

# Web semantico, linked data e studi letterari: verso una nuova convergenza

Fabio Ciotti

## La storia finora: luci e ombre delle Digital Humanities

Uno dei libri più influenti nella storia del rapporto tra informatica e scienze umane è stato senza dubbio il notissimo *Hypertext* di George Landow<sup>1</sup>. In quel volume lo studioso statunitense sosteneva che l'allora incipiente tecnologia dell'ipertesto avrebbe favorito una feconda convergenza tra la tradizione degli studi letterari (almeno per come questi si erano venuti evolvendo negli ultimi decenni del secolo scorso) e il dominio delle discipline computazionali e dei nuovi media digitali.

Dalla pubblicazione della prima edizione di quel libro sono passati ormai venti anni, e possiamo dire che l'era dell'ipertesto è ormai alle spalle: da una parte l'introduzione e l'evoluzione del Web ha per molti versi banalizzato l'ipertesto, rendendolo una forma/tecnologia di organizzazione delle informazioni di uso

---

<sup>1</sup> G. P. Landow, *Hypertext: the convergence of contemporary critical theory and technology*, Baltimore, Johns Hopkins University Press, 1992.

comune, al contempo attenuandone molte delle caratteristiche per così dire rivoluzionarie; dall'altra ha spostato l'attenzione sulla condivisione sociale del sapere, sulla creatività diffusa e sulla cooperazione tra gli individui mediata dalle tecnologie di comunicazione digitale. Naturalmente in questo contesto di profonda trasformazione delle modalità di creazione disseminazione e conservazione della conoscenza anche i saperi umanistici, i loro attori e i loro oggetti, hanno giocato un ruolo importante.

In questo modo l'Informatica umanistica, o *Humanities Computing* nella formulazione anglosassone, si è venuta progressivamente liberando della stigmatizzazione di disciplina di nicchia, riuscendo al contempo a ottenere una presenza rilevante nella didattica offerta dalla facoltà umanistiche (e questo anche in Italia, nonostante ritardi, ritrosie culturali e crisi dell'università in generale abbiano senza dubbio rappresentato e ancora rappresentino fattori di ostacolo); a conseguire importanti risultati sul piano della ricerca; a promuovere e consolidare infrastrutture e organizzazioni per la cooperazione scientifica a livello nazionale e internazionale che raccolgono e coordinano un numero ormai grandissimo di studiosi a livello planetario, organizzano convegni mastodontici e pubblicano monografie e periodici autorevoli<sup>2</sup>.

---

<sup>2</sup> Basti ricordare l'annuale conferenza «Digital Humanities», cui attendono centinaia di ricercatori, la storica rivista «Literary and Linguistic Computing», cui si sono aggiunte più recentemente «Digital Humanities Quarterly», «Text Technology» e «Digital Studies / Le champ numérique», e i due ponderosi volumi miscelanei editi dalla Blackwell: S. Schreibman, R. G. Siemens, J. Unsworth, *A companion to digital humanities*,

La recente e rapida diffusione del termine *Digital Humanities*<sup>3</sup> sancisce sul piano linguistico il successo di questo processo di consolidamento e generalizzazione, che ha attirato l'attenzione anche delle redazioni culturali della grande stampa, come testimonia la serie di articoli scritti dalla giornalista del *New York Times* Patricia Cohen, nel primo del quale si legge:

Members of a new generation of digitally savvy humanists argue it is time to stop looking for inspiration in the next political or philosophical "ism" and start exploring how technology is changing our understanding of the liberal arts. This latest frontier is about method, they say, using powerful technologies and vast stores of digitized materials that previous humanities scholars did not have<sup>4</sup>.

Una semplice rassegna dei risultati più notevoli conseguiti in questo campo richiederebbe di gran lunga troppo spazio per pensare di proporla in questa sede. Ma in questo panorama, se lo osserviamo senza l'ausilio del cannocchiale, quali sono le vette che si stagliano con maggiore rilievo? Quali sono i risultati più importanti che la ormai pluridecennale attività di ricerca nel dominio delle Digital Humanities ha prodotto? A nostro parere possiamo riassumerli nei seguenti punti:

---

Malden, Mass., Blackwell Pub., 2004 e R. G. Siemens, S. Schreibman, *A companion to digital literary studies*, Malden, MA, Blackwell Pub., 2007.

<sup>3</sup> Termine per il quale è invero arduo trovare una traduzione soddisfacente in italiano. Preferiamo pertanto mantenere la formulazione inglese.

<sup>4</sup> P. Cohen, "Digital Keys for Unlocking the Humanities' Riches", *New York Times*, novembre 16, 2010. [http://www.nytimes.com/2010/11/17/arts/17digital.html?\\_r=2&emc=eta1&](http://www.nytimes.com/2010/11/17/arts/17digital.html?_r=2&emc=eta1&).

- 1) La consapevolezza teorica e metodologica, ampiamente condivisa dai centri di elaborazione e dai singoli studiosi più avanzati e propulsivi della comunità, che vede nel rapporto con le metodologie informatiche un elemento epistemologica-mente e teoricamente rilevante e non un semplice fattore strumentale.
- 2) Il concetto di *modellizzazione* come attività intellettuale che caratterizza l'attività di indagine sugli oggetti e i fenomeni culturali mediante il computer, mediando tra il livello della teoria e quello dell'osservazione. Il livello al quale si colloca dunque l'informatica nella ricerca umanistica è quello propriamente del metodo, un metodo che è relativamente *theory-independent* ma che richiede alla teoria la qualità del rigore formale, l'esplicitazione degli enti teorici e delle relazioni tra tali enti che essa presuppone e l'indicazione di procedure per collegare tali enti ai dati osservativi (in ultima analisi al materiale linguistico testuale e a quello documentale o fattuale contestuale).
- 3) La predisposizione di linguaggi e standard condivisi per la modellizzazione, rappresentazione e disseminazione di risorse digitali di qualità, attività legata alla forte cooperazione con la comunità scientifica archivistica e biblioteconomica.
- 4) Le ampie campagne di digitalizzazione di fonti primarie e secondarie in formato testuale e/o immagine facsimilare e la predisposizione di vasti *repositories* on-line che ormai mettono a disposizione in forma libera e gratuita e con livelli di affidabilità linguistica piuttosto elevati una parte importante della tradizione testuale occidentale.
- 5) Lo sviluppo di importanti framework e infrastrutture software per effettuare *information retrieval*, analisi testuale e pubblicazione on-line di tali risorse testuali, in genere disponibili liberamente come prodotti *open source* o *web service*.

