nel campo della ricerca umanistica.

3. Il succedersi di queste diverse fasi nella disponibilità degli strumenti offerti dalla tecnologia non è stato però privo di conseguenze sulla pratica effettiva dell'applicazione di metodi informatici in campo umanistico. Possiamo cercare di illustrarne la portata con un esempio tratto dal campo della filologia e della critica testuale. A chi avesse cominciato ad occuparsi di edizioni digitali fin dalla fine degli anni ottanta, si offriva allora la possibilità di associare le trascrizioni dei testi alle immagini delle fonti manoscritte solo attraverso la disponibilità di potenti *workstation* in grado di offrire risorse di calcolo sufficienti all'elaborazione di immagini digitali ad alta risoluzione e tali da permettere una riproduzione adeguata dei documenti originali. Ma nei sistemi a quel tempo disponibili<sup>3</sup>, la possibilità di associare singole parti della trascrizione alle porzioni

---

<sup>3</sup> Cfr. M. Thaller, "Kleio: Ein Datenbanksystem", St. Katharinen, Max-Planck-Institut für Geschichte. In *Kommission bei Scripta Mercaturae*, Verlag, 1989.

corrispondenti delle immagini dei manoscritti da cui erano ricavate, presupponeva l'organizzazione del testo in elementi di informazione strutturata e l'integrazione di dati visuali in unico sistema di gestione dell'informazione, o *DataBase Management System* (DBMS).

L'edizione era concepita come un database<sup>4</sup> e, ciò che più importa ai fini della nostra trattazione, la trascrizione era concepita non più come un mezzo per riprodurre il documento originale, alla stregua di una tradizionale trascrizione diplomatica, utile soprattutto al fine di trasmettere al lettore «un'idea più fedele della natura della fonte» manoscritta,<sup>5</sup> ma come un dato fornito al computer «per essere ulteriormente elaborato»; così, in questa fase, la trascrizione

diventa un'attività di modellizzazione e di codifica dei dati, per estrarre dal manoscritto quanta più informazione possibile e per ricavarne analiticamente nuovi risultati. Da questo punto di vista, sia l'immagine che la trascrizione non sono più considerate come riproduzioni *fisiche* utili a ripresentare il documento originale, ma piuttosto come dati analitici miranti ad una nuova rappresentazione *logica* della fonte.

Mi sono deliberatamente riferito ad uno scritto di quegli anni, per mostrare come già nella fase d'uso del personal computer e fin nei primi tempi dell'espansione del web, la preoccupazione

---

<sup>4</sup> D. Buzzetti, P. Pari, A. Tabarroni, "Libri e maestri a Bologna nel XIV secolo: Un'edizione come database". In *Schede umanistiche*. n.s., 6, 1992(2), p. 163-169 e D. Buzzetti, "Masters and Books in 14th-century Bologna: An edition as a database". In F. Bocchi, P. Denley (eds.), *Storia & Multimedia, Proceedings of the Seventh International Congress of the Association for History and Computing*. Bologna, 29 August-2 September 1992, Bologna, Grafis Edizioni, 1994, p. 642-646.

<sup>5</sup> D. Buzzetti, "Image Processing and the Study of Manuscript Textual Traditions". In *Historical Methods*. 28, 1995(3), p. 145.

metodologica fosse ancora legata al lascito dell'esperienza maturata nel periodo precedente, contraddistinto dall'uso dei mainframe. L'insistenza sulla produzione di dati strutturati e la contrapposizione tra l'uso *logico* e l'uso *analogico* della rappresentazione digitale mettono in luce la permanenza di un'attenzione rivolta prevalentemente all'elaborazione dei dati e ad una concezione della rappresentazione dell'informazione funzionale soprattutto all'applicazione di procedure computazionali miranti all'elaborazione del contenuto informativo della fonte.

4. Mette conto insistere su questo aspetto perché fa risaltare, per converso, una tendenza prodottasi paradossalmente nel passaggio dai mainframe ai personal computer, per giungere infine all'attuale predominanza dei servizi web. La paradossalità del fenomeno consiste nello spostamento dell'enfasi dalla funzione che abbiamo definito "logica" alla funzione che abbiamo contraddistinto, metaforicamente, come "analogica" della rappresentazione digitale, che non è più vista, prevalentemente, come funzionale all'elaborazione del contenuto informativo del documento o dell'oggetto fisico rappresentato, ma che viene sempre più considerata come un mezzo per la riproduzione virtuale delle loro caratteristiche fisiche e materiali.

Nel campo dell'informatica umanistica questa tendenza al distacco dall'elaborazione del contenuto si è manifestata fin dalla fase di passaggio all'uso dei personal computer. Con l'affermarsi dei word processor e dei sistemi WYSIWYG (*What You See Is What You Get*) per la produzione dei documenti, nel campo stesso dell'infor-

matica, alla cosiddetta *data processing* o *database community* si è affiancata con sempre maggior peso la cosiddetta *document community*<sup>6</sup>.

Mentre l'interesse prevalente della *document community* mirava sostanzialmente a preservare l'invarianza del documento – e ciò valeva in particolare per i documenti di studio, oggetto della ricerca umanistica – l'interesse della *database community* mirava invece a preservare l'invarianza della semantica di dati. Così, mentre la *database community* «sceglieva di standardizzare la semantica» e quindi si preoccupava principalmente dell'elaborazione del *contenuto* dei dati, senza prestare particolare attenzione alla loro forma di *rappresentazione*, la *document community* «sceglieva di standardizzare la rappresentazione dei dati» e di generalizzare il *markup*, introducendo lo standard SGML e favorendo la marcatura non solo degli «aspetti tipografici dei documenti», ma anche di «proprietà più generali dei testi».

