

GLODIUM

Glossario di Informatica Umanistica

ANTONELLO FABIO CATERINO, MARCELLO BOLPAGNI, MARCO
PETOLICCHIO

draft : : 15 marzo 2019: : 12:54

Indice

1	Prefazione	1
	GLODIUM – Glossario di Informatica Umanistica. Cui prodest?	1
2	Archivio	1
	Bibliografia consultata	3
	Bibliografia consigliata	3
	Sitografia	3
3	ASCII	5
	Bibliografia consultata	5
4	Beta code	7
	Bibliografia consultata	8
	Sitografia	8
5	Biblioteca digitale	11
	Bibliografia consultata	13
	Bibliografia consigliata	14
	Sitografia	14
6	Big data	17
7	Browser	21
	Sitografia	22
8	Concordanze	23
	Bibliografia consigliata	24
	Sitografia	24
9	Corpus	27
10	Crowdsourcing	29
	Bibliografia consigliata	31
	Sitografia	31

11 CSS	33
Bibliografia consigliata	34
Sitografia	34
12 Data visualization	35
Bibliografia consigliata	36
Sitografia	36
13 Database	39
14 Digital Dark Age	41
Bibliografia consigliata	42
Sitografia	42
15 Digital Death	43
16 DOI	45
Bibliografia consigliata	45
Sitografia	46
17 DTD	47
Bibliografia consigliata	47
18 EAD	49
Bibliografia consigliata	50
Sitografia	50
19 Feed RSS	51
Sitografia	52
20 Filologia digitale	53
Bibliografia consigliata	55
Sitografia	55
21 FTP	57
Sitografia	58
21.1 Sitografia {-}:	58
21.2 Sitografia {-}:	58
Bibliografia consigliata	60
Sitografia	61
22 HTML	63
Bibliografia consigliata	64
Sitografia	65
23 Informatica umanistica	67
Bibliografia consultata	68
Bibliografia consigliata	68

Sitografia	69
24 Interfaccia utente (UI)	71
Sitografia	71
25 Machine learning	73
Bibliografia consigliata	75
Sitografia	76
26 Markup	79
27 Metadato	81
Bibliografia consigliata	84
Sitografia	84
28 N-gramma	87
Bibliografia consigliata	88
Sitografia	89
29 Oggetto digitale	91
Bibliografia consigliata	92
Sitografia	92
30 Ontologia	93
31 OPAC	95
Bibliografia consigliata	96
Sitografia	97
32 Open Access	99
Bibliografia consigliata	100
Sitografia	100
33 Paratesto	101
Bibliografia consigliata	103
Sitografia	103
34 PHP	105
35 Public History	107
36 Self-archiving	109
Bibliografia consigliata	111
Sitografia	111
37 SGML	113
Bibliografia consigliata	114
Sitografia	115

38 Tassonomia	117
Bibliografia consigliata	118
Sitografia	118
39 TEI	121
Sitografia	124
40 Thesaurus	125
Bibliografia consigliata	126
Sitografia	127
41 Web Semantico	129
Bibliografia consigliata	130
Sitografia	130
42 XSL	131
Bibliografia consigliata	132
Sitografia	132
43 XML	133
Bibliografia consigliata	134
Sitografia	134

Elenco delle tabelle

Elenco delle figure

GLODIUM – Glossario di Informatica Umanistica. Cui prodest?

Le grandi quantità di dati fanno paura a molte discipline: informatica, economia, matematica, statistica, filosofia. Perché non dovrebbero terrorizzare anche noi umanisti?

Ci avviciniamo con cautela a un futuro pieno di informazioni immediatamente disponibili, lasciandoci sempre più alle spalle un passato – che idealizziamo, senza troppa memoria storica – in cui i dati a disposizione erano davvero pochi, chiusi per di più nel cassetto di chi decideva di volta in volta come usarli e se concederne l'usufrutto.

Le paure si affrontano affrontando di volta in volta le situazioni che generano in noi questo senso di perdizione: non certo evitandole; qualsiasi esperto in psiche umana potrà comodamente dirci, in caso contrario, che non c'è niente di peggio dai *loop* derivati dalla paura della paura. Sembra proprio che la diffidenza verso l'informatica umanistica sia uno di questi *loop*: temiamo i dati, la loro sconvolgente capacità di confonderci, e dunque evitiamo di cimentarci con essi, rinunciando così sin troppe volte alle sconvolgenti opportunità di conoscere e migliorare che essi ci offrono. È facile gestire poche informazioni (e coronarsi maestri d'esse), ma è davvero una blasfemia ermeneutica rinunciare ad aumentare la conoscenza generale di una disciplina per paura di non riuscire a gestirne numero maggiore.

Eppure non abbiamo bisogno dell'informatica umanistica solo per orientarci nella selva dei *big data* che le nuove tecnologie – *obtorto collo* – ci offrono: ne abbiamo anche necessità per andare oltre i difetti convenzionali dell'isolazionismo disciplinare. I vari settori scientifico-disciplinari sono – nella sezione del sapere che convenzionalmente chiamiamo umanistica – sin troppo isolati gli uni dagli altri, forse perché è meno sentito tra gli umanisti il concetto di squadra (che rivive – ad esempio – in un laboratorio

scientifico). Tramite l'informatica sarebbe, invece, possibile creare piattaforme collaborative in cui ognuno è chiamato a svolgere una mansione, nell'ottica della ricerca generale; ciascuno secondo le proprie competenze e abilità di partenza, che mai e poi mai dovranno essere snaturate da ingenui principi di «tuttologia».

Orientarsi tra i *big data*, gestire più informazioni contemporaneamente, creare filtri sempre più precisi per i vari risultati delle ricerche, offrire piattaforme e modelli di collaborazione ai ricercatori: a fronte di tutti questi vantaggi, sembra ancora assurdo che manipoli di intellettuali duri e puri continui a tuonare contro la tecnologia applicata alla scienze umane; come se – tra parentesi – le scienze umane non fossero pioniere stesse del concetto di tecnologia. L'informatica è del resto bastata sul concetto stesso di grammatica, dunque quanto può essere impreciso – per non dire peggio – un umanista che rifugge una disciplina pronipote di una categoria logica di sua stessa competenza?!

Bisogna altresì spendere almeno un paio di parole sui profondi miglioramenti che l'informatica umanistica potrebbe offrire al concetto stesso di didattica. Le stesse piattaforme utili per i ricercatori con qualche modifica potrebbero essere utili per mantenere costanti rapporti tra studenti e docenti, per condividere sul momento materiali di studio, per offrire l'interazione necessaria perché l'istruzione abbia i suoi maggiori frutti, supponendo che il modello didattico più valido sia ancora quello basato su domanda/risposta. Anche da questo punto di vista si assiste, però, a una fiera dell'oscena cecità intellettuale, che culminano di solito in qualche crociata per l'abolizione dell'uso degli *smartphone* nelle scuole, come se prima del telefonino gli studenti non sapessero procacciarsi altrettante distrazioni dal loro percorso.

Queste sono poche considerazioni introduttive, che intendono offrire – a mo' di volo d'aquila – una panoramica generale sui possibili vantaggi di non chiudersi a riccio di fronte all'informatica umanistica. Non si vuole entrare nel particolare, ad esempio nel confronto tra i commenti ai testi editi in era pre-informatica coi commenti stilati con l'ausilio di *corpora* elettronici: sarebbe addirittura imbarazzante continuare a lodare la precisione dei nuovi strumenti. Certo, ci vogliono ricercatori e studiosi capaci di usare le nuove tecnologie senza mai considerarle una scorciatoia. I rovesci della medaglia esistono, del resto, da quando noi umani ci siamo resi conto che nel conio il minor numero di facce è due (pur supponendo che una delle due sia completamente vuota).

Linneo – grande profeta della classificazione tassonomica – si esprime eccellentemente dicendo *nomina si nescis, perit et cognitio rerum*. Ciò vale anche per l'informatica umanistica: non

conoscere i nomi dei concetti di cui si avvale equivale a fare una grandissima confusione.

Per concludere davvero: se l'informatica umanistica può essere utile, può essere ancor più utile avere a disposizione uno strumento di orientamento interno che sappia definire ogni istanza al suo interno, in modo da renderla un campo funzionante e funzionante a qualsiasi prospetto di studio o ricerca in ambito umanistico. Speriamo che il GLODIUM possa dare in tal senso una consistente mano.

GLODIUM nasce come strumento online¹ e come tale è stato inserito dall'EADH tra i progetti più significativi degli ultimi cinque anni² – in europa – nel campo delle *digital humanities*. L'intento della versione cartacea è principalmente manualistico: si vorrebbe offrire agli studenti, ovvero agli studiosi che muovono i primi passi all'interno della disciplina, uno strumento agile alla rapida consultazione, che contenga i una prima scelta dei lemmi più comuni in cui è possibile imbattersi, e per cui occorre una definizione chiara e lineare, senza eccessivi rimandi o appesantimenti.

AntonelloFabioCaterino

¹GLODIUM – Glossario di informatica umanistica. Progetto ideato e diretto da Antonello Fabio Caterino, a cura di Marcello Bolpagni, Antonello Fabio Caterino e Marco Petolicchio. ISBN 9788832173086. Al momento, il progetto è ospitato all'interno del *carnet de recherche Filologia Risorse Informatiche*, <https://fri.hypotheses.org>.

²Si rimanda al sito <https://eadh.org/projects/glodium-glossario-di-informatica-umanistica>.

Il termine *archivio* può avere numerose accezioni: rimandare a un'istituzione; a un'organizzazione; ad un luogo che ospita uno o più complessi documentali; al materiale documentale prodotto da uno specifico soggetto (magistrature, organi e uffici centrali e periferici dello Stato; enti pubblici; istituzioni private, famiglie o persone).

Focalizzando l'attenzione sull'ultima definizione, l'*archivio* è «il complesso dei documenti prodotti e acquisiti da un soggetto nell'esercizio delle sue attività». La natura del soggetto produttore si riflette nella tipologia d'*archivio*, caratterizzata poi da fattori quali l'ordine interno, la rilevanza giuridica, le modalità di formazione e sedimentazione.

Come scrive Monica Grossi: «preme sottolineare come elemento distintivo di qualunque archivio la sua natura organica, il suo essere costituito da più oggetti (i documenti) legati tra loro da relazioni complesse (espresse nel vincolo archivistico che lega le singole parti al tutto) e fondamentali per la piena comprensione di questo bene culturale vulnerabile e multiforme, crocevia di interessi pratici, giuridici e storici di una comunità eterogenea e diacronica».