A fronte di questi risultati, tutti di vasta portata ma collocati sul piano dei fondamenti teorici e metodologici e su quello delle infrastrutture generali per la ricerca, sta invece la oggettiva limitatezza dei risultati specifici (fatti salvi alcuni meritevoli controesempi) sul piano dei singoli campi disciplinari. Insomma il movimento delle Digital Humanities ha prodotto una notevole mole di risorse e strumenti digitali, ha acquisito una elevata autoconsapevolezza teorica (e più di un qualche riconoscimento istituzionale), ma di rado è riuscita a uscire dal circolo dei suoi addetti ai lavori, a stabilire una relazione scientifica con il *mainstream* della comunità scientifica umanistica; come ha scritto J. Unsworth:

We need (we still need) to demonstrate the usefulness of all the stuff we have digitized over the last decade and more – and usefulness not just in the form of increased access, but specifically, in what we can do with the stuff once we get it: what new questions we could ask, what old ones we could answer<sup>5</sup>.

Più recentemente, restringendo l'orizzonte della riflessione sugli studi letterari anche Willard McCarty ha evidenziato la questione della rilevanza critica dell'informatica letteraria

[...] literary computing is confined to providing evidence for or against what we already know or suspect. It is strongly inhibited in its capacity to surprise. Providing evidence seems justification enough, but evidence becomes increasingly

---

<sup>5</sup> J. Unsworth, "Tool-Time, or 'Haven't We Been Here Already?': Ten Years in Humanities Computing", presentato al *Transforming Disciplines: The Humanities and Computer Science*, Washington, D.C., gennaio 18, 2003. <http://www.iath.virginia.edu/~jmu2m/carnegie-ninch.03.html>.

problematic as the volume of data exceeds the norm for critical practices formed prior to the exponential growth of online resources. As this volume increases, so does the probability of arbitrary choice, and so the ease with which any statement may be connected to any other. Good critics may do better scholarship by finding more of what they need; bad critics may be swiftly becoming worse ones more easily. The point, however, is that literary computing has thereby served only as mutely obedient handmaiden, and so done nothing much to rescue itself from its position of weakness, from which it can hardly deliver the benefits claimed for it by the faithful. It has done little to educate scholars methodologically<sup>6</sup>.

Quali sono i motivi di questo tutto sommato insoddisfacente panorama? È possibile individuare dove l'informatica umanistica in generale e quella letteraria in particolare hanno fallito nel cogliere il punto? Certo, gli ultimi decenni sono stati caratterizzati da voghe culturali troppo lontane dal rigore formale e dall'idea di testo come oggetto linguistico: la Teoria, quella senza aggettivi per dirla con Culler<sup>7</sup>, non si presta facilmente a interagire con il certosino formalismo delle strutture dati e dei linguaggi informatici – salvo poi convolare felicemente a nozze con il "decostruito" ipertesto. Per non parlare poi dell'arena vasta e multiforme degli studi culturali, che spesso di tutto si occupano meno che del testo (e però molti di questi studi non poco si gioverebbero del contributo di alcune delle recenti

---

<sup>6</sup> W. McCarty, "Literary enquiry and experimental method: What has happened? What might?". In L. Dibattista, *Storia della Scienza e Linguistica Computazionale: Sconfinamenti Possibili*, Milano, Franco Angeli, 2009, pp. 40–41. <http://www.mccarty.org.uk/essays/McCarty,%20Literary%20enquiry%20and%20experimental%20method.pdf>.

<sup>7</sup> J. D. Culler, *Teoria della letteratura : una breve introduzione*, Roma, Armando, 1999.

tendenze innovative emerse nell'ambito delle Digital Humanities e di cui parleremo più avanti).

E tuttavia, come rileva lo stesso McCarty, proprio il problema del testo, pur così al centro della riflessione teorica e metodologica delle Digital Humanities, è rimasto tutto sommato scarsamente determinato nell'ambito delle concrete applicazioni dell'informatica agli studi letterari. Il problema è che non disponiamo di teorie del testo che si possano definire in senso stretto formali. Mentre la rappresentazione (e a maggior ragione l'elaborazione) informatica è ontologicamente formale in senso stretto. La lunga storia di quel sottodominio delle Digital Humanities rubricata sotto l'etichetta di *codifica testuale* è consistita nel tentativo non soddisfacente di superare questo duplice divario concettuale. E di conseguenza gli strumenti informatici per l'analisi e l'edizione scientifica dei testi (e i relativi risultati in termini di analisi ed edizioni) hanno quasi sempre deluso le aspettative e non sono riusciti ad acquisire un sufficiente riconoscimento nell'ambito delle discipline per così dire tradizionali. Cioè, nonostante si sia consapevoli del problema teorico, la predisposizione degli strumenti di rappresentazione e analisi concreti ha finora fatto assai poco i conti con le specificità e la complessità degli oggetti e delle procedure di analisi tipiche della ricerca letteraria.

Questo è dovuto anche e soprattutto al fatto che, nonostante le ripetute affermazioni teoriche, assai sporadico e di nicchia è stato l'investimento degli stessi cultori delle Digital Humanities nella definizione di nuovi modelli e linguaggi per la rappresentazione ed elaborazione formale dei complessi oggetti culturali cui si applicano. Più comunemente si sono ereditati e applicati modelli e linguaggi elaborati dall'informatica per finalità e domini diversi.

Paradigmatico il caso del linguaggio XML. Esso ha assunto un ruolo centrale nella costruzione di linguaggi standard per la rappresentazione di dati e metadati, divenendo una sorta di esperanto digitale; in virtù della sua flessibilità, robustezza e delle sue caratteristiche sintattiche è stato ampiamente adottato per la rappresentazione dei dati in ambito umanistico. Il problema è che XML da una parte impone l'adozione di un modello di dati ad albero che non sempre si adatta alla natura strutturale degli oggetti da rappresentare, dall'altra non è in grado di rappresentare adeguatamente i numerosi e complessi livelli semantici che caratterizzano un testo letterario. Anzi, in generale possiamo dire che XML non fornisce alcuna semantica ai dati in modo computazionalmente trattabile. Il comune fraintendimento per cui si parla di "markup semantico" deriva dal fatto che i marcatori sono leggibili e che, di norma, il vocabolario dei linguaggi XML usa termini delle lingue naturali. Ma la semantica "naturale" di tale vocabolario è del tutto inaccessibile a un elaboratore XML<sup>8</sup>.