Le conseguenze di questi diversi orientamenti furono decisive nelle scelte delle comunità di ricerca che si dedicavano all'elaborazione automatica dei testi. Le tre associazioni scientifiche che li rappresentavano, la *Association for Computational Linguistics* (ACL), la *Association for Literary and Linguistic Computing* (ALLC) e la *Association for Computers and the Humanities* (ACH) decisero di promuovere la *Text Encoding Initiative* (TEI) e di adottare lo standard SGML per la marcatura dei testi. Tuttavia, benché il problema della semantica dei dati «non fosse irrilevante per la *document community*», sembrava che la definizione di una semantica per il contenuto dei

---

<sup>6</sup> D. Raymond et al., "From Data Representation to Data Model: Meta-semantic issues in the evolution of SGML". In *Computer Standards & Interfaces*. 18, 1996, p. 25-36.

documenti «fosse un problema di difficile soluzione». Allo stesso modo, anche «i tentativi di definire una semantica nella comunità degli studiosi, e particolarmente nella Text Encoding Initiative, incontrarono molte resistenze». Così «la via indicata dallo SGML» sembrò allora la più «ragionevole» e si scelse

[di] insistere sul concetto dell'indipendenza dalle macchine e dai programmi e [di] fornire la base su cui in seguito si sarebbe potuta sviluppare la semantica, evitando però di proporre effettivamente una particolare specificazione<sup>7</sup>.

La conseguenza fu quella di spostare l'attenzione dall'elaborazione del contenuto alla mera rappresentazione dei dati. Delle due fondamentali questioni di cui si occupa l'informatica, e con essa l'informatica umanistica, la rappresentazione e l'elaborazione dell'informazione, l'elaborazione del contenuto è passata sempre più in secondo piano. L'improvvisa espansione del web e la possibilità di presentare sul web contenuti multimediali hanno ulteriormente accentuato questa tendenza.

5. I linguaggi su cui si fonda l'architettura del web, lo HTML e ora sempre più lo XML, sono, non a caso, linguaggi per la rappresentazione dei dati. Sono linguaggi che esprimono la struttura della rappresentazione, non la struttura di ciò che tale rappresentazione rappresenta; in altri termini, tali linguaggi non sono in grado di esprimere la struttura del contenuto informativo della rappresentazione, se non quando le due strutture siano isomorfe e possano essere poste in corrispondenza biunivoca. Di conseguenza, anche le

---

<sup>7</sup> Ivi, p. 28.

elaborazioni applicabili alle risorse web, sono in gran parte dipendenti dai formati dei documenti, ossia dalla struttura della rappresentazione dei contenuti informativi che essi veicolano. Le possibilità di elaborazione dei documenti digitali accessibili sul web dipendono dalla struttura che tali linguaggi assegnano ai documenti, ossia dai vincoli imposti da una struttura gerarchica ad albero. In effetti, il linguaggio usato per elaborare dati in formato XML, lo XSLT, «assume in ingresso (*input*) una struttura ad albero e genera in uscita (*output*) un'altra struttura ad albero»<sup>8</sup>. Esso è uno dei componenti della famiglia di linguaggi XSL (*Extensible Stylesheet Language*), introdotti «per definire la trasformazione e la presentazione di documenti XML»<sup>9</sup>, e se può essere certamente considerato «un linguaggio di programmazione completo», viene principalmente impiegato «per trasformare i dati d'origine nella forma di presentazione voluta». In realtà lo XSLT può servire anche per produrre «dati di ingresso (*input*) per un'altra applicazione»<sup>10</sup>, ma in ogni caso trasforma i dati d'origine XML in altre strutture di dati organizzate gerarchicamente ad albero; esso conserva così la struttura tipica del formato dei documenti digitali che veicolano l'informazione e non può dare risposta all'esigenza di elaborare la struttura propria del loro contenuto informativo.

---

<sup>8</sup> M. Kay, *What Kind of Language is XSLT? An analysis and overview*, <http://www.ibm.com/developerworks/library/x-xslt> (consultato il 23 marzo 2012).

<sup>9</sup> W3C, *The Extensible Stylesheet Language Family (XSL)*, <http://www.w3.org/Style/XSL> (consultato il 23 marzo 2012).

<sup>10</sup> M. Kay, *What Kind of Language is XSLT?*, cit.

Il web, d'altronde, nella sua forma originaria, era stato concepito per permettere l'accesso a distanza e la visualizzazione dei documenti ed è proprio per questa sua caratteristica che molti hanno potuto considerarlo come una potenziale biblioteca universale e che alcuni progetti miranti esplicitamente a questo scopo abbiano potuto essere intrapresi<sup>11</sup>. Spesso, tuttavia, è stato osservato che tali progetti hanno mirato più alla quantità che all'accuratezza e si è rafforzata l'opinione che l'«ossessione per i numeri debba cedere il passo alle preoccupazioni per la qualità»<sup>12</sup>. In effetti, ciò che viene prodotto da tali progetti di «digitalizzazione di massa» non è altro che «l'immagine di una copia del libro, una fotografia digitale delle sue pagine, associata a funzioni di ricerca molto limitate»<sup>13</sup>, che operano sui dati spesso imprecisi forniti da una lettura ottica automatica e priva di ulteriori controlli. Sicché, pur con l'aggiunta di un corredo essenziale di metadati, queste «enormi quantità di testo» vengono presentate «con un valore aggiunto minimale», se paragonato a quello delle edizioni digitali vere e proprie, in cui «piccole quantità

---

<sup>11</sup> Degni di nota «tre grandiosi tentativi», quali il Million Books Project, promosso nel 2002 dalla Carnegie Mellon University e chiamato anche Universal Digital Library (<http://www.ulib.org>, consultato il 27 marzo 2012); il Google Books Search, proposto nel 2002 da Larry Page, uno dei due fondatori di Google, e cominciato nel 2004 (<http://books.google.com>, consultato il 27 marzo 2012); e il MSN Book Search, della Microsoft, annunciato nel 2005 e cessato nel 2008: cfr. M. Deegan, K. Sutherland, *Transferred Illusions: Digital technology and the forms of print*, Ashgate, Farnham/Burlington, VT, 2009, p. 136 e sgg.

<sup>12</sup> J.-N. Jeanneney, *Google and the Myth of Universal Knowledge: A view from Europe*, Chicago, University of Chicago Press, 2007, p. 78.

<sup>13</sup> M. Deegan, K. Sutherland, *Transferred Illusions* cit., p. 146.

di dati testuali codificati» sono pubblicati «con un considerevole valore aggiunto», fatto di testimoni diversi del testo, annotazioni, immagini, glossari e link esterni, organizzati in ambienti di lavoro (*work-sites*) complessi e dotati di sofisticate funzionalità<sup>14</sup>.