Il trattamento e la cura degli archivi non possono inoltre prescindere dalle recenti modalità di produzione documentaria che con l'introduzione dei documenti digitali, che si è andata affiancando ai documenti cosiddetti analogici, ha modificato dai punti di vista amministrativo, normativo, tecnologico e culturale i rapporti tra supporto e testo del documento, tra forma e contenuto, oltre a incidere sulle relazioni tra i documenti e i loro autori e i processi di formazione, gestione e conservazione degli archivi.

Indipendentemente dal formato e dal supporto, comunque, la nostra teoria archivistica distingue tre fasi nel ciclo vitale dell'*archivio*: corrente, deposito e storico, cui corrispondono sul piano operativo ben precise e distinte attività. Giova sottolineare che questa distinzione ha un valore meramente operativo ma

non serba alcuna rilevanza sul piano teorico: il valore storico, e quindi l'interesse culturale del documento nasce sin dal formarsi degli archivi e coesiste sin dall'inizio con il fine pubblico di garantire la certezza del diritto.

La fase corrente è quella in cui l'*archivio* si forma e si organizza attraverso la produzione e la ricezione dei documenti e la formazione delle unità archivistiche. In questo stadio si perseguono fini di efficienza, garanzia del diritto e trasparenza amministrativa ma si impostano anche i presupposti per la tutela della futura memoria del soggetto produttore. Nell'*archivio* di deposito confluisce la documentazione che ha terminato la sua fase attiva e che, pur non risultando più necessaria all'espletamento dell'attività quotidiana, conserva interesse dal punto di vista operativo e non è ancora pronta ad essere destinata ad un uso prevalentemente culturale. L'*archivio* storico infine accoglie la documentazione relativa ad affari esauriti da almeno trent'anni. Tale documentazione, svuotata dalle esigenze di servizio, è destinata a conservazione permanente e alla consultazione a fini di studio e a fini culturali.

Questo collaudato modello è però oggetto di riflessione: le particolari caratteristiche degli archivi digitali impongono infatti un ripensamento della tradizionale articolazione del ciclo vitale e una nuova e particolare attenzione da destinare alla progettazione dei sistemi documentari e alla definizione dei requisiti descrittivi finalizzati alla gestione, alla selezione e alla conservazione dei documenti.

L'*archivio* inoltre è andato assumendo, sempre più nettamente, nuovi significati: è «luogo» di per sé simbolico di un passato collettivo, insieme di documenti-fonte per la riappropriazione da parte di singoli cittadini di tratti della loro storia individuale o per la rivendicazione di diritti alla trasparenza, all'informazione, alla tutela della privacy.

Gli stessi archivisti hanno sentito la necessità di rendere sempre maggiormente accessibile il patrimonio archivistico, superando la cerchia degli addetti ai lavori e delle funzionalità di mera ricerca scientifica. Tale esigenza ha trovato prime forme di esplicazione nelle mostre documentarie per poi allargarsi ai laboratori didattici, a quelli di scrittura narrativa e a quelli di teatralizzazione documentale, nonché all'istituzione di musei aggregati agli archivi.

Bibliografia consultata

Duranti L., *Il documento archivistico*, in Giuva L.- Guercio M., *Archivistica. Teorie, metodi, pratiche*, Roma, Carocci, 2014, pp. 19-33

Grossi M., *L'archivio in formazione*, in Giuva L.- Guercio M., *Archivistica. Teorie, metodi, pratiche*, Roma, Carocci, 2014, pp. 35-52

Bibliografia consigliata

Carucci P. - Guercio M., *Manuale di archivistica*, Roma, Carocci, 2008

Valacchi F., *Diventare archivisti*, Milano, Editrice Bibliografica, 2015

Sitografia

*Archivio (s.v.), in DGA. Direzione Generale Archivi, <<http://www.archivi.beniculturali.it/index.php/abc-degli-archivi/glossario>>

*Archivio (s.v.), in Lombardia Beni Culturali. Glossario, <<http://www.lombardiabeniculturali.it/archivi/glossario/>>

ICA. International Council on Archives, <<https://www.ica.org/en>>

ConcettaDamiani

(ingl. Acronimo per *American Standard Code for Information Interchange*)

È il primo standard, sviluppato nel 1968, che è ancora oggi l'unica pagina di caratteri che viene interpretata e trasmessa univocamente su qualsiasi sistema informatico. Il suo funzionamento è pressoché semplice, *ASCII* codifica ogni carattere in codice binario in modo da permettere al computer di comprendere ed immagazzinare quelle informazioni.

Alla sua nascita si basava su un sistema a 7 bit, ciò vuol dire che era in grado di codificare $2^7=128$ caratteri: da 0 a 32 e il 127 non sono stampabili, in quanto rappresentano dei caratteri di controllo che permettono di compiere alcune azioni, come l'andare a capo; da 65 a 90 rappresentano le maiuscole; da 97 a 122 sono le minuscole.

Nonostante la sua effettiva utilità *ASCII* si rivelò insufficiente per la circolazione di informazioni in un ambiente condiviso più ampio del sistema nazionale statunitense. Si cercò, quindi, di creare un sistema di codifica, chiamato «*ASCII* esteso», a 8 bit in modo da ampliare la gamma di caratteri codificabili ($2^8=256$). Nemmeno l'aggiunta di un bit nella codifica riuscì a risolvere i problemi causati da un sistema di scambio dell'informazione sempre più complesso.

Per questi motivi, nel 1991, venne sviluppato un nuovo sistema di codifica chiamato Unicode, il cui sistema non è esente da difetti, ma resta comunque lo standard universale con cui bisogna fare i conti nel momento in cui ci si avvicina all'informatica e all'editoria digitale.

Bibliografia consultata

Kreines D. C., *Oracle SQL: The Essential Reference*, USA, O'Reilly Media, 2000, p. 162

Tissoni F., *Lineamenti di Editoria Multimediale*, Milano, Edizioni Unicopli, 2009, p. 61

AlessiaMarini

Con l'espressione *Beta code* si fa riferimento ad un sistema di codifica che usa caratteri ASCII, ideato per la trascrizione di testi greci su piattaforme e browser. Si tratta di un pre-unicode standard e non va confuso con una semplice «romanizzazione» dell'alfabeto greco. Per riprendere l'ottima definizione proposta da Fusi (2017, p. 24), il *Beta code* è un esempio di metacodifica, «che sovrappone ad uno standard di codifica testuale un ulteriore livello di interpretazione»; esso «con tutto il necessario apparato di simboli specialistici e con una essenziale formattazione tipografica, rappresenta un modo per superare le limitazioni del repertorio di ASCII senza definire una codifica numerica completamente arbitraria» (Fusi, *ibid.*).

La prima convenzione di codifica e formattazione dei caratteri greci fu ideata da David W. Packard negli anni «70 e fu definita come 'Alpha code», nome che successivamente fu mutato in *Beta code*. Il *Beta code* fu adottato, nel 1981 dal TLG e – benché ritenuto da alcuni ormai obsoleto (cfr. *e.g.* Borgna-Musso 2017, p. 72, n. 5; ma, in merito ai limiti del *Beta code*, si rimanda anche a Fusi 2017, p. 28) – è utilizzato ancora oggi nell'ambito di diversi progetti, quali il *Perseus Project*, il *Packard Humanities Institute*, il *Duke Databank of Documentary Papyri et alii*, oltre il summenzionato TLG, che dispone di un'apposita «sezione» contenente il *Beta code manual*, nonché una comoda *Quick Reference Guide* scaricabile in formato pdf.

Nel *Beta code* le lettere greche sono rappresentate dai soli caratteri maiuscoli di ASCII; ove possibile, vi è una corrispondenza tra grafemi greci e grafemi romani, con l'ovvia eccezione delle lettere eta, theta, xi, psi, omega e digamma (che corrispondono rispettivamente ai caratteri H, Q, C, Y, W e V). Alcuni caratteri sono re-interpretati come segni diacritici: lo spirito dolce corrisponde a), quello aspro a (; l'accento acuto corrisponde a /, l'accento grave a \, l'accento circonflesso a =; la iota sottoscritta corrisponde al segno |.

È ormai divenuta topica l'esemplificazione dell'uso del *Beta code* attraverso la beta-coded version dei primi versi dell'*Iliade*:

*MH=NIN A)/EIDE, QEA/, *PHLHI+A/DEW *)AXILH=OS
OU)LOME/NHN, H(\ MURI/' *)AXAIOI=S A)/LGE' E)/QHKE

Il *Beta code* può anche essere facilmente convertito in altri formati (ad esempio nel formato Unicode), anche attraverso estensioni gratuite ed open source (cfr. *Greek Beta code converter*).

Bibliografia consultata

Borgna A.- Musso S., *Le sfide di una biblioteca digitale del latino tardoantico. Struttura, canone e questioni aperte di codifica e visualizzazione*, in P. Mastandrea (a cura di), *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, Venezia, Edizioni Ca' Foscari, 2017, pp. 69-94, <http://edizionicafoscari.unive.it/media/pdf/books/978-88-6969-183-6/978-88-6969-183-6_IhKmUbC.pdf>

Fusi D., *Tecnologie informatiche per l'umanista digitale*, Roma, Nuova Cultura, 2017, pp. 24-29

Sitografia

**ASCII* (s.v.), in *GloDIUM*, <<https://fri.hypotheses.org/glossario-di-informatica-umanistica>>

Battaglini G., *Note minime sul TLG: brevi cenni sulle «origini» del TLG e piccolo vademecum*, in *FRI – Filologia Risorse Informatiche (Carnet de recherche and online journal – Italian Digital Humanities)*, <<https://fri.hypotheses.org/1391>>

**Beta code* (s.v.), in *Wikipedia*, <https://en.wikipedia.org/wiki/Beta_Code>

Duke Databank of Documentary Papyri, <<https://papyri.info/docs/ddbdp>>

Greek Beta code converter, <<https://chrome.google.com/webstore/detail/greek-beta-code-converter/abelaepcjppjpkhpbeggnijcccphpnp>>

Packard Humanities Institute (official website), <<https://packhum.org/>>

**Packard Humanities Institute* (s.v.), in *Wikipedia*, <https://en.wikipedia.org/wiki/Packard_Humanities_Institute>

Perseus Project (official website), <<http://www.perseus.tufts.edu/hopper/>>

**Perseus Project* (s.v.), in *Wikipedia*, <https://en.wikipedia.org/wiki/Perseus_Project>

The Beta code manual, in *TLG*, <<http://www.tlg.uci.edu/encoding/>>

**Thesaurus Linguae Graeciae* (s.v.), in *Wikipedia*, <https://en.wikipedia.org/wiki/Thesaurus_Linguae_Graeciae>

TLG Beta code quick Reference Guide, <<https://www.tlg.uci.edu/encoding/quickbeta.pdf>>

TLG (sub-corpus del) in open access: Canon of Greek Authors and Works (Abridged TLG), <<http://stephanus.tlg.uci.edu/Iris/canon/csearch.jsp>>

TLG (ultima versione digitale del *TLG: Thesaurus Linguae Graeciae. Digital Library*. Ed. Maria C. Pantelia, University of California, Irvine), < <http://stephanus.tlg.uci.edu/>>

**Unicode (s.v.)*, in *Wikipedia*, <<https://it.wikipedia.org/wiki/Unicode>>

Giovanna Battaglino

L'espressione *biblioteca digitale* – come sottolinea Faba Pérez (1999) – si configura come un «concetto ambiguo, come la maggior parte dei concetti che si pongono in relazione con la tecnologia e con le informazioni». A lungo, il concetto di *biblioteca digitale* è stato connesso a quello di «biblioteca virtuale» (cfr. Basili-Pettenati 1994).