---

<sup>8</sup> F. Ciotti, "La rappresentazione digitale del testo: il paradigma del markup e i suoi sviluppi". In L. Perilli, D. Fiormonte (a cura di), *La macchina nel tempo: studi di informatica umanistica in onore di Tito Orlandi*, Firenze, Le Lettere, 2011



## **Nuove frontiere per le Digital Humanities**

Se questo è il quadro, quali sono le prospettive che si possono aprire per lo sviluppo dell'informatica umanistica e di quella letteraria in particolare? Senza dubbio, consolidare e se possibile estendere i risultati acquisiti è una missione validissima e anzi irrinunciabile. Gli archivi testuali vanno preservati ed implementati, la trascrizioni ed edizioni digitali basate sui formalismi attualmente disponibili moltiplicate, gli standard mantenuti, applicati e diffusi. Ma è giunto il momento ormai di individuare nuove linee di ricerca, di esplorare le tendenze innovative che potrebbero fornire un ulteriore salto di qualità e una più ampia giustificazione scientifica (ma anche istituzionale) all'incontro tra informatica e studi umanistici. Difficile dire a priori quali direzioni saranno le più proficue: il tempo lo dirà. Tra i numerosi campi di indagine aperti, ne segnaleremmo almeno due:

- 1) *Big Data*: lo sviluppo e l'applicazione di strumenti per l'analisi automatica delle ingenti masse di risorse testuali/documentali e di dati disponibili sulla rete e non solo, attraverso la sperimentazione di metodologie e tecnologie di *text mining* e *knowledge extraction*.
- 2) *Web 3.0*: la sperimentazione dei nuovi linguaggi e modelli di dati per la rappresentazione dei livelli semantici nelle risorse informative, delle tecnologie e delle architetture che vanno rubricate sotto le etichette di Web Semantico e Linked Data, ovviamente adattandole alle specificità degli oggetti culturali.

Per quanto riguarda la prima linea di ricerca, la cui analisi esula dagli obiettivi di questo lavoro, ci limitiamo a ricordare che si tratta di un ambito di ricerca complesso ma promettente che consiste nella applicazione di tecniche e strategie di *data mining*, ovvero di metodi computazionali, su base statistico/probabilistica, per la ricerca di regolarità e schemi ricorrenti impliciti e non osservabili a priori all'interno di grandi moli di dati strutturati e non strutturati:

Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques<sup>9</sup>.

La ricerca di tali pattern e regolarità si basa su complessi algoritmi probabilistici, i più noti dei quali sono fondati sull'analisi probabilistica bayesiana, che studia la probabilità di eventi non quantificabili a priori (ad esempio la probabilità che in un insieme di soggetti prevalga una certa aspettativa, o che un testo sia categorizzabile in base a un dato tema prevalente). Quando questi algoritmi sono applicati a dati testuali si parla più specificamente di *text mining*. In questa direzione si sono indirizzati alcuni importanti progetti di ricerca nell'ambito delle Digital Humanities, tra cui ricordiamo un importante progetto internazionale diretto da John Unsworth, il *Monk Project*<sup>10</sup>, e le ricerche condotte presso lo *Stanford*

---

<sup>9</sup> Gartner Group, "Data Mining Definition | Gartner", 2012, <http://www.gartner.com/it-glossary/data-mining>.

<sup>10</sup> J. Unsworth, M. Mueller, *The MONK Project Final Report*, Settembre 2, 2009. <http://monkproject.org/MONKProjectFinalReport.pdf>.

*Literary Lab* diretto da Franco Moretti<sup>11</sup>. Lo stesso Moretti ha teorizzato come queste tecnologie di ricerca possano essere il fondamento di un vero e proprio nuovo metodo di studio dei fenomeni letterari, che ha definito *distant reading* (giocando sulla opposizione con il *close reading* introdotto nella critica letteraria dal *New criticism*)<sup>12</sup>.

## Il Web Semantico, le ontologie e i Linked Data

In questa sede intendiamo piuttosto approfondire il discorso sulle potenzialità delle tecnologie del *Web Semantico* (WS) nelle Digital Humanities in generale e negli studi letterari in particolare. Il termine e la visione a cui esso allude sono state proposte da Tim Berners-Lee, l'inventore del Web, nel 1998<sup>13</sup>. L'idea consiste nell'associare alle risorse informative sul Web una descrizione formalizzata del loro significato intensionale mediante la sovrapp-

---

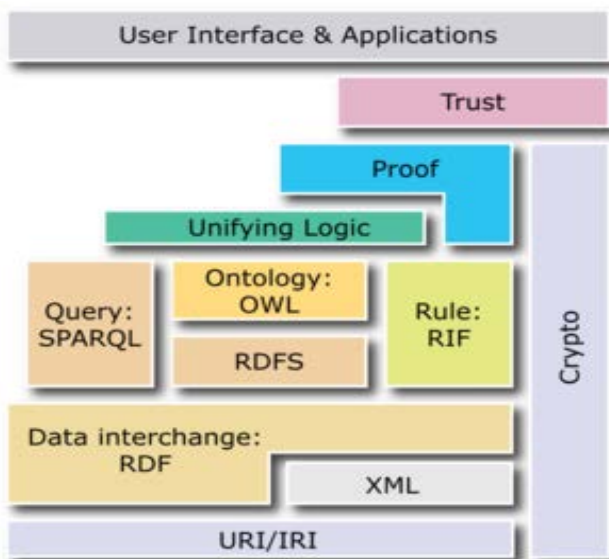
<sup>11</sup> R. Heuser, L. Le-Khac, Stanford Literary Lab, *A Quantitative literary history of 2,958 nineteenth-century British novels: the Semantic cohort method*, 2012. <http://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.

<sup>12</sup> Si vedano: F. Moretti, *Graphs, maps, trees: abstract models for a literary history*, London; New York, Verso, 2005; M. G. Kirschenbaum, "The remaking of reading: Data mining and the digital humanities". In *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*, Baltimore, MD, 2007. <http://www.cs.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>

<sup>13</sup> T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web". In *Scientific American*, vol. 284, fasc. 5, maggio 2001, pp. 34–43.; G. Antoniou, F. Van Harmelen, *A semantic Web primer*, Cambridge, Mass., MIT Press, 2008; E. Della Valle, I. Celino, D. Cerizza, *Semantic web: modellare e condividere per innovare*, Milano, Pearson Addison Wesley, 2008.

posizione di uno o più livelli di *metadati semantici*. Tali insiemi di metadati semantici sono espressi in formalismi che si collocano nella famiglia dei sistemi di rappresentazione della conoscenza a suo tempo sviluppati nell'ambito dell'intelligenza artificiale, e dunque possono essere elaborati automaticamente a diversi livelli di complessità: si va dalla semplice visualizzazione o consultazione per scorrimento di indici strutturati; alla interrogazione e ricerca supportata da motori inferenziali; alla classificazione e collegamento automatico; fino alla derivazione di nuove conoscenze implicite mediante inferenza logica e alla valutazione di attendibilità mediante il computo logico di asserti di fiducia.

Figura 1.