Di conseguenza, tali grandi progetti di digitalizzazione sono stati accolti con molte perplessità. Deegan and Sutherland hanno osservato che «ponendo mente alla ricerca», spesso non ci si guadagna altro che l'accesso a «contenuti superati, giustamente trascurati e inaffidabili» e si finisce col «risuscitare e riciclare un sapere stantio e non più attuale»<sup>15</sup>. Inoltre, esse fanno notare che «lo spazio informativo (*information space*) di Internet è di tipo completamente diverso da quello di una biblioteca»<sup>16</sup>, sicché ora «il paradigma della biblioteca universale è quello di Internet e niente affatto quello di una biblioteca»<sup>17</sup>. In questo spazio informativo aperto non vengono trasferiti i principi di organizzazione dell'informazione propri della tradizione biblioteconomica e quindi «si perde uno dei maggiori benefici che le biblioteche recano a quello spazio quasi senza confine che è costituito dalla nostra cultura erudita: l'ordine»<sup>18</sup>. Infine, sempre dal punto di vista della ricerca, il tipo di accesso ai testi imposto dal web rischierebbe di compromettere la capacità di interagire criticamente con essi, significativamente e in profondità, al punto da far sorgere una nuova forme di apprensione. Questa

---

<sup>14</sup> Ivi, p. 142.

<sup>15</sup> Ivi, p. 145.

<sup>16</sup> Ivi, p. 150.

<sup>17</sup> Ivi, p. 151.

<sup>18</sup> Ivi, p. 149

preoccupazione affatto nuova (*new anxiety*) risiede in questo: è possibile che l'accesso istantaneo all'informazione, l'elaborazione automatica del testo e quel genere di stimoli visivi e di immedesimazione con lo schermo (*screen absorption*) che sostituiscono la lettura compro-mettano l'esercizio delle preziose capacità riflessive proprie del processo cerebrale della lettura (*reading brain*)?<sup>19</sup>

Si tratta semplicemente dell'inquietudine per il nuovo o di un problema reale, che modifica le nostre stesse attitudini mentali?

Si può argomentare che i grandi progetti di digitalizzazione non sono stati ideati come applicazioni proprie dell'informatica umanistica, ma che nascono più specificamente dalle soluzioni che la moderna tecnologia è in grado di offrire per facilitare l'accesso al materiale librario. Infatti, di *libri* si tratta e non propriamente di testi: il libro è solo il «veicolo fisico» del testo<sup>20</sup>, l'«architettura dell'informazione»<sup>21</sup> che esso contiene, e come Deegan e Sutherland non mancano di far notare, il progetto «Google Book Search (si noti – dicono – il corsivo aggiunto) non mette a nostra disposizione del testo elettronico, ma mette a nostra disposizione dei libri»<sup>22</sup>. Tuttavia, anche le edizioni digitali e altri veri e propri prodotti della ricerca nel campo dell'informatica umanistica sono stati condizionati

---

<sup>19</sup> Ivi, p. 173. Il riferimento di Deegan e Sutherland al *reading brain* rimanda al saggio di M. Wolf, *Proust and the Squid: The story and science of the reading brain*, Cambridge, Icon Books, 2008, in cui si mette in risalto la funzione essenziale e la portata evolutiva della lettura nello sviluppo delle capacità mentali della specie umana.

<sup>20</sup> Ivi, p. 91.

<sup>21</sup> B. Cope, A. Phillips (ed.), *The Future of the Book in the Digital Age*, Oxford, Chandos Publishing, 2006, p. 8.

<sup>22</sup> M. Deegan, K. Sutherland, *Transferred Illusions* cit., p. 147.

dalle nuove funzionalità che il *Web* ha permesso di utilizzare, tanto che l'impatto delle nuove tecnologie di rete ha prodotto effetti quasi paradossali, che hanno agito in controtendenza rispetto allo sviluppo dei metodi computazionali applicabili alle discipline umanistiche. La comparsa del testo elettronico sul web sembra aver favorito in misura maggiore la «simulazione di specifici prodotti di stampa»<sup>23</sup>, le cui «caratteristiche libresche», anche se non nella forma dell'«oggetto fisico», sono state «conservate in modo trionfale»<sup>24</sup>, senza peraltro contribuire a stimolare ed estendere la nostra interazione critica col testo. A giudizio delle due studiose, il «pensare in profondità» è strettamente connesso col «leggere in profondità», ma «né l'una né l'altra attività viene agevolata dall'impazienza e dall'applicazione ininterrotta delle funzionalità offerte dalla rete – l'accesso istantaneo, i collegamenti multipli (*multi-linking*), l'elaborazione simultanea (*multi-tasking*) e la ridotta interazione con i diversi spazi informativi»<sup>25</sup>. E come ci si può chiedere se «gli strumenti di *text mining* possano effettivamente servire a dar senso alla letteratura»<sup>26</sup>, allo stesso modo, nel campo delle edizioni digitali, si può pensare che «le potenzialità del *computer* come strumento di visualizzazione abbiano prevalso sui suoi usi propriamente 'computazionali'»<sup>27</sup>. Così, l'attenzione è stata rivolta soprattutto ai problemi della rappresentazione del testo, trascurando o ponendo in secondo piano lo sviluppo di procedure automatiche per l'elabo-

---

<sup>23</sup> Ivi, p. 27.

<sup>24</sup> Ivi, p. 116-117.

<sup>25</sup> Ivi, p. 180.

<sup>26</sup> Ivi, p. 176.

<sup>27</sup> Ivi, p. 75.

razione del suo contenuto informativo. Insomma, la rappresentazione dell'*espressione*, o della materialità del testo, ha prevalso sull'*elaborazione* del suo *contenuto*, ovvero dell'informazione che esso trasmette, e sebbene la rappresentazione del testo materiale sia costituita da un testo digitale, la *rappresentazione* del testo materiale ha prevalso sull'*elaborazione* dell'informazione contenuta nell'oggetto digitale che lo rappresenta.