Ad oggi, non esiste una definizione univoca, giacché molteplici sono gli approcci, muovendo dai quali si è provato a definire una *biblioteca digitale*; fra essi è possibile annoverare, accanto ad un approccio prettamente informatico (che valorizza la dimensione della digitalizzazione), l'approccio seguito dalla biblioteconomia (che valorizza quanto attiene alla fruizione ed alla organizzazione documentaria), nonché numerosi approcci di ricerca, sovente legati a specifici progetti di digitalizzazione.

La prima definizione è stata elaborata dalla Borgman (1993), che definisce la biblioteca digitale come una «struttura», nella quale si combinano ed interagiscono un «servizio», una «architettura di rete», un insieme di «risorse informative» – costituite da banche-dati testuali e numeriche, immagini statiche o dinamiche –, nonché un insieme di strumenti per localizzare e reperire informazioni. La studiosa, in un successivo contributo (Borgman 1999; ma cfr. anche Borgman 2000 e 2003), riprende e precisa tale definizione, descrivendo la *biblioteca digitale* sia come «estensione e potenziamento dei sistemi di conservazione delle informazioni» (in quanto «set» di risorse elettroniche e competenze tecniche, finalizzate alla creazione, alla ricerca e all'uso delle informazioni), sia come «estensione e potenziamento di tutti i luoghi atti a conservare, organizzare e diffondere informazioni», quali musei, biblioteche, archivi, scuole, laboratori. È evidente che la definizione di *biblioteca digitale* rechi con sé, da un punto di vista concettuale, sia un aspetto informativo-digitale/strumentale che un aspetto più propriamente informativo-sociale.

Tra le prime definizioni va annoverata quella elaborata dalla Digital Library Federation (DLF, 1998), ancora chiaramente orien-

tata su una dimensione biblioteconomica: «le biblioteche digitali sono organizzazioni che forniscono risorse, compreso il personale specializzato, per selezionare, organizzare, interpretare, distribuire, preservare l'integrità delle collezioni digitali, assicurandone la persistenza nel tempo, sì che esse possano essere prontamente ed economicamente accessibili per essere usate da una specifica comunità o da una serie di comunità». Tale definizione è stata successivamente (DLF, 2014) ripresa ed ampliata.

Le successive formulazioni definitorie – quali quella di Oppenheim-Smithson (1999) e di Arms (2000) –, al contrario, puntano soprattutto alla valorizzazione della dimensione informativa. Pertanto, la *biblioteca digitale* è concepita sostanzialmente come «sistema informativo», le cui «risorse», disponibili in formato digitale, possono essere acquisite, archiviate, conservate/preservate, recuperate, attraverso il ricorso alle «tecnologie digitali».

Una prima definizione più articolata è quella proposta da Marchionini-Fox (1999), i quali affermano che «il lavoro di una biblioteca digitale si realizza in uno «spazio» determinato dalla convergenza di quattro dimensioni: comunità (di utenti), tecnologia, servizi e contenuti». La dimensione comunitaria poggia sul bisogno tipicamente umano di cercare informazioni per conoscere e per comunicare; la dimensione tecnologica fa riferimento ai progressi tecnologici, che hanno trasformato anche il modo di conservare e condividere le informazioni; per quanto concerne i servizi, i due studiosi fanno riferimento alla facilitazione nel reperire informazioni, anche attraverso l'aiuto in linea; infine, la dimensione contenutistica consta dei documenti cercati, che possono presentarsi in varie forme/media.

Anche in Italia il concetto di *biblioteca digitale* (che ha cominciato ad affermarsi non prima della fine degli anni «90) comincia a divenire noto soprattutto in relazione alla sua dimensione «strutturale»: in Italia, il sintagma è stato introdotto da Malinconico (1998), il quale parla di «nuova infrastruttura dell'informazione».

Successivamente anche il dibattito italiano relativo all'ontologia di una biblioteca digitale si concentra sulla dimensione informativa, come emerge dalla definizione offerta da Salarelli-Tammaro (2000): «la biblioteca digitale è uno spazio informativo in cui le collezioni digitali, i servizi di accesso e le persone interagiscono a supporto del ciclo di creazione, preservazione, uso del documento digitale». Naturalmente, tale «spazio» è caratterizzato da una specifica organizzazione, funzionale alla dimensione informativa, come emerge dall'ampia definizione elaborata da Ciotti-Roncaglia (2002), che descrivono la *biblioteca digitale* come una «collezione

di documenti digitali strutturati (sia prodotti mediante digitalizzazione di originali materiali, sia realizzati *ex novo*), dotati di un'organizzazione complessiva coerente di natura semantica e tematica, che si manifesta mediante un insieme di relazioni interdocumentali e intradocumentali e mediante un adeguato apparato meta-informativo».

Nel 2005, la Tammaro propone una definizione – per così dire – «problematizzata»: la studiosa sottolinea che perché una *biblioteca digitale* sia tale (e sia efficace) non basta la conversione digitale della documentazione cartacea, ma occorre la concomitanza di una serie di elementi, quali la presenza di un preciso punto di accesso alle risorse digitali in rete, la definizione di una chiara finalità di servizio (mission), la dichiarazione di una precisa politica di sviluppo della collezione digitalizzata, l'adeguatezza dell'organizzazione dell'informazione digitale resa disponibile e la presenza di nuovi servizi d'accesso, che facilitino l'utenza nella fase di ricerca.

Pur nella *varietas* e nella problematicità delle definizioni proposte, è, senza dubbio, possibile trarre alcune conclusioni, che mirano non ad «imbrigliare» il concetto di *biblioteca digitale*, ma ad enunciarne gli aspetti imprescindibili. È evidente che il concetto di *biblioteca digitale* non implichi una semplice trasposizione digitale di una biblioteca reale. Essa è funzionale ad offrire ad una comunità di utenti più o meno ampia, per il tramite delle tecnologie informatiche, un contenuto culturale di validità scientifica. Ciò è posto in atto grazie alla sinergia di «assi strategici», quali la digitalizzazione, la tesaurizzazione (ed organizzazione) digitale e l'accessibilità on-line.

Bibliografia consultata

Borgman C. L., *Digital libraries and the continuum of scholarly communication*, in *Journal of Documentation*, 56, 4, 2000, pp. 412-430

Borgman C. L., *National electronic library report*, in Fox E. A. (ed.) *Sourcebook on Digital Libraries: report for the national science foundation*, Blacksburg, Computer Science Department, 1993, pp. 126-147

Borgman C. L., *The invisible library: paradox of the global information infrastructure*, in *Library trends*, 51, 4, 2003, p. 652

Borgman C. L., *What are digital libraries? Competing visions*, in *Pergamon. Information Processing and Management*, 35, 1999, pp. 227-243,

<<https://pdfs.semanticscholar.org/d0e6/90b74b3b9513d9d1f97cf366e31a3920a4bf.pdf>>

Faba Pérez C., *Biblioteche digitali: concetto e principali progetti*, in *Investigación Bibliotecológica*, 13, 26, 1999, pp. 64-78

Marchionini G. - Fox E. A., *Progress towards digital libraries: augmentation through innovation*, in *Information processing and management*, 35, 3, 1999, pp. 219-225

<https://ac.els-cdn.com/S0306457398000582/1-s2.0-S0306457398000582-main.pdf?_tid=25a411f0-fe0b-11e7-9e69-00000aacb35d&acdnat=1516471137_5866310621e56d197c1a9a06dce0917d>

Oppenheim C.-Smithson D., *What is the hybrid library?*, in *Journal of Information Science*, 25, 2, 1999, pp. 97-112

Bibliografia consigliata

Arms M. Y., *Digital Libraries*, Cambridge, MIT Press, 2000

Basili C. - Pettenati C., *La biblioteca virtuale*, Milano, Editrice Bibliografica, 1994

Ciotti F. - Roncaglia G., *Il mondo digitale: introduzione ai nuovi media*, Roma-Bari, Laterza, 2002

Malinconico S. M., *Biblioteche digitali: prospettive e sviluppo*, in *Bollettino AIB. Rivista Italiana di biblioteconomia e scienze dell'informazione*, 38, 3, 1998 <<http://bollettino.aib.it/article/view/8394/7498>>

Salarelli, A. - Tammaro, A. M., *Biblioteca digitale*, Milano, Editrice Bibliografica, 2000

Tammaro A. M., *Che cos'è una biblioteca digitale?*, in *Digitali Web. Rivista del Digitale nei Beni Culturali*, 1, 2005, <<http://digitalia.sbn.it/article/view/325/215>>

Sitografia

**Biblioteca digitale* (s.v.), in *Treccani*, <http://www.treccani.it/enciclopedia/biblioteca-digitale_%28Enciclopedia-Italiana%29/>

**Biblioteca virtuale* (s.v.), in *Treccani*, <http://www.treccani.it/enciclopedia/biblioteca-virtuale_%28XXI-Secolo%29/>

BNCF - Biblioteca Nazionale Centrale di Firenze, <<http://thes.bncf.firenze.sbn.it/termine.php?id=30683>>

DLF - Digital Library Federation, <<https://old.diglib.org/about/dldefinition.htm>>

DLF - Digital Library Federation, <<https://www.diglib.org/?s=definition+of+digital+librar>>

ICCU, Studio di fattibilità per la realizzazione della Biblioteca Digitale: <<http://www.iccu.sbn.it/upload/documenti/BDI-SDF.pdf>> (sezione prima); <<http://www.iccu.sbn.it/upload/documenti/BDI-SDF-Prog.pdf>> (sezione seconda)

Giovanna Battaglini

Il concetto di *Big Data* non è ancora oggetto di una definizione univoca e gli ambiti di applicazione non sono ancora stati perimetrati in maniera definitiva. Nell'Oxford Dictionary, si legge «Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions». Si tratta di una definizione esplicitamente riferita ai dati, in senso quantitativo; l'insieme dei dati è definito «dataset». In letteratura, si usano espressioni quali «Big Dataset» o in alcuni casi «Big Digital Objects». I dati possono comprendere sia risorse strutturate (es. i database, vd. database) che non strutturate (ess. immagini, testi nei social network, vd. social network).