L'architettura generale del Web Semantico è comunemente raffigurata mediante un diagramma a pila che ne specifica le componenti a diversi livelli di astrazione, a partire dagli oggetti informativi a cui si applica, tecnicamente denominate *risorse*<sup>14</sup>. Il significato di questo termine è assai ampio: una risorsa informativa può essere un oggetto informativo accessibile sul Web – dal singolo documento, a sue parti, a collezioni di documenti – un oggetto reale o un oggetto astratto. Il primo problema che si pone è quello della 'identificazione' delle risorse in modo non ambiguo e indipendente dall'universo del discorso. Le URI (*Uniform Resource Identifiers*) sono i formalismi che svolgono tale ruolo: degli identificativi univoci e persistenti che permettono che una data risorsa possa essere menzionata e individuata nello spazio informativo del Web<sup>15</sup>.

Se una risorsa è identificata in modo univoco è possibile esprimere su di essa asserti che ne descrivono il contenuto sotto un qualche rispetto, esprimono ciò che un utente pensa su tale contenuto, ne specificano proprietà e relazioni. Questi asserti sono i *metadati semantici*. Affinché i metadati semantici siano utilizzabili dai

---

<sup>14</sup> In effetti al momento sono stati definiti modelli architetture e linguaggi solo fino al livello delle ontologie e dei sistemi a regole. Gli strati più 'alti' del WS sono ancora avvolti da una aura quasi mistica, e molti esperti nutrono forti dubbi che potranno mai essere tradotti in qualcosa di funzionante, almeno sulla scala totalizzante prevista dal disegno di Berners-Lee. Ci sono ad esempi dei seri problemi formali e matematici che impediscono di applicare algoritmi di dimostrazione automatica a questo livello.

<sup>15</sup> La forma più comune di URI son gli indirizzi delle pagine Web (URL) per cui esiste un consolidato protocollo di dereferenziazione, ma non sono le uniche (e peraltro la loro funzione è spuria in quanto svolgono sia il ruolo di identificatori sia quello di localizzatori).

computer, è necessario che vengano espressi in un linguaggio che sia computazionalmente trattabile. È questo il fine del *Resource Description Framework* (RDF) sviluppato presso *World Wide Web Consortium*<sup>16</sup>. RDF è un metalinguaggio dichiarativo per formalizzare asserti che esprimono proprietà e relazioni tra risorse, il cui modello di dati è basato su tre elementi:

- Risorse.
- Proprietà.
- Asserti.

Le risorse come abbiamo visto sono tutto ciò che può essere soggetto di descrizione: pagine Web, documenti, persone, istituzioni, concetti. Le proprietà sono coppie attributo-valore associate alla risorsa. Ogni proprietà ha un significato specifico, una serie di valori leciti ed è associabile a uno o più tipi di risorsa. Proprietà e valori possono essere espressi da URI (e dunque da altre risorse) o da valori letterali (valori diretti).

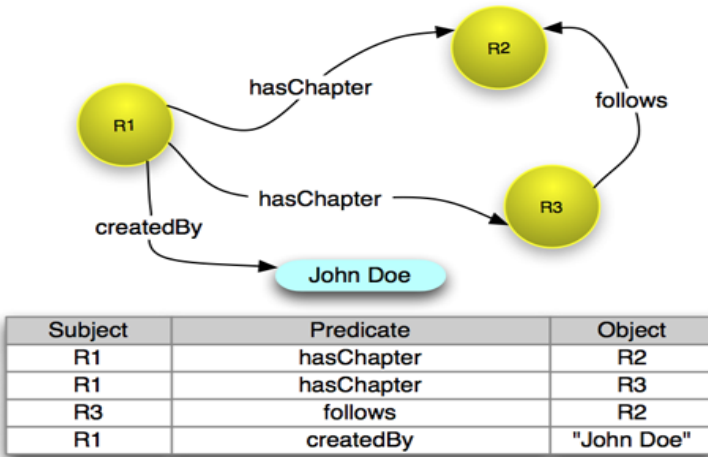
Gli asserti (*statement*) sono la struttura predicativa che esprime l'associazione di una proprietà a una risorsa. Ogni asserto ha una struttura soggetto – predicato – oggetto. Un asserto specifica una

---

<sup>16</sup> Il Web Consortium (W3C, <http://www.w3c.org>) è una organizzazione *no profit* che promuove e coordina lo sviluppo delle tecnologie di base per il Web in modo indipendente dai singoli attori privati a esso interessati, rilasciando standard e linee guida in regime aperto. Le specifiche di RDF sono in *Resource Description Framework (RDF): Concepts and Abstract Syntax*, K. G., Carroll J. (Ed.), W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-concepts>, che definisce il modello dei dati e la sintassi per esprimere asserti RDF.

relazione predicativa tra soggetto e oggetto (in RDF sono consentite solo relazioni binarie). Gli asserti sono anche noti come triple e gli insiemi di asserti si possono rappresentare come *grafi etichettati orientati aciclici*, come si vede nella figura seguente.

Figura 2.



RDF in quanto tale non fornisce un vocabolario predefinito e a priori di proprietà e di relazioni sotto cui sussumere e organizzare le risorse. Si tratta di un modello di dati semplice e rigoroso per specificare proprietà di risorse, qualsivoglia esse siano. In un contesto ampio ed eterogeneo come il Web possono esistere numerosi schemi e vocabolari semantici, basati su diverse concettualizzazioni di particolari domini, su diverse terminologie e lingue. In linea generale si può assumere che esistano anche concettualizzazioni mutuamente contraddittorie e/o mutevoli nel tempo. Al fine di rendere utilizzabili queste concettualizzazioni in modo

computazionale (almeno in parte) è necessario conseguire un ulteriore livello di formalizzazione: quello delle ontologie formali.

La definizione classica di questo concetto è stata fornita da Gruber<sup>17</sup>: "An ontology is an explicit specification of a conceptualization". Il termine ontologia, ereditato dalla metafisica classica dove, sin dalla sistemazione aristotelica, denotava la teoria dell'essere e delle sue categorie, è oggi adottato a denotare una ampia e diversificata classe di oggetti che vanno dai vocabolari controllati, ai thesauri fino alle ontologie formali vere e proprie. Queste, oltre a fissare una terminologia strutturata per gli enti di un dato dominio, ne fissano anche la semantica condivisa da una data comunità, in termini logico-formali:

In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application<sup>18</sup>.

Esistono numerosi linguaggi formali per specificare ontologie formali. A un primo e più semplice livello di complessità e capacità espressiva si pone RDF Schema (RDFS), che permette di definire formalmente:

---

<sup>17</sup> T. R. Gruber, "A translation approach to portable ontology specifications". In Knowledge Acquisition, vol. 5, fasc. 2, 1993, p. 199.

<sup>18</sup> T. R. Gruber, "Ontology", Encyclopedia of Database Systems, Springer-Verlag, 2009.