6. È possibile cercare di comprendere questa situazione apparentemente paradossale e di rispondere in modo adeguato alla *new anxiety* avvertita dai critici più attenti della portata innovativa delle nuove tecnologie e del loro impatto sullo sviluppo delle nostre pratiche culturali? L'indizio di un nuovo punto di vista da cui considerare la questione può essere colto nell'uso, apparentemente non intenzionale, di particolari metafore riferite allo scambio di informazioni in rete. Esse illustrano in modo incisivo la fase di passaggio dallo scambio bilaterale e diretto dell'informazione tra sistemi isolati alla diffusione distribuita dell'informazione nel nuovo sistema informativo integrato costituito dal web.

Questo passaggio può essere descritto come un passaggio dal *Gopher* a *Lynx*. Il *Gopher* (1991) era un protocollo progettato per distribuire, ricercare e recuperare documenti in Internet facendo uso di server appositamente predisposti. *Lynx* (1992) era un *web browser* per terminali funzionanti in modalità carattere e privi di interfaccia grafica, in grado di supportare diversi protocolli, tra cui il *Gopher* e il protocollo HTTP che era stato introdotto per lo scambio di dati nel *WorldWideWeb*. È evidente che ci troviamo, con ciò, in una fase di passaggio tra la sperimentazione di sistemi isolati e il successo del nuovo sistema distribuito per la diffusione dell'informazione. L'uso metaforico dei termini, *gopher* e *lynx*, si fonda, rispettivamente,

sull'omofonia con le espressioni 'go for' (o 'go far') e 'links,' che alludono al modo di operare dei *gopher* e alla navigazione nel web attraverso i collegamenti (*links*) ipertestuali. È singolare però che si tratti di metafore biologiche che fanno leva sul nome di due specie animali, un roditore della famiglia dei geomidi (il *gopher*) e la lince (*lynx*), un genere della famiglia dei felini. Ma le metafore biologiche con le specie animali richiamano l'idea dell'evoluzione e proprio in questi termini viene portato a ragionare chi trova scritto che «il protocollo *Gopher* è ancora usato da alcuni entusiasti» e che «rimane ancora una piccola *popolazione* di *server* mantenuti in funzione». <sup>28</sup>

7. Che cosa ci permette di cogliere, a proposito del nostro problema, una prospettiva di tipo evolutivo? In primo luogo l'importanza dello sviluppo delle *funzioni* dell'organismo –nel nostro caso, dell'oggetto digitale – insieme all'evoluzione del suo *aspetto*. Se dunque, da una parte, gli oggetti digitali accessibili sul web possono essere visualizzati e presentati come riproduzioni fedeli di oggetti fisici corrispondenti, dall'altra, essi possono essere valutati sulla base delle possibilità di elaborazione del loro contenuto informativo. Da questo punto di vista, si possono riscontrare sviluppi convergenti in campi di ricerca diversi e interconnessi.

Nel campo dell'informatica umanistica si sta affermando la consapevolezza che i suoi metodi e le sue procedure convergono con quelli propri della rappresentazione della conoscenza, praticati nel campo dell'intelligenza artificiale:

---

<sup>28</sup> Cfr. "Gopher (protocol)". In *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, [http://en.wikipedia.org/wiki/Gopher\\_\(protocol\)](http://en.wikipedia.org/wiki/Gopher_(protocol)) (consultato il 5 aprile 2012).

D. Buzzetti, *Che cos'è, oggi, l'informatica umanistica?*

Per l'informatica umanistica, la rappresentazione della conoscenza è un modo convincente, illuminante, e produttivo di fare ricerca umanistica – e per molti versi è ciò che l'informatica umanistica è venuta facendo, implicitamente, per anni.<sup>29</sup>

John Unsworth, che ne è l'autore, trae questa affermazione da un confronto tra le funzioni svolte dalla rappresentazione della conoscenza nella pratica dell'intelligenza artificiale e le diverse forme di rappresentazione adottate nel campo dell'informatica umanistica. Alla domanda su che cosa sia l'informatica umanistica, Unsworth risponde nel modo seguente:

L'informatica umanistica è una pratica di rappresentazione, una forma di modellizzazione o, come afferma Wallace Stevens, di imitazione. È anche (come Davis e i suoi coautori affermano) un modo di ragionare e un insieme di assunti ontologici, e la sua pratica di rappresentazione è stata determinata dal bisogno di computazione efficiente, da un lato, e di comunicazione umana dall'altro.<sup>30</sup>

Questa caratterizzazione dell'informatica umanistica è ricavata dall'analogia risposta alla domanda su che cos'è la rappresentazione della conoscenza in intelligenza artificiale che Randall Davis, Howard Shrobe e Peter Szolovits propongono «nei termini di cinque

---

<sup>29</sup> J. Unsworth, "Knowledge Representation in Humanities Computing", presentato a *eHumanities*, NEH Lecture Series on Technology & the Humanities, Lecture I, Washington, DC, April 3, 2001, <http://people.lis.illinois.edu/~unsworth/KR> (consultato il 5 aprile 2012).

<sup>30</sup> Id., "What is Humanities Computing and What is Not?". In *Jahrbuch für Computerphilologie*, 4, 2002, p. 71-84, <http://computerphilologie.tudarmstadt.de/jg02/unsworth.html> (consultato il 5 aprile 2012).

importanti e distintamente diversi ruoli che una rappresentazione svolge» e secondo la quale la rappresentazione può essere considerata come (a) «un surrogato», ovvero un modello – dice Unsworth – dell'oggetto rappresentato, (b) «una teoria frammentaria di ragionamento intelligente», (c) «un insieme di assunzioni ontologiche», (d) «un mezzo (*means*) per una computazione pragmaticamente efficiente» e, infine, (e) «un mezzo (*medium*) di espressione umana». <sup>31</sup>

Come si può osservare, le forme di rappresentazione praticate dall'informatica umanistica sono qui definite in funzione dell'efficienza computazionale. Rappresentazione ed elaborazione dell'informazione non restano quindi separate, ma la forma di rappresentazione risulta progettata in funzione dell'applicazione di procedure computazionali volte a produrre risultati utili sul piano dell'elaborazione dell'informazione. Le conseguenze di questa concezione sono tratte dallo stesso Unsworth in un articolo presentato nel gennaio 2003 ad un convegno sul tema *Transforming Disciplines: Computer Science and the Humanities*, dove scrive:

Noi dobbiamo (dobbiamo ancora) dimostrare l'utilità di tutti i materiali che abbiamo digitalizzato nel corso degli ultimi dieci anni ed oltre – e dimostrarne l'utilità non solo per quanto riguarda la maggiore facilità di accesso, ma in modo più preciso, per quanto riguarda *ciò che ne possiamo fare* quando li abbiamo a nostra disposizione [...] <sup>32</sup>

---

<sup>31</sup> R. Davis, H. Shrobe, P. Szolovits, "What is a Knowledge Representation?". In *AI Magazine*. 14, 1993(1), p. 17-33.