Quando tuttavia si fa riferimento ai *Big Data*, si tende a comprendere non solo la quantità, quanto anche l'insieme delle tecnologie e delle procedure per la gestione di flussi enormi di informazioni, a grandi velocità (*Big Data Analytics*). Il trattamento massivo dei dati infatti deve comprendere sia una quantità di informazione importante, sia una ragionevole velocità di processazione delle informazioni stesse. Solitamente, quando si fa riferimento ai *Big Data*, si considerano ordini di grandezze enormi, comprendenti anche migliaia di terabyte (zettabyte).

Le finalità sono legate allo studio di comportamenti e tendenze umane in una data comunità, a scopo descrittivo o predittivo. In questo senso, i *Big Data* hanno visto uno sviluppo significativo certamente nel campo delle scienze sociali, ma si stanno espandendo in molti settori, ad esempio medico e farmaceutico.

Nelle Digital Humanities è in piena attuazione un processo di codificazione degli ambiti di ricerca; il quadro non è organico, proprio perché i *Big Data* richiedono un mutamento sia da un punto di vista tecnologico – in altre parole, sistemi capaci di organizzare enormi quantità di informazione, agendo anche su migliaia di server – sia dal punto di vista delle procedure. Consideriamo la nostra conoscenza in merito a enormi dataset di tipo culturale: i

milioni di volumi digitalizzati da Google, le informazioni da Google Maps, i miliardi di foto digitalizzate, le informazioni estraibili dai social network utilizzabili. Per un singolo studioso è difficile delimitare mentalmente quantitativi di informazione tanto elevati, per la maggior parte sconosciuti e potenzialmente sempre in espansione. Le enormi sfide epistemologiche che pongono si affrontano anche considerando i diversi passaggi del processo di digitalizzazione degli oggetti, ma i metodi per un loro efficace utilizzo sono ancora da esplorare, tanto dai data scientist che dagli umanisti. Una disciplina simile non sarebbe sbagliato considerarla a parte rispetto alle DH propriamente dette.

Esiste però un altro punto di vista per considerare i *Big Data* nell'ambito delle Digital Humanities, pensando alla rivalutazione delle metodiche di tipo quantitativo. I *Big Data* possono essere tali anche in senso relativo, rispetto all'oggetto di studio e al fatto che non possano essere collezionati e analizzati seguendo un approccio tradizionale. I *Big Data* lavorano inoltre in condizione di interoperabilità o comunque sono collegati tra di loro. Gli studi che adottano questo approccio sono necessariamente data-driven (vd. data-driven) e la loro diffusione ha portato anche a considerare l'avvento di un nuovo paradigma scientifico, nel quale i dati saranno in quantità tali da «parlare da sé», portando a obsolescenza concetti quali «ipotesi», «teoria», «modello». Un'ipotesi su cui in realtà converrebbe essere estremamente cauti, per due importanti problemi da affrontare: le difficoltà dietro alla selezione di materiali non strutturati e all'interpretazione degli stessi. Per chiarire meglio: sicuramente la linguistica è la disciplina che attualmente guarda ai *Big Data* con enorme interesse, proprio considerando la sua attitudine a lavorare con i dati, rinforzata dal sempre maggiore utilizzo dei corpora (vd. *Corpora*). Eppure, in un'ideale scala di valori dei dati, l'informazione estraibile da Google Books o dai social network e quella da collezioni di testi densamente annotati e delimitati a un determinato scopo si troverebbero sicuramente ai due estremi opposti.

In queste condizioni, la sfida con i *Big Data* si giocherà probabilmente su diversi fronti: da un lato, la realizzazione di strumenti sempre più sofisticati per estrarre valore dai dati; dall'altro, preparazione e conoscenza approfondita della tematica oggetto di studio, per un'interpretazione consapevole anche dal punto di vista qualitativo; inoltre i modi in cui creare interazione tra dataset e utente, attraverso l'uso di diverse interfacce. Infine, non da ultimo, una profonda riflessione interna sulle singole discipline, come la filologia, che richiedono per loro statuto un approccio di tipo qualitativo.

Esempio: MARTIN (*Monitoring and Analysing Real-time Tweets in Italian Natural language*) è uno strumento sviluppato dalla Fondazione Bruno Kessler per estrarre informazione da un dataset non strutturato come è il social network Twitter, comparare e analizzare il linguaggio dei tweet, le co-occorenze e i concetti-chiave. Questo software è stato presentato in una sua versione durante la sesta conferenza dell'AIUCD, per un'indagine sul sentimento del dibattito su determinati eventi storici su Twitter.

FlaviaSciolette

Si tratta di un'applicazione (software) utilizzata per accedere e usufruire dei siti Web. Tra i più comuni si trovano Microsoft Internet Explorer, Google Chrome, Mozilla Firefox e Apple Safari. Le principali funzioni del *browser* sono: recuperare l'informazione per l'utente, presentare l'informazione e accedere ad altre informazioni. Gli oggetti presenti sul Web vengono infatti identificati attraverso un URL che il *browser* si occupa di identificare e rendere disponibile alla lettura.

Per svolgere queste funzioni il *browser* si occupa di convertire il linguaggio HTML e gli altri formati che compongono una pagina web, come PNG o JPEG. In questo modo il file di testo e le immagini che compongono la pagina vengono visualizzati nella finestra del *browser*. Per eseguire questa funzione il *browser* utilizza un motore di rendering (engine), che interpreta le informazioni in ingresso codificate secondo uno specifico formato e, elaborandone, ne fornisce una rappresentazione grafica. Il modo in cui il *browser* interpreta e elabora graficamente i file HTML dipende dall'HTML stesso e dalle specifiche CSS contenute nella risorsa web.

Per svolgere la sua terza funzione, la navigazione, il *browser* si appoggia sempre ad un motore di ricerca (search engine) che analizza un insieme di dati e restituisce un indice dei contenuti disponibili classificandoli in modo automatico. I *Web browser* entrano in comunicazione con i *Web server* principalmente tramite il protocollo HTTP che gli permette di inviare informazioni e recuperare le pagine che contengono tali informazioni.

I primi *browser* come Netscape Navigator e Mosaic potevano semplicemente convertire il file HTML e aggiungere segnalibri per ricordare le pagine visitate. Le loro iterazioni più moderne supportano diversi formati HTML (come HTML 5 o XHTML), sistemi di encrypting per l'accesso a siti sicuro e funzioni Javascript. Lo sviluppo di dei *browser* ha permesso la creazione di siti più interattivi e con sistemi di visualizzazione più avanzata da un punto di vista grafico.

Sitografia

**Browser* (s.v.), in *Tech Terms*, <<https://techterms.com>>

How Browsers Work: Behind the scenes of modern web browsers, <https://www.html5rocks.com/en/tutorials/internals/howbrowserswork/#The_browser_main_functionality>

**Web Browser* (s.v.), in *Intro to ITC*, <http://openbookproject.net/courses/intro2ict/web/web_browser/>

AntonioMarsonFranchini

(o *Concordances*)

Per *concordanza verbale* (o spoglio lessicale) si intende il repertorio alfabetico delle parole presenti in un'opera o nelle opere di uno stesso autore (cfr. *Dizionario di Italiano*, Sabatini-Coletti), accanto alle quali sono indicati i passi in cui esse ricorrono adeguatamente concordate al contesto; si definisce, invece, *concordanza reale* il repertorio di passi che rimandano a concetti o argomenti ordinati tematicamente.

Se in passato le *concordanze* necessitavano di schedature lemmatiche eseguite a mano a notevoli costi di tempo e denaro, a partire dagli anni '90, la disponibilità sempre più ampia di *concordanze* informatizzate ha reso possibile ricerche sistematiche di tipo lessicale che destituissero l'ancestrale *auctoritas amanuense* a favore dei programmi informatici (come *Concordance*) di facile e rapida consultazione, capaci di ottenere in un tempo minimo le *concordanze* cercate. In un primo momento solo i testi sacri e pochi profani (i classici greci e latini, Shakespeare o Joyce) furono sottoposti alla selezione delle *concordanze*. La leggenda, infatti, narra che la prima *concordanza* fu realizzata da Hughes de St-Cher, erudito domenicano e membro della facoltà dell'Università di Parigi, per la *Vulgata*, con la collaborazione di cinquecento monaci che terminarono la schedatura nel 1230 con il titolo di *Concordantiae Sacrorum Bibliorum*; lavoro ventennale fu invece quello di John Bartelett e sua moglie per le *concordanze* dell'intera opera shakespeariana.

Un simile dispendio di tempo – e di denaro – durò fino a quando Padre Roberto Busa pensò di indicizzare l'*opera omnia* di Tommaso d'Aquino con il supporto dell'International Business Machines (meglio nota come IBM), sancendo l'atto di nascita delle Digital Humanities.

Stilare un repertorio di *concordanze*, più che un mero processo meccanico, quale certamente in parte è, è un *modus cogitandi* e *operandi* che stimola lo studioso a ricercare nelle liste di frequenza

dei termini contenuti in un'opera (o in opere) di uno stesso autore dei segnali forti sulla sua psicologia, la poetica e gli intenti.

Disporre di un repertorio di *concordanze* assolve quindi a varie funzioni: osservare gli usi di un termine, esaminare i contesti (semantici, sintattici o testuali) in cui esso è inserito, analizzare la regolarità con cui questo si accompagna ad altri termini nel suo contesto ma, soprattutto, contribuire all'esegesi dell'autore e della sua produzione.

L'introduzione di una *concordanza* elettronica, superiore a quella stampata, ha migliorato *sine dubio* le prestazioni della ricerca essenzialmente per due ragioni: per le sue dimensioni, dal momento che può includere un'ampia quantità di testo e permettere allo studioso di analizzare la *concordanza* adeguatamente inserita in un brano di ampio respiro, e perché può essere facilmente modificata, mentre una *concordanza* stampata è fissa. Spesso confuse con l'*index*, le *concordanze* sono sì un repertorio alfabetico di termini, ma, diversamente dall'indice – che associa a ciascuna parola il numero di pagina in cui è inserita – sono affiancate dall'indicazione dei passi (del capitolo e/o il versetto) in cui esse ricorrono, nonché dal passo stesso, assolvendo tutt'al più alla funzione di «indice contestualizzato».