- Classi (o tipi) di risorse
- Classi di proprietà
- Relazioni tra classi di risorse e proprietà (Ex.: classe -> sottoclasse)
- Domini e range di proprietà

I vincoli espressi da RDFS, tuttavia, non sono sufficienti per esprimere interamente i vincoli ontologici necessari agli obiettivi del WS. Occorre un sistema per specificare le relazioni logico-semantiche (equivalenza, specificazione, generalizzazione, istanziazione, cardinalità, simmetria etc.) tra oggetti e proprietà di un medesimo schema e di schemi diversi. Ad esempio, la relazione di "autorialità" potrebbe essere indicata dalla proprietà "essere autore" dove l'autore sta in funzione di soggetto e il cui oggetto è un dato documento. In uno schema differente, al contrario, potremmo avere che il soggetto è il documento di cui si predica la proprietà "essere scritto da" che ha come oggetto un esponente della classe degli autori. Evidentemente si sta parlando dello stesso insieme di individui e relazioni (dominio), ma in modo simmetrico.

Nel contesto del WS il linguaggio deputato a conseguire questo secondo livello di formalizzazione è *Web Ontology Language* (OWL). OWL, ora giunto alla versione 2.0, può essere espresso in diverse notazioni equivalenti e ha due possibili interpretazioni semantiche: una (OWL 2 DL) basata sulla semantica modellistica di una variante

di *description logic* completa e computabile<sup>19</sup>; e una più potente (OWL 2 Full) basata su un generalizzazione del modello a grafo di RDF/S, ma non decidibile. Nelle logiche descrittive la modellizzazione ontologica è distinta in due parti: la cosiddetta *Tbox* (*terminological box*) descrive le classi e i concetti generali, specifica le loro relazioni e ne definisce proprietà; la *Abox* (*assertion box*) contiene gli asserti fattuali, che elencano gli individui del dominio, ne individuano i ruoli e indicano a quali classi e concetti definiti nella *Tbox* essi appartengano.

Questa struttura semplifica la modellazione concettuale di domini fortemente popolati, poiché gli individui possono essere descritti con una relativa autonomia dalla formalizzazione concettuale di alto livello. Le ontologie formali basate su *description logic* hanno anche il vantaggio di potere essere utilizzate da sistemi di ragionamento automatico abbastanza efficienti (quali il motore inferenziale Racer o il più antico e consolidato linguaggio Prolog), mediante i quali si possono eseguire numerosi processi di elaborazione e gestione delle basi di conoscenza:

Logic reasoning is one possible application for ontologies. It is probably helpful (i) to check consistency during ontology development, (ii) to enable semi-automatic merging of (domain) ontologies as well as (iii) to deduce hidden information contained in the ontology. These three tasks can be applied to all elements of ontologies, classes as well as instances

---

<sup>19</sup> F. Baader, *The description logic handbook: theory, implementation, and applications*, Cambridge, UK; New York, Cambridge University Press, 2003.

[...] logic reasoning can fulfill different purposes in the phase of creating an ontology and in the phase of using it [...], e.g. investigating the structure of categories/ concepts, or testing if every object is used in the intended and not contradictory way. In different situations of a work process, logic reasoning can be used to avoid or to solve problems: If several persons build together an ontology, new included elements can be checked for inconsistency or redundant information can be detected<sup>20</sup>.

Il progetto del WS (o Web 3.0, formulazione adottata dopo e per certi versi in alternativa al diffondersi della moda culturale/ tecnologica del Web 2.0) nella sua generalità richiede numerose e rilevanti innovazioni dal punto di vista tecnico, da quello delle competenze richieste e soprattutto da quello dei comportamenti sociali e culturali degli utenti del Web. Molti esperti e studiosi nutrono forti dubbi sul fatto che tale progetto nella sua versione più ambiziosa e universale potrà mai realizzarsi. Come accennato esistono numerosi e validi ostacoli tecnici e teorici: inconsistenza tra ontologie; incompletezza dei sistemi deduttivi per le versioni più espressive di OWL e RDFs; complessità computazionale degli algoritmi inferenziali applicati a un numero di asserti potenzialmente enorme; criticità della Assunzione di Mondo Aperto, secondo la quale è falso solo ciò che si può dimostrare esplicitamente tale, alla base delle logiche descrittive. Ma forse più rilevanti sono i dubbi circa la sua reale necessità, almeno per gli scopi e gli obiettivi

---

<sup>20</sup> A. Zöllner-Weber, "Ontologies and Logic Reasoning as Tools in Humanities?". In *Digital Humanities Quarterly*, vol. 3, fasc. 4, 2009. <http://www.digitalhumanities.org/dhq/vol/3/4/000068/000068.html>.

per cui è stato ideato e progettato, che sarebbero assai meglio conseguiti dalle tecnologie e dai sistemi di cooperazione decentrata e sociale introdotti dal cosiddetto Web 2.0: si pensi ad esempio al meccanismo delle *folksonomie* (contrapposte alle tassonomie e ai thesauri) e al *social filtering*<sup>21</sup>.

Diverso il discorso relativo all'applicazione di tecnologie del WS a domini specifici e in contesti controllati e locali. In questi contesti vengono meno molte delle problematiche tecniche e vengono valorizzate le capacità di organizzazione delle conoscenze da parte di esperti pur potendo usufruire di strumenti assai più flessibili e dinamici rispetto ai tradizionali strumenti di controllo semantico dell'informazione. In questa direzione si muovono anche le recenti sperimentazioni che vanno sotto l'etichetta di *Linked Data*<sup>22</sup>.

Con questo termine ci si riferisce a un insieme di soluzioni per la pubblicazione e l'interconnessione di dati strutturati sul Web mediante tecnologie del WS. L'idea è stata introdotta ancora da Tim Berners-Lee al fine dare concretezza al progetto del WS rendendo disponibile nei suoi formalismi e protocolli la sterminata quantità di

---

<sup>21</sup> C. Shirky, *Ontology is Overrated: Categories, Links, and Tags*, 2005. [http://www.shirky.com/writings/ontology\\_outrated.html](http://www.shirky.com/writings/ontology_outrated.html); K. H. Veltman, "Towards a Semantic Web for Culture". In *Journal of Digital Information*, vol. 4, fasc. 4, 2004. <http://jodi.tamu.edu/Articles/v04/i04/Veltman>.

<sup>22</sup> T Berners-Lee, "Linked Data - Design Issues", 2006. <http://www.w3.org/DesignIssues/LinkedData.html>; C. Bizer, T. Heath, T. Berners-Lee, "Linked Data - The Story So Far", *International Journal on Semantic Web and Information Systems*, vol. 5, fasc. 3, 33 2009, pp. 1-22.; T. Heath, C. Bizer, "Linked Data: Evolving the Web into a Global Data Space", *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, fasc. 1, febbraio 2011, pp. 1-136.

dati contenuti nei database dei sistemi informativi già ora presenti sul Web. Questo deve avvenire seguendo alcuni semplici principi di base:

The term *Linked Data* refers to a set of best practices for publishing and interlinking structured data on the Web. These best practices were introduced by Tim Berners-Lee in his Web architecture note *Linked Data* and have become known as the *Linked Data principles*. These principles are the following:

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

The basic idea of Linked Data is to apply the general architecture of the World Wide Web to the task of sharing structured data on global scale. In order to understand these Linked Data principles, it is important to understand the architecture of the classic document Web<sup>23</sup>.