<sup>32</sup> J. Unsworth, "Tool-Time, or 'Haven't We Been Here Already?' Ten Years in Humanities Computing", presentato al convegno *Transforming Disciplines: The*

Si tratta di un «invito evidente ad andare 'oltre il rappresentare'»<sup>33</sup>, come recita esplicitamente il titolo di uno dei suoi articoli pubblicati successivamente, dove dichiara in modo ancora più diretto:

Abbiamo speso una generazione costruendo furiosamente biblioteche digitali e sono sicuro che ora costruiremo strumenti da usare in quelle biblioteche altrettanto furiosamente, per almeno un'altra generazione, e non vedo l'ora che ciò si faccia. Sono sicuro che il testo non verrà dimenticato, mentre saremo occupati nella costruzione dei nostri strumenti – ma sono anche certo che i nostri strumenti ci porranno in relazione con i testi in forme del tutto nuove.

Nello stesso saggio, Unsworth coglie distintamente il succedersi delle diverse fasi di sviluppo dell'informatica umanistica, che qui sono state messe in relazione con il parallelo sviluppo della tecnologia e con l'introduzione di importanti innovazioni tecnologiche. Così, infatti, egli afferma:

siamo, credo, all'inizio di quella che mi sembra la terza fase principale dell'informatica umanistica (*humanities computing*), che è passata dagli strumenti, negli anni '50, '60 e '70, alle fonti primarie, negli anni '80 e '90, e ora sembra ritornare di nuovo agli strumenti.

---

*Humanities and Computer Science*, Washington DC, 17-18 gennaio 2003, <http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html> (consultato il 5 aprile 2012).

<sup>33</sup> D Buzzetti, "Oltre il rappresentare: Le potenzialità del markup". In *La macchina nel tempo: Studi di informatica umanistica in onore di Tito Orlandi*, a cura di L. Perilli e D. Fiormonte, Firenze, Le Lettere, 2011, p. 40.

Sicché, continua, «credo che stiamo arrivando ad un momento in cui la forma di attenzione che prestiamo alle fonti primarie si stia spostando dalla digitalizzazione all'analisi, [...] dalla rappresentazione all'astrazione», e non potremmo trovare una formulazione più chiara del passaggio dalla preoccupazione prevalente per la rappresentazione delle fonti a quella per l'elaborazione del contenuto.<sup>34</sup>

8. Ed è proprio dal punto di vista dell'elaborazione del contenuto informativo degli oggetti digitali che si realizza la convergenza dell'informatica umanistica con le tendenze in atto in altri e diversi campi contigui di ricerca. Così nel campo specifico dell'organizzazione della conoscenza e della catalogazione a soggetto, Vanda Broughton, esperta di classificazione a faccette e di costruzione di thesauri, afferma:

Il lavoro di collaborazione attualmente in corso con studiosi operanti nel campo dell'informatica umanistica suggerisce che l'uso combinato di metodi di analisi a faccette e metodi di codifica del testo possono offrire una soluzione per migliorare la praticabilità degli strumenti di metadattazione e fornire mezzi più ingegnosi e sofisticati per la rappresentazione dei soggetti.<sup>35</sup>

---

<sup>34</sup> J. Unsworth, "Forms of Attention: Digital Humanities Beyond Representation", presentato al terzo convegno del Canadian Symposium on Text Analysis (CaSTA), *The Face of Text: Computer-Assisted Text Analysis in the Humanities*, McMaster University, 19-21 novembre 2004, <http://www3.isrl.uiuc.edu/~unsworth/FOA> (consultato il 6 aprile 2012).

<sup>35</sup> V. Broughton, Finding Bliss on the Web: Some problems of representing faceted terminologies in digital environment, in C. Gnoli, F. Mazzocchi (ed.), "Paradigms and Conceptual Systems". In *Knowledge Organization, Proceedings of the Eleventh*

Ed è importante osservare che la convergenza si realizzi proprio sul piano dell'analisi del contenuto, diretta, in questo caso, ai fini dell'indicizzazione delle materie trattate.

Questo lavoro di metadatazione e di descrizione semantica del contenuto delle risorse digitali, che interessa tanto l'informatica umanistica, quanto l'organizzazione della conoscenza, converge poi significativamente con l'introduzione delle tecnologie che vengono attualmente sviluppate per la costruzione del semantic web. Così, è proprio l'evoluzione tecnologica del web che permette di recuperare quell'aspetto parzialmente negletto dell'informatica umanistica che la sua iniziale costruzione aveva momentaneamente e quasi paradossalmente contribuito a oscurare. Programmaticamente, infatti, il web semantico si pone l'obiettivo dell'elaborazione dell'informazione veicolata dalle risorse digitali che attraverso il web possono essere attinte e visualizzate: «l'elaborazione dei contenuti semantici sta al cuore della visione del *Semantic Web*»<sup>36</sup>. È quindi naturale che questo sia il terreno di convergenza delle più recenti tendenze della ricerca nel campo dell'informatica umanistica e nelle aree più vicine. La sfida posta dall'evoluzione tecnologica riporta l'informatica umanistica alle sue fondamentali motivazioni originarie.

---

*International ISKO Conference, Rome, 23-26 February 2010 (Advances in Knowledge Organization, 12), Würzburg, Ergon-Verlag, 2010, p. 193.*

<sup>36</sup> N. Calzolari, "Language resources in the semantic Web vision". In C. Zong (ed), *International Conference on Natural Language Processing and Knowledge Engineering: Proceedings*, Beijing, IEEE Press, 2003, p. 16.