Bibliografia consigliata

Gigliozzi G., *Introduzione all'uso del computer negli studi letterari*, a cura di Fabio Ciotti, Milano, Bruno Mondadori, 2003

Schweickard W. (a cura di), *Nuovi media e lessicografia storica. Atti del colloquio in occasione del settantesimo compleanno di Max Pfister*, Tübingen, Max Niemeyer Verlag, 2006

Stella F., *Testi letterari e analisi digitale*, Roma, Carocci editore, 2018

Sitografia

**Concordanza* (s.v.), in *Dizionario italiano Sabatini Coletti*, <http://dizionari.corriere.it/dizionario_italiano/C/concordanza.shtml>

**Concordance* (s.v.), in *Wikipedia* (EN), <[https://en.wikipedia.org/wiki/Concordance_\(publishing\)](https://en.wikipedia.org/wiki/Concordance_(publishing))>

The rise of the machines in Humanities. The Magazine of the National Endowment for the humanities, <<https://www.neh.gov/humanities/2013/julyaugust/feature/the-rise-the-machines>>

AlessandraDiMeglio

(lat. plur. *corpora*)

Genericamente, un *corpus* è un insieme di testi rappresentativi di un certo ambito. In ambito digitale, il termine ha un significato specifico: insieme di testi (orali, scritti o multimediali) disponibile su supporto elettronico digitale – online o da terminale – divisi singolarmente nelle loro unità minime (dette «token»), annotati secondo le convenzioni di un linguaggio detto di markup (il più diffuso è XML, per gli scopi delle DH nello standard TEI) e interrogabili attraverso un'interfaccia pensata per introdurre stringhe di caratteri od operatori logici. L'annotazione permette alla macchina di interpretare il testo secondo le sue specificità testuali (es. prosa o verso). I *Corpora* sono uno strumento di studio della Linguistica cosiddetta «dei *corpora*», basata su testi giudicati autentici e rappresentativi; tuttavia è interessante notare come, posizione ben argomentata da Francesco Sabatini, la linguistica italiana abbia sempre avuto un approccio basato su *Corpora* testuali, pur se non digitale, fin dai tempi del Dizionario della Crusca.

Esempio: CT – *Corpus Taurinense* (direzione: Manuel Barbera): un *corpus* di testi in fiorentino del secondo Duecento molto annotato.

FlaviaSciolette

Mutuato dall'ambito economico, il termine *crowdsourcing*, usato per la prima volta da Jeff Howe nell'articolo *The Rise of Crowdsourcing* apparso nel 2006 sulla rivista *Wired*, individua una nuova organizzazione del lavoro che consente la formazione di una rete capace di mettere in contatto numerosi volontari, hobbisti, o esperti senza che siano fisicamente presenti nello stesso luogo. Ciononostante questi possono partecipare all'ideazione e alla realizzazione di progetti lavorativi mediante l'uso degli strumenti informatici e proporre soluzioni innovative ad uno stesso problema. Il termine *crowdsourcing*, nato dalla fusione di «crowd» (folla) e «(out)sourcing» (esternalizzazione dell'attività fuori dalla propria impresa), in breve tempo, – anche grazie alla pubblicazione nel 2008 del libro *Crowdsourcing – Why the power of the crowd the future of business* di Howe – è entrato a pieno titolo nella terminologia corrente del settore finanziario, estendendo il suo campo semantico alle Digital Humanities.

Howe distingue quattro tipologie di *crowdsourcing*:

1. l'intelligenza collettiva o saggezza della folla, che si giova della conoscenza dei gruppi, in quanto superiore a quella dei singoli;
2. crowd-creation: si serve della creatività della folla per lo svolgimento delle attività;
3. crowd-voting: ordina le informazioni in base alle scelte della folla;
4. crowd-funding, che permette ai gruppi di raccogliere autofinanziamenti.

Henk van Ess ha sostenuto che il *crowdsourcing* consiste nella possibilità, data ad esperti, di risolvere problemi da condividere liberamente con chiunque; Estellés e González nel 2012 hanno definito il *crowdsourcing* una tipologia di attività online nella quale una persona, istituzione o organizzazione, senza scopo di lucro, propone ad un gruppo di individui, mediante un annuncio aperto, la

realizzazione libera e volontaria di un compito specifico. Il prodotto ottenuto dal *crowdsourcing* è, secondo alcuni, altamente specializzato e ad alto contenuto tecnologico, pari a quello ottenibile con l'utilizzo di professionisti pagati per eseguire la stessa prestazione, e in tempi anche minori (Antonio Dini); secondo altri, esso non garantirebbe alcuna professionalità né qualità, essendoci aperta partecipazione da parte di chiunque, anche di chi non ha adeguate competenze. I vantaggi del *crowdsourcing* sono vari e vanno dai costi, limitatamente bassi, alla possibilità di avvalersi di un nutrito numero di esperti (o appassionati) interessati alla ricerca, mentre sono da considerarsi degli svantaggi il rischio di fallimento, l'assenza di un contratto e/o la difficoltà a collaborare con un gran numero di persone. Tra i primi esempi di *crowdsourcing* nell'ambito delle Digital Humanities c'è il progetto *Transcribe Bentham*, iniziato nel 2010, che consiste nella digitalizzazione dei manoscritti di Jeremy Bentham, ma anche *Ancient Lives*, che trascrive i papiri greci, etc.

Il *crowdsourcing* trova applicazione soprattutto nell'ambito creativo e nel finanziamento di progetti (o crowdfunding) che diversamente rischierebbero di non essere attuati a causa del tempo e delle esigue risorse economiche. *Wikipedia*, ad esempio, è un progetto di *crowdsourcing* che si avvale di numerosi professionisti sul territorio nazionale e che è, oggi, tra quelli di maggior successo. Bruno Pellegrini, tra i più autorevoli esponenti del *crowdsourcing* europeo, ha realizzato l'*Italian Crowdsourcing Landscape*, una mappa che raccoglie i vari progetti attivi in Italia, che chiunque può aggiornare mediante un'app per iOS o Android. A questo si aggiungono numerosi altri progetti, nazionali ed europei, che attestano l'importanza del *crowdsourcing*. Ad esempio, Greg Cane, direttore del *Perseus Project*, ha lanciato un appello agli studenti: *Give Us Editors!*, invitandoli a collaborare alla digitalizzazione e alla modifica dei libri scansionati avendo questi competenze adeguate a questo compito. Tuttavia, come sostengono Hedges e Dunn: «in un tempo in cui il web sta trasformando simultaneamente il modo in cui le persone collaborano e comunicano, fondendo gli spazi che le comunità accademiche e non accademiche occupano, non è mai stato più importante considerare il ruolo che le comunità pubbliche – collegate o meno – giocano nella ricerca accademica umanistica», giacché anche all'esterno delle accademie la collaborazione finalizzata alla ricerca risulta essere di fondamentale importanza per il suo progresso. A tal proposito, Chris Blackwell e Tom Martin hanno sostenuto – non a torto – che gli studenti universitari potrebbero – e dovrebbero – pubblicare online articoli informativi capaci di accrescere il valore delle discipline umanii-

stiche e di migliorare la formazione e la qualità degli interventi di chi è interessato.

Certo è che, dati gli enormi vantaggi economici e di tempo, il *crowdsourcing* rappresenta oggi il più diffuso approccio lavorativo. Ampiamente adottato nel settore delle Digital Humanities esso contribuisce quotidianamente al progresso della ricerca, alla realizzazione di progetti mastodontici e al raggiungimento di risultati altrimenti impensabili.

Bibliografia consigliata

Andro M., *Digital Libraries and Crowdsourcing*, 5, London, Wiley, 2018

Brabham D. C., *Crowdsourcing*, Cambridge, The MIT press Essential Knowledge Serie, 2013

Estellés Arolas E.- González Ladrón-de-Guevara F., *Towards an integrated crowdsourcing definition*, in *Journal of Information Science*, 38, 2, 2012, pp. 189-200

Terras M., *Crowdsourcing in the Digital Humanities*, in Schreibman S.- Siemens R.- Unsworth J. (eds.), *A New Companion to Digital Humanities*, Chichester, Wiley-Blackwell, 2016, pp. 420-439

<http://discovery.ucl.ac.uk/1447269/1/MTerras_Crowdsourcing%20in%20Digital%20H

Sitografia

**Crowdsourcing* (s.v.), in *Wikipedia*, <<https://it.wikipedia.org/wiki/Crowdsourcing>>

Frost D. R., *Crowdsourcing, Undergraduates, and Digital Humanities Projects*, <<https://rebeccafrostdavis.wordpress.com/2012/09/03/crowdsourcing-undergraduates-and-digital-humanities-projects/>>

Howe J., *The rise of crowdsourcing*, <<https://www.wired.com/2006/06/crowds/>>

Mazzini E., *Il crowdsourcing tra necessità di coordinamento e perdita di controllo*/Capitolo 1 – *Il Crowdsourcing*, <https://it.wikisource.org/wiki/Il_crowdsourcing_t

Pezzali M., *Crowdsourcing: quando la rete... trova la soluzione*, in *Il Sole 24 ore*, <<https://www.ilsole24ore.com/art/SoleOnline4/Econom>

ia%20e%20Lavoro/2009/02/crowdsourcing-rete-soluzione.shtml>

Van Ess H., *Harvesting Knowledge. Success criteria and strategies for crowdsourcing*, <<https://www.slideshare.net/StifoPers/presentatie-henk-van-ess>>

AlessandraDiMeglio

(ingl. Acronimo per *Cascading Style Sheets*)

Rappresenta un linguaggio utilizzato per definire la formattazione di un documento HTML, XHTML o XML. Essendo uno standard patrocinato dal W3C, le regole per il suo utilizzo (Recommendation), emanate nel 1996, sono liberamente consultabili sul loro sito.

I *CSS* basano la loro sintassi su una gerarchia ben definita di marcatori, non modificabili o estendibili, che utilizzano per definire la visualizzazione dei contenuti di un documento. L'utilizzo di questi fogli di stile a cascata, si è resa necessaria sia per separare i contenuti della pagina HTML dalle impostazioni sulla formattazione in modo da ottenere un documento dalla codifica più snella e dalle dimensioni più leggere. La separazione del foglio di stile dal documento, avvenuta nel 1999 con la creazione di HTML4, rende tutta la fase di programmazione del documento (cioè tutte le impostazioni circa la visualizzazione della pagina, cosa va dove e come deve essere visualizzato dal browser) nettamente più ordinata e di più facile lettura sia dal programma che dagli autori stessi.