La creazione di Linked Data, insomma, costituisce un percorso graduale dall'attuale Web come *rete di documenti* il cui contenuto di conoscenza è prodotto interamente dall'interpretazione umana a un Web come *rete di dati* che veicolano in modo formalizzato frammenti di semantica processabili da un elaboratore. I Linked Data

---

<sup>23</sup> T. Heath, C. Bizer, *op.cit.*

forniscono inoltre una soluzione al problema della identificazione condivisa dei concetti (intesi nel senso più generale possibile), grazie all'adozione condivisa di URI. Un'applicazione esemplare e molto nota di questa architettura è costituita da *GeoNames* ([www.geonames.org](http://www.geonames.org)), una ontologia (qui il termine è usato nel senso ampio di vocabolario formalizzato) che descrive circa 10 milioni di toponimi, organizzati per attributi e relazioni geografico/territoriali. Un altro *repository* di Linked Data assai noto è *DBPedia* ([www.dbpedia.org](http://www.dbpedia.org)) che costituisce una riformulazione formalizzata di parte del contenuto della nota enciclopedia collaborativa Wikipedia<sup>24</sup>.

## **Studi letterari e tecnologie semantiche: verso il Web Semantico Letterario**

Dato il contesto teorico e tecnico che abbiamo delineato sopra, quale convergenza ci può essere tra le tecnologie del Web Semantico e gli studi umanistici e letterari in particolare? Ha un senso parlare di un *Web Semantico Letterario*? La risposta positiva a questo quesito la forniscono implicitamente le già numerose sperimentazioni e progetti di ricerca in atto, o in via di elaborazione, in questa area. Non ci soffermeremo in questa sede su una descrizione analitica di ciascuna di esse. Intendiamo piuttosto fornire una visione a larga scala, tracciare tendenze per future esplorazioni e sperimentazioni.

---

<sup>24</sup> C. Bizer et al., "DBpedia - A crystallization point for the Web of Data", *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, fasc. 3, settembre 2009, pp. 154-165.

Un primo punto di convergenza consiste nella creazione e pubblicazione di Linked Data Set in ambito umanistico e letterario. A un livello molto generalista questa attività ovviamente si potrebbe avvalere di convergenze interdisciplinari anche all'interno dello stesso campo umanistico. Si pensi infatti alla possibilità di predisporre ontologie terminologiche (sullo stile di GeoNames) per:

- luoghi e spazi geografici di epoche e culture diverse;
- persone ed eventi storici;
- autori e opere;
- luoghi e spazi geografici immaginari e funzionali;
- entità e personaggi finzionali;
- temi e motivi letterari;
- temi e motivi iconologici;
- figure retoriche;
- generi e stili letterari e artistici;

Questa lista potrebbe allungarsi a piacimento, ma appare evidente come la creazione di questi repertori sotto forma di Linked Data, organizzati da un livello soggiacente di ontologie formali (a loro volta capaci di interconnettersi ad altre ontologie generali come il CIDOC CRM - <http://www.cidoc-crm.org>) permetterebbe di stabilire relazioni, interconnessioni e *mash-up* tra concetti e nozioni di diversi domini; la loro esplorazione mediante sistemi di ragionamento e dedizione automatici porterebbe a individuare e studiare in modo sistematico fenomeni altrimenti invisibili o solo

intuibili. Peraltro già la stessa costruzione di una ontologia, tanto per fare un esempio, di temi e motivi letterari richiederebbe uno sforzo di analisi concettuale che porterebbe *ipso facto* un importante contributo agli studi di critica tematica.

Tuttavia predisporre repertori, ancorché strutturati *sub specie* di ontologie e Linked Data sarebbe una piccola rivoluzione rispetto alla disponibilità dei corrispettivi repertori cartacei<sup>25</sup>. Così come costruire *repositories* e biblioteche digitali di intere tradizioni letterarie in codifica XML/TEI non ha *finora* portato a grandi passi avanti nella conoscenza critico-letteraria.

Un vero e proprio "salto gestaltico" si può invece ottenere mettendo in relazione ontologie/set di Linked Data e collezioni digitali di testi. Le ontologie possono infatti fornire una semantica computabile per i documenti digitali. Questa integrazione aprirebbe la strada a indagini e ricerche su vasta scala decisamente innovative. Si pensi alla possibilità di incrociare in un dato insieme testuale e in modo esaustivo le interazioni tra determinate forme metriche e figure o tropi; oppure alla possibilità di analizzare in modo sistematico come un tema si fenomenizzi e migri tra testi di varie epoche; o ancora studiare l'evoluzione intertestuale di un personaggio e la sua relazione con generi e temi.

La realizzazione di una simile macro architettura è ovviamente assai complessa tecnicamente e altrettanto onerosa in termini di

---

<sup>25</sup> Si pensi a opere come: R. Ceserani, M. Domenichelli, P. Fasano, *Dizionario dei temi letterari*, Torino, UTET, 2007; A. Ferrari, *Dizionario dei luoghi letterari immaginari*, Torino, UTET, 2007.



tempo e risorse (e tuttavia, se venti anni fa qualcuno avesse detto che oggi avremmo avuto a disposizione in formato digitale, opportunamente codificate in TEI, intere tradizioni letterarie, non si sarebbe forse manifestato il medesimo scetticismo?).

Per le problematiche tecniche, una soluzione consiste nell'adozione di una strategia di annotazione semantica *multilayer* dei documenti digitali basata sul paradigma dello *stand-off markup*, invece del tradizionale e notoriamente problematico approccio di *inline markup*. La segmentazione del testo al fine di codificare i frammenti linguistico-testuali che fenomenizzano i molteplici livelli semantici incorrerebbe inevitabilmente nel problema delle "gerarchie sovrapposte" che caratterizza XML<sup>26</sup>.

Nello *stand-off markup* i marcatori sono in parte o in tutto esterni rispetto alla sequenza lineare dei caratteri, eliminando alla radice la questione delle sovrapposizioni sintattiche. Naturalmente si pone il problema di esprimere formalmente il collegamento tra i metadati semantici esterni (di norma triple RDF o asserti in OWL) e i brani di testo a cui sono applicati, preservando la portabilità e la (teorica) leggibilità umana dell'insieme di documenti digitali risultante. Una elegante soluzione interamente basata su tecnologie standard W3C è stata proposta da Di Iorio e Vitali nel formalismo ontologico EARMARK<sup>27</sup>.