9. Possiamo così concludere queste considerazioni accennando ad un esempio significativo di questa convergenza e ad alcune iniziative di ricerca specifiche nel campo dell'informatica umanistica, il cui orientamento si rivolge manifestamente nella stessa direzione. Il primo esempio riguarda lo sviluppo delle cosiddette *semantic digital libraries*<sup>37</sup>. L'idea nasce dall'esigenza di sviluppare ed estendere i sistemi di gestione delle biblioteche digitali (DLMS) integrando tecnologie proprie del web semantico (3.0) e del *social networking* (2.0)<sup>38</sup>. La nuova concezione delle "biblioteche digitali semantiche" si fonda sul presupposto che «le tecnologie semantiche» e dei *social networks* «siano in grado di offrire soluzioni più efficienti per la costruzione di procedure affidabili e facili da usare per accedere ai contenuti e ai metadati»<sup>39</sup>. In questa prospettiva, le biblioteche digitali sono destinate ad evolversi in «sistemi complessi che integrano servizi avanzati di biblioteca digitale» e nuovi servizi di «sostegno alle comunità di pratica e di ricerca». Per ottenere questo risultato, «in una biblioteca digitale, l'organizzazione della conoscenza entra in gioco in forme diverse e strettamente interrelate» e, più specificamente, come gestione dell'«informazione su termini e concetti e sulle loro relazioni», oppure come «organizzazione di informazione ontologica e lessicale»<sup>40</sup>.

---

<sup>37</sup> Cfr. *Semantic Digital Libraries: Bringing Libraries to Web 3.0*, <http://semdl.info/> (consultato il 7 aprile 2012).

<sup>38</sup> Ivi, <http://semdl.info/tutorials/6/abstract> (consultato il 7 aprile 2012).

<sup>39</sup> S. R. Kruk et al., *Architecture of Semantic Digital Libraries*, Digital Enterprise Research Institute (DERI), National University of Ireland, Galway, Work Package 4, 30 June 2008, p. 1.

<sup>40</sup> D. Soergel, "Digital Libraries and Knowledge Organization". In S. R. Kruk and B. McDaniel (ed.), *Semantic Digital Libraries*, Berlin, Springer, 2009, p. 9 e 21.

Il progetto delle biblioteche digitali semantiche si propone come una forma di risposta alla 'nuova preoccupazione' provocata dal web e al timore di perdere quell'«ordine»<sup>41</sup> proprio dei sistemi bibliotecari tradizionali, che lo spazio informativo offerto dal web non riuscirebbe a ricostituire. Si può tuttavia ribattere, a questo proposito, che le tecnologie del web semantico, ovvero di quello che è stato chiamato il *Web 3.0*<sup>42</sup>, offrono «tecniche di ricerca efficienti nel nuovo e interconnesso spazio informativo» delle risorse accessibili in rete. Infatti, lo spazio informativo del *Web* non è uno spazio completamente aperto e privo di struttura, ma possiede una propria forma di organizzazione costituita dai *link* che collegano risorse diverse, o parti diverse di risorse, o documenti digitali. Inoltre, il ricorso alle ontologie, la tecnologia semantica dotata di maggiore potenza inferenziale, produce nuove forme di organizzazione dell'informazione che non si riducono ad «una mera specificazione di schemi di metadati» già definiti e consolidati nella pratica di catalogazione e descrizione bibliografica tradizionale, ma permettono ai metadati «di diventare più aperti, destrutturati e, ciò che più importa, altamente interconnessi». Infatti,

---

<sup>41</sup> M. Deegan, K. Sutherland, *Transferred Illusions*, Arnham ; Burlington, Ashgate, 2009, pp. 149 e 173.

<sup>42</sup> «'Semantic Web' è talvolta usato come sinonimo di 'Web 3.0', anche se la definizione di ciascun termine varia», "Semantic Web". In *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, [http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web) (consultato il 5 aprile 2012); cfr., p. es., J. Markoff, "Entrepreneurs See a Web Guided by Common Sense". In *The New York Times*, 12 novembre 2006, <http://turing.cs.washington.edu/NYT-KnowItAll.htm> (consultato il 9 aprile 2012): «Web 3.0 – o 'Web semantico', dall'idea di aggiungere significato».

le ontologie non introducono vincoli sui metadati, ma permettono piuttosto di ragionare su concetti interconnessi, in quanto assegnano un significato agli oggetti informativi e alle loro relazioni. Quindi, nelle biblioteche digitali semantiche, i metadati diventano una rete di concetti interconnessi e dotati di senso.<sup>43</sup>

Le biblioteche digitali semantiche intendono inoltre avvalersi dell'integrazione delle tecnologie del web semantico con le forme di indicizzazione e classificazione praticate da comunità di utenti nel cosiddetto *Web 2.0*<sup>44</sup> e si presentano anche come *social semantic digital libraries*<sup>45</sup>. Così, ad esempio, le *folksonomies*, ossia la «pratica di categorizzazione collaborativa mediante parole chiave scelte liberamente», vengono valutate come «un meccanismo leggero (*lightweight*) e facilmente accessibile per la creazione di ontologie e metadati», tale da permettere l'affiorare di una «semantica emergente (*emergent semantics*)» e rendere concreta «l'aspettativa che le interazioni individuali di un gran numero di agenti razionali producano effetti globali valutabili come una semantica»; in questo modo, «le ontologie diventerebbero un effetto emergente del

---

<sup>43</sup> S. R. Kruk et al., "Architecture of Semantic Digital Libraries". In S. R. Kruk., McDaniel (ed.), *Semantic Digital Libraries*, cit., pp. 78-79.

<sup>44</sup> «Un sito Web 2.0 permette agli utenti di interagire e collaborare tra loro, dialogando in un *social medium* come produttori-consumatori (*prosumer*) di contenuti generati dagli utenti in una comunità virtuale» ("Web 2.0", in *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, [http://en.wikipedia.org/wiki/Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0); (consultato il 9 aprile 2012).