Ad oggi i *CSS*, e soprattutto la loro separazione dal documento stesso, risultano estremamente utili per definire separatamente le visualizzazioni che una stessa pagina web può (e deve) assumere sui diversi device: Computer, Tablet e Smartphone essendo dotati di schermi dalle dimensioni e dalle possibilità differenti, hanno bisogno ognuno della sua formattazione in modo da rendere l'interfaccia utente il più rispondente possibile ai principi di chiarezza e usabilità, basilari per qualsivoglia sito web.

Con un singolo foglio di stile è possibile:

1. Impostare la posizione dei diversi contenuti interni alla pagina web (porzioni di testo, immagini, link esterni, video, banner animati, eccetera);

2. definire lo stile, il carattere, il colore e il font delle titolazioni (definite in HTML come `<h1>` `<h2>` `<h3>`) dei paragrafi e di tutte le peculiarità tipografiche interne al documento.

Bibliografia consigliata

Duckett J., *HTML e CSS. Progettare e costruire siti web. Con Contenuto digitale per download e accesso on line*, Milano, Apogeo, 2017

Tisconi F., *Lineamenti di editoria multimediale*, Milano, Edizioni Unicopli, 2009

Sitografia

*W3CSS (s.v.), in *w3schools.com*, <http://www.w3schools.com/>

AlessiaMarini

L'enorme quantità di dati – small data, big data o smart data – che oggi si offrono alla continua fruizione degli utenti del Web e che li distolgono dal discernere i dati significativi da quelli superflui, trova una pratica consultazione nella rappresentazione grafica, infografica, di liste o mappe che trasforma suddetti dati in «parlanti». L'insieme delle tecniche atte a rappresentare i dati in maniera interattiva è detto *Data Visualization*.

L'esigenza di semplificare mediante la rappresentazione visiva un quantitativo di informazioni altrimenti caotica e disorganizzata è di antica datazione. La sostituzione del dato scritto con quello visivo ha sì semplificato la ricezione delle informazioni, ma ha anche consentito una loro più rapida elaborazione; ha aumentato la trasparenza della comunicazione; ha fornito innovativi punti di vista grazie a nuovi input di analisi e ha consentito una più svelta risoluzione delle problematiche.

Coniata da William Fetter per la prima volta nel 1960, l'espressione *computer graphics* inaugura la *Data Visualization* digitale che si è evoluta negli anni esponenzialmente. I primi progetti di rappresentazione digitale nell'ambito delle Digital Humanities – come l'analisi testuale fatta da John Burrow dei versi del XVII e del XVIII secolo – utilizzavano la rappresentazione grafica per interpretare la mole di dati disponibili all'indagine umanistica. Un progetto di *Data Visualization* attualmente in vigore è il *Mapping the Republic of Letters*, che utilizza rappresentazioni grafiche per aiutare gli studiosi ad analizzare oltre 55.000 lettere e documenti archiviati digitalmente nel proprio database (<http://republicofletters.stanford.edu/index.html>).

Tra i principali tool delle *Data Visualization* si annoverano:

1. Tableau: (a pagamento) in grado di produrre grafici e visualizzazioni interattive avanzate;
2. Infogram: utile per la creazione di infografiche, semplice e veloce da utilizzare;

3. Carto: adatto alla pubblicazione di mappe online;
4. Google Chart Tools: che consente la creazione di un'ampia gamma di grafici embeddabili all'interno dei siti web.

L'efficacia delle *Data Visualization* trova conferma, oltre che in quello scientifico, anche nel settore delle DH. La possibilità di individuare connessioni, che normalmente non potrebbero essere viste, ha infatti incentivato l'indagine pionieristica in questo campo, per consentire anche ai ricercatori umanistici di servirsi di nuovi mezzi all'avanguardia. Il TACC (*Texas Advanced Computing Center*), ad esempio, si propone la creazione di software grafici specifici per gli studi umanistici e promuove la combinazione di testo, video, grafica 3D, animazione, audio etc. per rimediare, in questo modo, alla scarsità di mezzi di cui dispongono gli umanisti nel campo grafico (<https://www.tacc.utexas.edu/research-development/tacc-projects/a-thousand-words>).

Bibliografia consigliata

Fischetti T. *et al.*, *Data Analysis and Visualization*, Birmingham, Packt, 2016

Murray S., *Interactive Data Visualization for the Web*, Sebastopol (CA), O'Reilly Media, 2013

Stephen A. T., *Data Visualization with Javascript*, USA, No Starch Press, 2015

Telea A. C., *Data Visualization. Principles and Practice*, Boca Raton (FL), CRC, 2015

Sitografia

**Data Visualization* (s.v.), in *Wikipedia*, <https://en.wikipedia.org/wiki/Data_visualization>

Mapping the Republic of Letters, <<http://republicofletters.stanford.edu/index.html>>

Minguzzi G., *Data Visualization: quando i numeri raccontano le storie*, <<https://www.merita.biz/data-visualization/>>

Minto P., *Data Visualization, una storia/1*, <<https://www.rivistastudio.com/data-visualization-una-storia/>>

Minto P., *Data Visualization, una storia/2*, <<https://www.rivistastudio.com/data-visualization-2/>>

TACC - *Texas Advanced Computing Center*, <<https://www.tacc.utexas.edu/research-development/tacc-projects/a-thousand-words>>

Zambon M., *Glossario: Data Visualization*, <<https://www.tagmanageritalia.it/glossario-data-visualization/#gref>>

Zepel T., *Visualization as Digital Humanities?*, <<https://www.hastac.org/blogs/tzepel/2013/05/02/visualization-digital-humanities>>

Alessandra Di Meglio

(ingl. Acronimo: *DB*)

Sinonimi: «base di dati», «banca dati»; fr. *base de données*.

Un *database* è una collezione di dati omogenei per contenuto e formato, organizzati secondo una determinata struttura, scelta da chi sviluppa il progetto (es. il modello relazionale, uno dei più diffusi nell'ambito della progettazione). Un *DB* è solitamente interrogabile da un'interfaccia visuale da schermo – online o da terminale – pensata per introdurre stringhe di caratteri od operatori logici: un linguaggio di interrogazione per *DB* è detto «query language» (linguaggio di ricerca). La classe di software per la creazione e la gestione dei *DB* è detta DBMS (*Database Management System*); uno dei più diffusi e maggiormente supportati è l'open-source Mysql, rilasciato dalla Oracle.

Esempio: *BEDT – Bibliografia Elettronica dei Trovatori* (Direzione Scientifica: Stefano Asperti, Direzione tecnica: Luca De Nigro) una base di dati su modello relazionale dedicata alla poesia provenzale.

FlaviaSciolette

L'espressione *Digital Dark Age*, di conio alquanto recente, è attestata per la prima volta nel 1997 in una relazione presentata nella Conferenza dell'*International Federation of Library Associations and Institutions* (IFLA) ed è stata successivamente ripresa e «valorizzata» nel Convegno intitolato *Time and Bits* a cura della *Long Now Foundation* e del *Getty Conservation Institute* nel 1998. L'espressione è parzialmente mutuata dalla storiografia, in seno alla quale si ricorre all'espressione *Dark Age* per indicare un periodo di assenza di documentazione scritta con riferimento al Medioevo Ellenico o all'Alto Medioevo.

La definizione di *Digital Dark Age* è legata alla dicotomia concettuale «digital preservation» ~ «digital obsolescence». In particolare, il concetto di Medioevo Digitale si configura come il possibile, nefasto futuro esito dell'obsolescenza digitale, vale a dire l'indisponibilità ed inaccessibilità di dati in supporto digitale, dovuta o al deterioramento dei supporti (hardware) o alla troppo rapida evoluzione degli strumenti informatici atti a creare, modificare e leggere i dati (software). In particolare, il deterioramento dei supporti è dovuto al fatto che essi non hanno una durata illimitata, perché i supporti magnetici (hard disk, pen drive) possono risentire di campi elettromagnetici esterni, mentre i supporti ottici (CD, DVD) possono essere danneggiati dalla luce o dalla temperatura. Ad esso si affianca il problema della «usability»: è necessario che i digital data siano sempre in un formato accessibile, aspetto questo di non facile gestione.

La realtà, apparentemente paradossale, risiede nel fatto che le collezioni digitali facilitano l'accesso alle informazioni, ma non la loro conservazione. Sono stati avviati programmi di ricerca, finalizzati ad elaborare (analytic) tools – come il *cloud-based storage* – e strategie di preservazione delle informazioni conservate in formato digitale – strategie legate alla «interoperability» tra software e servizi (SaaS).

Bibliografia consigliata

Howell A., *Perfect One Day – Digital The Next: Challenges in Preserving Digital Information*, in *Australian Academic & Research Libraries*, 31, 4, 2013, pp. 121-141 <<https://www.tandfonline.com/doi/pdf/10.1080/00048623.2000.10755130>>

Ross H. D., *Preserving Digital Materials*, München, Rowman & Littlefield, 2005

Sitografia

**Digital Dark Ages* (s.v.), in *LisWiki*, <https://liswiki.org/wiki/Digital_Dark_Ages>

IFLA - International Federation of Library Associations and Institutions, <<https://www.ifla.org/>>

Kuny T., *A Digital Dark Ages? Challenges in the Preservation of Electronic Information*, relazione presentata al *Workshop: Audiovisual and Multimedia joint with Preservation and Conservation, Information Technology, Library Buildings and Equipment, and the PAC Core Programme*, September 4, 1997, <<https://archive.ifla.org/IV/ifla63/63kuny1.pdf>>

**Obsolescenza digitale* (s.v.), in *Treccani*, <http://www.treccani.it/enciclopedia/obsolescenza-digitale_%28Lessico-del-XXI-Secolo%29/>

**Obsolescenza digitale* (s.v.), in *Wikipedia*, <https://it.wikipedia.org/wiki/Obsolescenza_digitale>

Rusbridge C., *Excuse me... some Digital Preservation Fallacies?*, in *Web Magazine for Informations Professionals*, <<http://www.ariadne.ac.uk/issue46/rusbridge/>>

Shave L., *It's a digital world. Records and information Management Professionals. Guide to avoid Digital Dark Ages*, <<http://rimpa.com.au/assets/2015/09/eBook-Avoiding-the-digital-dark-age.pdf>>

Giovanna Battaglino

Con il termine *Digital Death* si intende solitamente l'insieme delle questioni che riguardano, in primo luogo, i modi in cui è cambiato il rapporto tra il singolo individuo e la morte a causa dello sviluppo delle nuove tecnologie informatiche e mediatiche, a partire soprattutto dalla diffusione popolare del Web. In secondo luogo, le conseguenze che ne derivano per quanto concerne la costruzione della propria identità personale e il suo legame con la memoria in seguito alla morte di sé o di un altro individuo.