---

<sup>26</sup> Su questo rimandiamo a F. Ciotti, *op.cit.* e alla bibliografia ivi contenuta.

<sup>27</sup> A. Di Iorio, S. Peroni, F. Vitali, "A Semantic Web approach to everyday overlapping markup". In *J. Am. Soc. Inf. Sci. Journal of the American Society for Information Science and Technology*, vol. 62, fasc. 9, 2011, pp. 1696–1716; A. Di Iorio, S. Peroni, F. Vitali,

The basic idea is to model EARMARK documents as collections of addressable text fragments, and to associate such text content with OWL assertions that describe structural features as well as semantic properties of (parts of) that content. As a result EARMARK allows not only documents with single hierarchies (as with XML) but also multiple overlapping hierarchies where the textual content within the markup items belongs to some hierarchies but not to others. Moreover EAMARK makes it possible to add semantic annotations to the content through assertions that may overlap with existing ones<sup>28</sup>.

I frammenti di testo sono identificati come range di caratteri oppure mediante puntatori XPath e XPointer. I tre autori hanno anche sviluppato diversi strumenti per applicare il modello formale, il quale è comunque utilizzabile con qualsiasi sistema di interrogazione e ragionamento compatibile con OWL e SPARQL<sup>29</sup>.

Un'ulteriore applicazione delle tecnologie del Web Semantico nello studio dei fenomeni testuali e letterari consiste nella modellizzazione ontologica delle strutture narrative. Come noto si tratta di uno dei pochi settori delle scienze del testo soggetto già in passato e indipendentemente dall'informatica, a numerosi tentativi

---

"Handling Markup Overlaps Using OWL". In P. Cimiano, H. Pinto (a cura di), *Knowledge Engineering and Management by the Masses*. «Lecture Notes in Computer Science», vol. 6317, Springer Berlin Heidelberg, 2010, pp. 391–400 [http://dx.doi.org/10.1007/978-3-642-16438-5\\_29](http://dx.doi.org/10.1007/978-3-642-16438-5_29).

<sup>28</sup> A. Di Iorio, S. Peroni, F. Vitali, *op.cit.*, p. 1704.

<sup>29</sup> SPARQL (*Simple Protocol and RDF Query Language*) è un linguaggio di interrogazione per basi dati semantiche espresse in RDF sviluppato dal W3C.

di formalizzazione più o meno compiuta, sin dalle origini della narratologia negli anni '60 del secolo scorso. I modelli a suo tempo elaborati da Bremond, da Todorov e poi da Greimas<sup>30</sup> cercavano di fornire una vera e propria grammatica e semantica formale della narrazione, sul modello delle grammatiche generative della lingua a suo tempo proposte da Chomsky. Successivamente diversi autori hanno ripreso ed esteso le prime formalizzazioni inserendole nel contesto più generale della linguistica testuale (ricordiamo in particolare J. Petöfi, T. Van Dijk, W. Dressler, R. de Beaugrande<sup>31</sup>) o degli studi sui mondi possibili narrativi e sulla teoria logica dell'azione e le loro applicazione nell'intelligenza artificiale<sup>32</sup>. Una menzione speciale, visto il contesto, va fatta ai lavori di Giuseppe Gigliozzi, che negli anni '80 e '90 del secolo scorso sperimentò e realizzò diversi modelli computazionali dei fenomeni narrativi, come la creazione di *story grammar* per fiabe e novelle e la descrizione formale dei personaggi e dei ruoli narrativi, realizzando anche diversi programmi basati sul linguaggio Lisp, come SEB e SEBNET<sup>33</sup>.

Questa ricca tradizione rappresenta una solida base teorica su cui innestare le più recenti e flessibili (nonché supportate dal punto

---

<sup>30</sup> C. Brémond, *Logica del racconto*, Milano, Bompiani, 1977; Algirdas Julien Greimas, *La semantica strutturale; ricerca di metodo*, Milano, Rizzoli Editore, 1968.

<sup>31</sup> T. A. van Dijk, *Text and context*, 1977; W. U. Dressler, *Current trends in textlinguistics*, Berlin; New York, W. de Gruyter, 1978.

<sup>32</sup> M.-L. Ryan, *Possible worlds, artificial intelligence, and narrative theory*, Bloomington, Indiana University Press, 1991; J. C. Meister, *Computing action: a narratological approach*, Berlin, Walter de Gruyter, 2003.

<sup>33</sup> Si vedano G. Gigliozzi, *Studi di codifica e trattamento automatico di testi*, Roma, Bulzoni, 1987; G. Gigliozzi, *Saggi di informatica umanistica*, Milano, UNICOPLI, 2008.

di viste delle implementazione software) tecnologie del Web Semantico. Si muove in questa direzione l'ontologia OWL per la descrizione formale dei personaggi letterari proposta da Zöllner-Weber; da ricordare in questa ambito anche un progetto di ricerca recentemente avviato presso il Language Technology Lab del DFKI (Deutsche Forschungszentrum für Künstliche Intelligenz) che ha come obiettivo l'individuazione e il riconoscimento dei personaggi più rilevanti nei racconti popolari attraverso l'uso combinato di ontologie e sistemi di elaborazione del linguaggio naturale. Ovviamente tali ontologie di aspetti e fenomeni intratestuali possono essere integrate con quelle intertestuali che abbiamo discusso sopra, e a loro volta connesse direttamente con i testi, fino a formare un *Semantic Web Letterario* che apre nuove prospettive per l'avanzamento del sapere teorico e critico sul patrimonio letterario.

Resta da considerare un ultimo ma non marginale aspetto: chi e con quali risorse economiche e di tempo potrebbe procedere alla costruzione di questo nostro visionario Semantic Web Letterario? La risposta è: tutta la comunità degli studi letterari. La storia e l'evoluzione del Web ha dimostrato che non solo è possibile costruire sistemi, anche di enorme complessità, attraverso un processo pubblico incrementale e cooperativo, ma che tale strategia si dimostra assai più efficiente ed efficace di quelle private, monolitiche e centralizzate. Moltissimo lavoro nella costruzione della macroarchitettura che qui proponiamo potrebbe essere condotto usando sistemi di cosiddetto *crowdsourcing* guidato, posto che esistano le opportune infrastrutture abilitanti. Il modello del *social tagging*, opportunamente corretto mediante sistemi basati su ontologie che ne

orientino e controllino l'applicazione, permetterebbe di coinvolgere studiosi esperti ma anche giovani ricercatori e cultori nel costruire e popolare le ontologie. Un simile sforzo intellettuale e tecnologico non potrebbe che essere condotto in questo modo. E il prodotto di un tale impresa non potrà che essere un bene comune, un contenuto aperto e disponibile per tutta la comunità degli studi.