<sup>45</sup> Cfr. S. R. Kruk et al., *The anatomy of a Social Semantic Digital Library*, [http://www.contexta.cl/mediawiki/images/7/72/The\\_anatomy\\_of\\_a\\_Social\\_Semantic\\_Digital\\_Library.pdf](http://www.contexta.cl/mediawiki/images/7/72/The_anatomy_of_a_Social_Semantic_Digital_Library.pdf) (consultato il 9 aprile 2012).

sistema, anziché costituire una convenzione fissa e limitata della maggioranza». <sup>46</sup>

Di particolare interesse, a questo proposito, sono le prime implementazioni di sistemi di *social tagging* che permettono di produrre annotazioni nella forma di *linked data*<sup>47</sup>. Si tratta di ontologie cosiddette "leggere" (*lightweight*) sviluppate per l'annotazione, «ontologie di annotazione (*tag ontologies*)» che «permettono di ottenere rappresentazioni interpretabili automaticamente (*machine-readable*) capaci di essere condivise da diversi sistemi di annotazione collaborativa (*social tagging*)»<sup>48</sup>. In questo modo, si realizza l'integrazione tra le tecnologie del web 2.0, il web orientato all'utente, e il web 3.0, il web orientato alle macchine e all'elaborazione automatica dell'informazione. Il risultato più significativo di questa convergenza tra l'inserimento di metadati semantici aggiunti dagli utenti e la loro acquisizione ed elaborazione in un sistema integrato di organizzazione della conoscenza è appunto costituito dalla produzione di *Open Linked Data*, un processo attualmente in corso, di natura quasi evolutiva, paragonabile alla crescita e all'espansione

---

<sup>46</sup> P. Mika, "Ontologies Are Us: A unified model of social networks and semantics". In *Journal of Web Semantics*, 5, 2007(1), p. 5-7.

<sup>47</sup> Il termine «descrive un metodo di pubblicazione di dati strutturati in modo che possano essere interconnessi tra loro e poter essere meglio utilizzati», "Linked data". In *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, [http://en.wikipedia.org/wiki/Linked\\_data](http://en.wikipedia.org/wiki/Linked_data) (consultato il 10 aprile 2012).

<sup>48</sup> H. L. Kim et al., "The state of the art in tag ontologies: A semantic model for tagging and folksonomies". In *DCMI '08: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*, Dublin Core Metadata Initiative, 2008, <http://dc2008.de/wp-content/uploads/2008/09/kim-scerri-breslin-decker-kim.pdf> (consultato il 10 aprile 2012).

iniziale del web, il cui sviluppo assume forme del tutto analoghe a un fenomeno biologico di diffusione virale.<sup>49</sup>

10. Questi recenti sviluppi delle tecnologie informatiche favoriscono il ritorno dell'informatica umanistica alla sua ispirazione originaria: l'«attenzione» che in una successiva fase di sviluppo si era rivolta principalmente «alla rappresentazione delle fonti primarie» torna a rivolgersi «alla costruzione degli strumenti» per l'elaborazione del loro contenuto informativo<sup>50</sup>. Anche per le discipline umanistiche «il *Web* semantico» rappresenta «il futuro» e l'informatica umanistica è chiamata a fornire «rappresentazioni formali del lascito documentario dell'umanità (*human record*)», adatte cioè ad essere elaborate automaticamente. Infatti, come ancora ricorda John Unsworth,

queste rappresentazioni – ontologie, schemi, rappresentazioni della conoscenza, comunque le si voglia chiamare – dovrebbero essere prodotte da persone formate negli studi umanistici. E la disciplina che li produce richiede una formazione umanistica, unita ad una conoscenza di elementi di matematica, logica, ingegneria e informatica. [...] C'è una grande quantità di lavoro da fare – e indubbiamente non tutto di natura tecnica. Nella costruzione di questa grande mappa del

---

<sup>49</sup> Cfr. il *Linking Open Data (LOD) Project Cloud Diagram*, <http://richard.cyaniak.de/2007/10/lod/> (consultato il 9 aprile 2012); per il *Linking Open Data community project*, cfr. [http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project\\_Description](http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description) (consultato il 9 aprile 2012). Nelle intenzioni dei promotori dell'iniziativa, «i *Linked Data* si propongono di usare il Web per collegare dati in relazione tra loro», *Linked Data: Connect Distributed Data across the Web*, <http://linkeddata.org> (consultato il 10 aprile 2012).

<sup>50</sup> J. Unsworth, *Forms of Attention*, cit.

sapere una grande parte sarà costituita da lavoro collaborativo (social work), creazione del consenso, compromesso. Ma anche quest'attività dovrà essere affidata a persone che sappiano come il consenso possa essere raggiunto ed espresso in un *medium* di natura computazionale.<sup>51</sup>

Possiamo ora, in conclusione, ritornare al nostro esempio iniziale che ci invitava a riflettere sulla natura computazionale delle edizioni digitali, per cercare di illustrare le sfide che attendono l'informatica umanistica nella fase presente, contrassegnata dal diffondersi delle attuali tecnologie semantiche. Dopo lo sviluppo delle tecnologie per la *rappresentazione* del testo con l'impiego di marcatori inseriti all'interno della stringa di caratteri che ne costituisce la rappresentazione codificata di base (*embedded markup*), il problema dell'*elaborazione* del contenuto informativo dei dati testuali si presenta in forme nuove.

Attualmente, la rappresentazione critica dei testi si attiene soprattutto al canone impostosi con l'affermarsi e il diffondersi della pratica di codifica testuale fondata sulle norme della Text Encoding Initiative (TEI). Il problema dell'*elaborazione* del contenuto informativo dei dati testuali si pone quindi come il tentativo di mettere in relazione la descrizione delle caratteristiche testuali ottenuta con l'impiego di una forma di *embedded markup*, qual è quella rappresentata dalla codifica TEI, con le possibili forme di indicizzazione o di rappresentazione semantica del contenuto del testo. Il problema non è di facile soluzione, perché le strutture e le proprietà della rappresentazione del testo non sono isomorfe alla struttura e agli elementi della rappresentazione semantica del suo

---

<sup>51</sup> Id., *What is Humanities Computing*, cit.

contenuto. Non è ancora stato perfezionato e tanto meno generalmente accettato un modello funzionale adeguato, che permetta una mappatura idonea e computabile tra gli elementi delle due strutture<sup>52</sup>. Si possono tuttavia richiamare due recenti proposte che possono fornire indicazioni utili alla soluzione del problema e che rappresentano due esempi significativi del rinato interesse per lo sviluppo di strumenti utili all'elaborazione dell'informazione veicolata dai dati testuali.