Tre i problemi specifici su cui si concentrano gli studiosi della *Digital Death*: 1) gli effetti che la morte di un singolo individuo produce all'interno della realtà digitale e, quindi, nella vita «reale» di chi soffre la perdita e il lutto; 2) gli effetti che la perdita di oggetti e informazioni digitali personali producono all'interno della realtà fisica di un singolo individuo. In altre parole, le conseguenze di una perdita definitiva di informazioni, fotografie, messaggi, quando si rompe il proprio computer senza aver fatto un preventivo backup del materiale lì contenuto; 3) l'inedito significato che assume il concetto di immortalità in relazione tanto al singolo individuo quanto agli oggetti e alle informazioni digitali personali.

Questi problemi implicano un'attenzione particolare, a livello interdisciplinare, da parte degli studiosi per tutti i progetti digitali che cercano di farci sopravvivere alla nostra morte quali «spettri digitali» (Eter9, Eterni.me, LifeNaut, ecc.) e per le questioni inerenti a concetti come «patrimonio digitale», «memoria digitale», «eredità digitale», i quali prendono forma a partire dall'uso decennale dei social network da parte degli utenti del Web.

Davide Sisto

(Digital Object Identifier)

Definito anche identificatore persistente, il *DOI* identifica un oggetto digitale, i suoi dati e i metadati (autore, titolo, editore, data di pubblicazione, ma anche numero del fascicolo di una rivista, l'elenco degli articoli in essa contenuti, etc.), in maniera persistente e secondo uno schema strutturato ed estensibile.

L'associazione del *DOI* all'oggetto è totalmente univoca: attraverso il *DOI*, infatti, si accede alla pagina Internet contenente l'oggetto e se l'oggetto viene spostato si sposta anche il *DOI* che lo identifica. Differentemente da ISBN e ISSN, esso è immediatamente azionabile e si configura come siringa alfanumerica divisa in prefisso e suffisso.

Dato un *DOI* di questo tipo: 10.1491/napoli081995, il prefisso è 10.1491, di cui 10 identifica il *DOI*, mentre 1491 indica il registrante; napoli081995 è il suffisso che coincide con l'oggetto.

A occuparsi della registrazione dei *DOI* c'è la *multilingual European DOI Registration Agency*, meglio nota come mEDRA, l'*Agenzia Europea di Registrazione del DOI*, nominata dall'IDF (*International DOI Foundation*, www.doi.org) il primo luglio 2003.

Il successo ottenuto da mEDRA e il forte interesse mostrato dagli editori scientifici e professionali, universitari/scolastici, ma anche dalle università e dai centri di ricerca, dalle biblioteche e dalle pubbliche istituzioni, ha incentivato una più ampia diffusione del *DOI*, eliminando le barriere economiche, consentendone l'utilizzo anche ai piccoli editori e incentivandone, in questo modo, il processo di standardizzazione.

Bibliografia consigliata

ISO 26324, *International Standard. Information and documentation – Digital Object Identifier system*, Switzerland, ISO, 2012

Sitografia

**DOI-Digital Object Identifier (s.v.), in Wikipedia, <https://it.wikipedia.org/wiki/Digital_object_identifier>*

DOI. Un código esencial para citas bibliográficas y búsquedas científicas, <<http://infobib.blogspot.com/2009/04/doi-un-codigo-esencial-para-citas.html>>

mEDRA, <<https://www.medra.org/it/DOI.htm>>

mEDRA – multilingual DOI Registration Agency, <http://www.aie.it/Portals/_default/Skede/Allegati/Skeda105-1466-2007.5.2/Medra%20brochure%20applicazioni%20IT.pdf>

Servizi SBA: Informazioni sul DOI, <<http://biblioteche.unige.it/node/255>>

AlessandraDiMeglio

(ingl. Acronimo per *Document Type Description*)

La *DTD* è un linguaggio per definire la grammatica che descrive la composizione degli elementi che fanno parte di una certa classe di documenti XML. È un retaggio del SGML, in quanto già se ne serviva per strutturare formalmente e gerarchicamente tutti i componenti del linguaggio.

La *DTD* è un insieme di dichiarazioni che principalmente determinano quali elementi devono e/o possono essere usati, quali attributi possono essere aggiunti agli elementi e soprattutto quali relazioni legano o separano i vari elementi. Le *DTD* risultano, quindi, molto utili per la comprensione e la validazione del documento XML, per la creazione di fogli di stile e di interfacce dinamiche e per l'interscambio di documenti. In parole povere altro non è che un documento in cui vengono sia inseriti tutti gli elementi e gli attributi che il codificatore potrà utilizzare nel documento XML, sia specificati i tipi di rapporti che intercorrono tra di loro.

Ogni *DTD* può essere, arbitrariamente, interna o esterna (in questo caso può diventare un file autonomo e interscambiabile) al documento XML, ma in entrambi i casi essa deve essere necessariamente dichiarata nel documento stesso, subito dopo la dichiarazione XML.

Bibliografia consigliata

Ausiello G. *et al.*, *Modelli e linguaggi dell'informatica*, Milano, McGraw-Hill, 1991

Ciotti F., *Il testo e l'automa. Saggi di teoria e critica computazionale dei testi letterari*, Roma, ARACNE Editrice s.r.l., 2007

Møller A. - Schwartzbach M.I. - Gaburri S. (a cura di), *Introduzione a XML*, Milano, Pearson, 2007

Tisconi F., *Lineamenti di editoria multimediale*, Milano, Edizioni Unicopli, 2009

AlessiaMarini

(ingl. Acronimo per *Encoded Archival Description*)

L'*Encoded Archival Description* (EAD) è uno strumento per la conversione e la pubblicazione in formato elettronico di strumenti di ricerca archivistici prodotti originariamente su supporto cartaceo, nonché per l'elaborazione e lo scambio di descrizioni archivistiche in formato nativamente digitale. L'adozione di tecnologie come l'*eXtensible Markup Language* (XML), volte alla conservazione e alla comunicazione dei dati, indipendentemente da specifiche piattaforme hardware e software, consente di garantire la persistenza della struttura e del contenuto delle descrizioni, nonché la loro accessibilità e validità nel tempo.

Lo standard prende forma nell'ambito del *Berkeley Finding Aids Project* (BFAP), avviato nel 1993 presso la Berkeley University della California, con l'intento di sviluppare un modello non proprietario per la codifica in formato digitale di strumenti di ricerca quali inventari, elenchi sommari, indici utilizzando lo *Standard Generalised Mark-up Language* (SGML). La versione 1.0 di *EAD* è stata rilasciata nel 1998, accompagnata dalla pubblicazione della *Tag library* e delle *Application guidelines* sul sito ufficiale del progetto, ospitato in quello della Library of Congress che era diventata l'agenzia responsabile del mantenimento di *EAD*.

Nel 2002, dopo un attento lavoro di revisione da parte della comunità scientifica internazionale, è stata rilasciata una nuova versione del modello che è stata allineata, in alcune sue componenti, all'edizione del 2000 dello standard di descrizione archivistica ISAD(G), *General International Standard Archival Description*.

In anni recenti, anche alla luce di importanti novità – dal rilascio dell'*Encoded Archival Context - Corporate bodies, Persons, and Families* (EAC-CPF), alla comparsa di collection management system, alle nuove opportunità cui sembrano aprire i *Linked Open Data* – si è sentita la necessità di un'ulteriore revisione dello standard finalizzata ad accrescerne chiarezza e coerenza, a garantire una maggio-

re interoperabilità e un generale miglioramento delle funzionalità in ambienti internazionali e multilingue.

Il sito ufficiale dello standard rende disponibile il modello dati, espresso come DTD, XML Schema, Relax NG, Schematron, unitamente alla *Tag Library* e al foglio di stile che dovrebbe garantire il passaggio da EAD 2002 a EAD3.

Nell'estate del 2015 la versione EAD3 è stata adottata come standard dalla *Society of American Archivists* (SAA).

Bibliografia consigliata

Carucci P. - Guercio M., *Manuale di archivistica*, Roma, Carocci, 2008

Michetti G. (a cura di), *EAD: Encoded archival description tag library = Descrizione archivistica codificata dizionario dei marcatori*, Roma, ICCU, 2005

Pala F., *Lo standard EAD3 per la codifica dei dati archivistici: qualche novità e molte conferme*, in *J-LIS*, 8, 3, 2017, pp. 148-176

Pitti D., *Encoded Archival Description: The Development of an Encoding Standard for Archival Finding Aids*, in *The American Archivist*, 60, 3, 1997, pp. 268-283

Sitografia

EAD - Encoded Archival Description (official site), <<http://www.loc.gov/ead/index.html>>

ICAR - Istituto Centrale per gli Archivi, <<http://www.icar.beniculturali.it/index.php?id=52>>

ConcettaDamiani

(ingl. Acronimo per *Rich Site Summary/ RDF Site Summary o Really Simple Syndication*)

Quando si parla di *Feed RSS* ci si riferisce a un flusso, di un'unità di informazioni formattata secondo regole decise precedentemente per rendere interpretabili e interscambiabili una serie di contenuti frequentemente aggiornati tra diverse applicazioni e piattaforme. Si tratta di un metodo per fornire e distribuire i contenuti sul Web attraverso un formato standard XML e, come si desume dall'acronimo, spesso in un formato che assomiglia ad un sommario con link che rimandano al sito web originale. Si tratta quindi di una applicazione di XML che deve conformarsi a XML 1.0 secondo le specifiche pubblicate dal W3C (*World Wide Web Consortium*).

Nel corso del tempo il formato ha subito diverse evoluzioni e, ad oggi, esistono differenti versioni di *RSS*. Una prima versione 0.90 venne sviluppata da Netscape nel 1999, seguita da una versione semplificata, la 0.91. Il vero sviluppo avvenne però per mano del team di «content management» di Userland che, partendo dal formato RDF (*Resource Description Framework*), diede vita a *RSS 1.0*. Dave Winer, membro del team di Userland ha poi sviluppato una nuova versione semplificata dal nome di *RSS 2.0*. La vasta popolarità del formato *RSS* e la sua lunga storia hanno fatto in modo che il termine *Feed RSS* venga utilizzato spesso come sinonimo di «web feed». Si tratta però di un uso incorretto del termine in quanto non tutti i formati di *feed* sono *RSS*, ad esempio oggi un altro formato di vasto utilizzo è Atom, un diretto concorrente di *RSS*.