## Riferimenti Bibliografici

- Antoniou, Grigoris, Frank Van Harmelen, *A semantic Web primer*, Cambridge, Mass., MIT Press, 2008.
- Baader, Franz., *The description logic handbook: theory, implementation, and applications*, Cambridge, UK; New York, Cambridge University Press, 2003.
- Barnard, David T., Lou Burnard, Jean-Pierre Gaspard, Lynne A. Price, Michael Sperberg-McQueen, Giovanni Battista Varile, "Hierarchical Encoding of Text: Technical Problems and SGML Solutions". In *Computers and the Humanities*, vol. 29, fasc. 3, 1995, pp. 211–231.
- Berners-Lee, Tim, "Linked Data - Design Issues", 2006. <http://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, Tim, James Hendler, Ora Lassila, "The Semantic Web". In *Scientific American*, vol. 284, fasc. 5, maggio 2001, pp. 34–43.
- Bizer, Christian, Tom Heath, Tim Berners-Lee, "Linked Data - The Story So Far". In *International Journal on Semantic Web and Information Systems*, vol. 5, fasc. 3, 33 2009, pp. 1–22.
- Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann, "DBpedia - A crystallization point for the Web of Data". In *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, fasc. 3, settembre 2009, pp. 154–165.
- Brémond, Claude, *Logica del racconto*, Milano, Bompiani, 1977.
- Buzzetti, Dino, "Digital Representation and the Text Model". In *New Literary History*, vol. 33, fasc. 1, S.d., pp. 61–88.

- Ceserani, Remo, Mario Domenichelli, Pino Fasano, *Dizionario dei temi letterari*, Torino, UTET, 2007.
- Ciotti, Fabio, "La rappresentazione digitale del testo: il paradigma del markup e i suoi sviluppi", Lorenzo Perilli, Domenico Fiorimonte (a cura di). In *La macchina nel tempo: studi di informatica umanistica in onore di Tito Orlandi*, Firenze, Le lettere, 2011.
- Cohen, Patricia, "Digital Keys for Unlocking the Humanities' Riches", *New York Times*, novembre 16, 2010.  
[http://www.nytimes.com/2010/11/17/arts/17digital.html?\\_r=2&emc=eta1&](http://www.nytimes.com/2010/11/17/arts/17digital.html?_r=2&emc=eta1&).
- Culler, Jonathan D., Francesco Muzzioli, Gian Paolo Castelli, *Teoria della letteratura : una breve introduzione*, Roma, Armando, 1999.
- Della Valle, Emanuele, Irene Celino, Dario Cerizza, *Semantic web : modellare e condividere per innovare*, Milano, Pearson Addison Wesley, 2008.
- Dijk, Teun A. van, *Text and context*, 1977.
- Dressler, Wolfgang U., *Current trends in textlinguistics*, Berlin; New York, W. de Gruyter, 1978.
- Ferrari, Anna, *Dizionario dei luoghi letterari immaginari*, Torino, UTET, 2007.
- Gartner Group, "Data Mining Definition | Gartner", 2012.  
<http://www.gartner.com/it-glossary/data-mining>.
- Gigliozzi, Giuseppe, *Saggi di informatica umanistica*, Milano, UNICOPLI, 2008.

F. Ciotti, *Web semantico, linked data e studi letterari*

— — —, *Studi di codifica e trattamento automatico di testi*, Roma, Bulzoni, 1987.

Greimas, Algirdas Julien, *La semantica strutturale; ricerca di metodo*, Milano, Rizzoli Editore, 1968.

Gruber, Thomas R., "A translation approach to portable ontology specifications", *Knowledge Acquisition*, vol. 5, fasc. 2, 1993, pp. 199–220.

— — —, "Ontology", *Encyclopedia of Database Systems*, Springer-Verlag, 2009.

Heath, Tom, Christian Bizer, "Linked Data: Evolving the Web into a Global Data Space". In *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, fasc. 1, febbraio 2011, pp. 1–136.

Heuser, Ryan, Long Le-Khac, Stanford Literary Lab, *A Quantitative literary history of 2,958 nineteenth-century British novels: the Semantic cohort method*, 2012. <http://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.

Di Iorio, Angelo, Silvio Peroni, Fabio Vitali, "A Semantic Web approach to everyday overlapping markup". In *J. Am. Soc. Inf. Sci. Journal of the American Society for Information Science and Technology*, vol. 62, fasc. 9, 2011, pp. 1696–1716.

— — —, "Handling Markup Overlaps Using OWL". In P. Cimiano, H. Pinto (a cura di), *Knowledge Engineering and Management by the Masses*. «Lecture Notes in Computer Science», vol. 6317, Springer Berlin Heidelberg, 2010, pp. 391–400 [http://dx.doi.org/10.1007/978-3-642-16438-5\\_29](http://dx.doi.org/10.1007/978-3-642-16438-5_29).

Kirschenbaum, M. G., "The remaking of reading: Data mining and the digital humanities". In *The National Science Foundation*



*Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD, 2007.* <http://www.cs.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>.

Koleva, Nikolina, Thierry Declerck, Hans-Ulrich Krieger, "An Ontology-Based Iterative Text Processing Strategy for Detecting and Recognizing Characters in Folktales", Jan Christoph Meister (a cura di). In *Digital Humanities 2012 Conference Abstracts*, 467–470, Hamburg, Hamburg University Press, 2012.

Landow, George P., *Hypertext: the convergence of contemporary critical theory and technology*, Baltimore, Johns Hopkins University Press, 1992.

McCarty, Willard, *Humanities computing*, Basingstoke [England]; New York, Palgrave Macmillan, 2005.

— — —, "Literary enquiry and experimental method: What has happened? What might?", in Liborio Dibattista, *Storia della Scienza e Linguistica Computazionale: Sconfinamenti Possibili*, 32–54, Milano, Franco Angeli, 2009. <http://www.mccarty.org.uk/essays/McCarty,%20Literary%20enquiry%20and%20experimental%20method.pdf>.

McGann, Jerome, "Marking Texts of Many Dimensions", Susan Schreibman, Ray Siemens, John Unsworth (a cura di), *A Companion to Digital Humanities*, 198–217, Oxford, Blackwell, 2004. <http://www.digitalhumanities.org/companion>.

Meister, Jan Christoph, *Computing action: a narratological approach*, Berlin, Walter de Gruyter, 2003.

F. Ciotti, *Web semantico, linked data e studi letterari*

Moretti, Franco, *Graphs, maps, trees: abstract models for a literary history*, London; New York, Verso, 2005.

Ryan, Marie-Laure, *Possible worlds, artificial intelligence, and narrative theory*, Bloomington, Indiana University Press, 1991.

Schreibman, Susan, Raymond George Siemens, John Unsworth, *A companion to digital humanities*, Malden, Mass., Blackwell Pub., 2004.

Shirky, Clay, *Ontology is Overrated: Categories, Links, and Tags*, 2005.