La prima iniziativa è costituita dal progetto Humanities Research Infrastructure and Tools (HRIT) promosso dal Center for Textual Studies and Digital Humanities della Loyola University di Chicago, un'iniziativa che si propone di sviluppare strumenti e servizi infrastrutturali per la creazione di un ambiente collaborativo funzionale al lavoro di preparazione di edizioni critiche digitali. Tra gli obiettivi dichiarati del progetto spicca quello di «costruire un ambiente *open-source on-line*, efficiente e collaborativo, in cui aggregare, collegare ipertestualmente (*link*) o con riferimenti incrociati, e ancora emendare e condividere testi esaminati criticamente (*vetted*) di fonti documentarie primarie – corredati con apparati critici e con analisi e commenti, in forma di codifica (*markup*), annotazioni, indicizzazione con parole chiave, collegamenti ipertestuali (*linking*), e così via»<sup>53</sup>. Come si vede, il progetto mira sia alla rappresentazione dei testi, nella forma di edizioni critiche digitali codificate, sia all'elaborazione e all'analisi del loro contenuto, sviluppando

---

<sup>52</sup> Per una discussione più approfondita del problema, cfr. D. Buzzetti, *Oltre il rappresentare*, in *La macchina nel tempo*, cit., pp. 39-62.

<sup>53</sup> HRIT Wiki, <https://sites.google.com/a/ctsdh.luc.edu/hrit-intranet/> (consultato il 10 aprile 2012).

strumenti funzionali idonei, compatibili e integrabili con tecnologie semantiche e collaborative.

Secondo Desmond Schmidt, «uno degli obiettivi primari del progetto HRIT», al quale partecipa, «è quello di modificare la tecnologia del *markup* in modo da permettere la collaborazione mediante la condivisione e il riuso dei contenuti». <sup>54</sup> La soluzione proposta consiste in una rappresentazione esterna (*standoff*) ed espressa come una stringa di caratteri, di *proprietà* assegnate a pericopi (*ranges*) di testo. Tale soluzione permette di rappresentare, in un unico formato, sia proprietà testuali o dell'*espressione* del testo, sia proprietà interpretative o del *contenuto* del testo. Le variazioni di entrambi gli insiemi di proprietà possono essere rappresentate mediante «un grafo orientato, con un nodo iniziale e un nodo finale» – il formato Multi-Version Document, o MVD<sup>55</sup> – e ci si può chiedere se ci siano vincoli reciproci che connettono le variazioni di entrambi i grafi, per riuscire a mettere in relazione la rappresentazione delle proprietà testuali, solitamente espressa con forme di *embedded markup*, con la rappresentazione delle proprietà interpretative del testo, solitamente espressa con forme di *standoff markup*. Lo stesso Desmond Schmidt riconosce che la proposta di rappresentare le caratteristiche testuali con proprietà esterne (*standoff properties*) alla stringa di caratteri in cui consistono i dati testuali «non è un'idea originale», ma si richiama «in misura maggiore o minore» ad altre implementazioni come, tra l'altro, «i sistemi di annotazione

---

<sup>54</sup> D. Schmidt, "Standoff Properties". In *HRIT: Facilitating interoperability by overcoming overlap*, abstract presentato al DH2012 (in corso di stampa).

<sup>55</sup> Id., "The Inadequacy of Embedded Markup for Cultural Heritage Texts". In *Literary and Linguistic Computing*. 25, 2010(3), p. 350.

semantica» e il cosiddetto «modello della stringa ampliata (*extended string*)»<sup>56</sup> proposto da Manfred Thaller.<sup>57</sup> Ed è questa, in effetti, la seconda linea di ricerca alla quale si intendeva fare riferimento per segnalare, con le parole di Unsworth, nuove «forme di attenzione» allo sviluppo di «strumenti» per l'elaborazione del contenuto informativo dei testi e delle altre risorse digitali accessibili in rete.

La proposta di Manfred Thaller fa riferimento ad un modello per la rappresentazione dinamica dell'informazione trasmessa da un «testo storico», ricavato dall'«*equazione infologica* di Langefors». Nell'implementazione del modello, i «segni materiali (*tokens*) che veicolano l'informazione», sono tenuti «nettamente e inequivocabilmente separati» dal «significato» che viene loro assegnato, sia esso un'istruzione per la loro «resa» grafica, oppure «un descrittore semantico su cui possa operare qualche sistema di recupero dell'informazione (*information retrieval system*)». I diversi tipi di informazione che vengono associati ai segni materiali che la veicolano, ovvero ai dati che la trasmettono, sono rappresentati in uno «spazio interpretativo a *n* dimensioni», che permette di considerare il «testo» come una «stringa di caratteri diversi, ognuno dei quali può essere descritto con un numero arbitrario di proprietà»<sup>58</sup>. Nel quadro del progetto PLANETS, dedicato alla conservazione

---

<sup>56</sup> HRIT (*Humanities Research Infrastructure and Tools*), *Standoff Properties*, <http://dhtestbed.ctsdh.luc.edu/hritinfrastructure/index.php/standoff-properties> (consultato il 12 aprile 2012).

<sup>57</sup> Cfr. M. Thaller, "Strings, Texts and Meaning". In *Digital Humanities 2006: Conference Abstracts*, Paris, CATI - Université Paris-Sorbonne, 2006, p. 212-14.

<sup>58</sup> Id., *What Is a Text within the Digital Humanities, or Some of Them, at Least?*, abstract presentato al DH2012 (in corso di stampa).

D. Buzzetti, *Che cos'è, oggi, l'informatica umanistica?*

digitale<sup>59</sup>, questo modello astratto è stato concretamente applicato allo «sviluppo di strumenti»<sup>60</sup> per confrontare il contenuto informativo di rappresentazioni diverse, secondo diversi formati, di uno stesso oggetto culturale<sup>61</sup>. Anche in questo caso, come si vede, la rappresentazione digitale dell'oggetto e l'elaborazione del suo contenuto informativo vengono chiaramente distinte e considerate degne di pari attenzione.

Lo sviluppo dell'informatica umanistica sembra decisamente avviarsi, nella fase attuale, in questa direzione.

---

<sup>59</sup> Cfr. <http://www.planets-project.eu/about/> (consultato il 12 aprile 2012).

<sup>60</sup> M. Thaller, *What Is a Text*, cit.

<sup>61</sup> Cfr. Id., *The eXtensible Characterisation Languages – XCL*, Hamburg, Kovač, 2009.