Nella sua forma più semplice un *Feed RSS* è composto da un elemento principale `<rss>` seguito da un elemento `<channel>` e dai suoi sotto elementi `<item>`. L'elemento `<channel>` deve contenere tre informazioni irrinunciabili: un titolo, un link, una breve descrizione che informi l'utente delle informazioni contenute nell'aggiornamento. Gli elementi opzionali spaziano dalla lingua delle informazioni contenute nel feed alla mail della persona responsabile del sito web che ha creato il *Feed RSS*.

La funzione principale dei *Feed RSS* è garantire la semplicità di informazione tramite Internet rendendo automatico l'invio degli aggiornamenti dei siti o blog che vengono visitati dall'utente. Per usufruire dei feed generati automaticamente dai siti che interessano il lettore è necessario fare uso di *Feed Reader* o aggregatori in grado di monitorare tutti i siti richiesti dall'utente e restituire gli aggiornamenti, per l'appunto, in forma di sommario facilmente consultabile.

Sitografia

**RSS (s.v.)*, in *Tech Terms*, <<https://techterms.com>>

RSS Advisory Board, <<http://www.rssboard.org/rss-specification#whatIsRss>>

Rss20AndAtom10Compared, <<http://www.intertwingly.net/wiki/pie/Rss20AndAtom10Compared>>

Antonio Marson Franchini

La *filologia digitale* si configura come il punto di intersezione fra l'ecdotica e l'informatica. Pertanto, la definizione di *filologia digitale* non può essere scissa dalla definizione di *edizione digitale*. In prima istanza, la *filologia digitale* potrebbe essere definita come l'uso combinato di strumenti informatici e specifici metodi, al fine di allestire una edizione digitale.

Come sottolinea Andrews (2012, 2) è difficile definire in maniera non «equivoca» una edizione digitale: ciò dipende dal fatto che il sintagma può contestualmente riferirsi sia alla trascrizione di un testo critico (elaborato seguendo l'approccio tradizionale) in formato elettronico – definizione, invero, un po' banalizzante e priva di una vera portata innovativa (cfr. Fiormonte 2007) –, sia alla pubblicazione di un testo critico allestito tramite il ricorso a specifici tools e software digitali. Pertanto, la Andrews (2012, 2) propone una definizione, per così dire, «conciliante», asserendo che «we might call «digital philology» an approach to textual editing that welcomes the aid of technology wherever possible and which will usually, but not necessarily, result in a digital publication». In tempi ancor più recenti, Rosselli Del Turco (2017, 3) propone di intendere, in via generale, la *filologia digitale* come «l'uso di strumenti e metodi dell'informatica applicati all'ecdotica con l'obiettivo di creare un'edizione critica (o diplomatica) di un testo».

Una edizione digitale presenta diversi vantaggi rispetto ad una edizione critica tradizione (cfr. Rosselli Del Turco 2010 e 2017), fra i quali – da un punto di vista precipuamente filologico – vanno annoverati la possibilità di una gestione «elastica» delle varianti testuali, senza limiti di spazio per quanto concerne l'estensione dell'apparato critico; l'intertestualità, intesa come possibilità di prevedere collegamenti con testi supplementari (quali commenti, edizioni critiche precedenti *et similia*); la possibilità di consultare i vari testimoni. *Brevibus verbis*, l'edizione critica digitale si presenta come un'edizione dinamica: «l'obiettivo del supporto digitale è di sottrarre all'immobilità della stampa la vivacità del testo criticamente stabilito» (Giacomelli 2012, contributo al quale si rimanda

anche in merito alle problematiche della digitalizzazione di fonti papiracee; in merito al «dinamismo» di una edizione digitale ed al suo carattere ipertestuale, cfr. Tomasi 2018).

L'approccio filologico digitale si fa, dunque, latore di innovazioni che, naturalmente, non si esauriscono nella *facies* digitale: «la filologia digitale si avvale di strumenti informatici, ma non è identificabile con essi» (Rosselli Del Turco 2010). La «digitalizzazione» del lavoro del filologo non consente soltanto di rendere più efficienti alcune fasi del lavoro di allestimento di una edizione critica (quali, in particolare, la *recensio* e la *collatio codicum*). In tutti i contributi che hanno proposto riflessioni di natura epistemologica (e conseguenti proposte di definizione) in merito alla *filologia digitale* emerge chiaramente la consapevolezza che «il passaggio del testo dalla modalità pre-informatica e gutemberghiana a quella segnata dall'informatica o digitale è tale da cambiare sostanzialmente non solo il concetto di testo, ma anche la natura stessa della filologia stessa» (Mordenti 2012, 37). Il filologo digitale adotta un nuovo approccio, che sposta l'attenzione dal testo «ideale» (per ricostruire il quale si individua il testimone la cui *auctoritas*, per vari motivi, è superiore a quella degli altri testimoni) al testo «reale». In altre parole, la *filologia digitale* si interessa alla individualità e alle variazioni di ogni singolo testimone, anziché fornire un «unificato» *textus receptus*. Come ben sintetizza Mattioda (2007), «l'ambiente digitale permette un passo in avanti nella visualizzazione e nel confronto dei testi, ma soprattutto supera l'idea del testo definitivo: non solo il filologo non sceglie un testo più importante, ma modifica il suo lavoro grazie ai linguaggi di codifica e ai codici di mark up. Il testo che il filologo dovrà preparare

..

non è soltanto un testo univoco, ma un ipotesto che può contenere in sé varie trascrizioni».

Infine, non bisogna credere che l'approccio filologico digitale annulli o ridimensioni l'azione umana. La possibilità di includere in una edizione critica digitale un apparato critico contenente tutte le varianti e/o i collegamenti a tutti i testimoni non annulla la responsabilità ecdotica individuale del filologo, il quale opera sulla base di precisi criteri editoriali (da lui selezionati e chiaramente indicati), sulla base dei quali dà l'avvio alla propria esegesi dei dati raccolti. La filologia, seppure in un contesto digitale, resta sempre sostanzialmente «interpretazione». Contestualmente, si complessifica il compito del lettore, il quale dev'essere in grado di ricostruire il percorso interpretativo del filologo (potendo natural-

mente anche dissentire in merito a determinate scelte). Ciò rende l'edizione digitale un vero e proprio strumento di ricerca.

Bibliografia consigliata

Aa.Vv., *Accademia Nazionale dei Lincei*, Tavola rotonda sul tema *Filologia Digitale: problemi e prospettive*, 135, Roma, Bardi Edizioni, 2017

Cotticelli Curras P., *Linguistica e filologia digitale*, Alessandria, Edizioni dell'Orso, 2011

Fiormonte D., *Scrittura e filologia nell'era digitale*, Torino, Bollati Boringhieri, 2003

Fiormonte D., *Note sul problema della filologia digitale*, in *Parole online. Quale editoria e filologia nell'era digitale? – Nuova Informazione Bibliografica*, II, pp. 355-362, <https://infolet.it/files/2009/04/per_rcicala4.pdf>

Mattioda E., *Alcune considerazioni su filologia e testi digitali*, in *Philologie et Politique*, 7, 2007, pp. 163-172, <<http://journals.openedition.org/laboratoireitalien/143>>

Mordenti R., *Filologia digitale (a partire dal lavoro per l'edizione informatica dello Zibaldone Laurenziano di Boccaccio)*, in *Humanist Studies & the Digital Age*, 2, 1, 2012, pp. 37-56 <<http://journals.oregondigital.org/hsda/article/download>>

Spinazzè L., *Filologia digitale. Dalla ricerca alla didattica. L'informatica umanistica al servizio delle scienze dell'antichità*, Trento, Tangram Edizioni Scientifiche, 2015

Sitografia

Andrews T. L., *The third way: philology and critical edition in the digital Age*, <https://boris.unibe.ch/43071/1/variants_postprint.pdf>

Giacomelli I., *Edizione digitale di fonti primarie. Elegie tirtaiche su papiro: un esempio*, in *Griselda online – Portale di letteratura* (2012), <<http://www.griseldaonline.it/informatica/edizione-digitale-di-fonti-primarie.html>>

**Informatica umanistica* (s.v.), in *Treccani*, <http://www.treccani.it/enciclopedia/informatica-umanistica_%28XXI-Secolo%29/>

Rosselli Del Turco R., *Filologia digitale: ragioni, problemi, prospettive di una disciplina*, III incontro di filologia digitale, Verona 3-5 marzo 2010, <<http://www.dfl.univr.it/documenti/Iniziativa/dall/dall639204.pdf>>

Rosselli Del Turco R., *L'edizione scientifica digitale: strumenti e progetti*, Verona 20-21 aprile 2017, <<http://filologiadigitale-verona.it/wp-content/uploads/2017/04/Introduzione-alla-Filologia-Digitale-1.pdf>>

Tomasi F., *Le nuove frontiere della filologia*, in *Argo*, 2, 2018 <http://www.argonline.it/argo/argo-n-2/argo-n-2_-francesca-tomasi-nuove-frontiere-filologia/>

Giovanna Battaglino

(ingl. Acronimo per *File Transfert Protocol*)

È il protocollo TCP/IP di collegamento ad una rete che consente il trasferimento di file tra cui quelli ASCII, EBCDIC e binari. Si tratta di un linguaggio studiato per facilitare la comunicazione all'interno di una architettura client-server. La prima versione venne sviluppata dal MIT nel 1971 e di proponeva degli obiettivi primari:

1. promuovere la condivisione di file (programmi o dati);
2. incoraggiare l'uso indiretto o implicito di computer remoti;
3. risolvere in maniera trasparente incompatibilità tra differenti sistemi di stoccaggio file tra host;
4. trasferire dati in maniera affidabile ed efficiente.

Il protocollo utilizza delle connessioni di tipo TCP (*Transition Control Protocol*), distinte tra di loro, per trasferire i documenti e controllare lo stato dei trasferimenti stessi; per garantirne ulteriormente la sicurezza ad ogni accesso viene richiesta l'autorizzazione del client attraverso un nome utente ed una password. Negli anni è stata sviluppata una ulteriore versione del *FTP*, integrandolo con un sottostrato SSL/TLS (*Secure Sockets Layer/ Transport Layer Security*) che riduce il rischio di accessi non autorizzati alle comunicazioni client-server.

La sua logica di funzionamento è relativamente semplice: *FTP* utilizza due connessioni diverse per la gestione di dati e comandi ed è eseguibile solamente tramite l'intercessione di un software chiamato client. Attraverso il client ci si conatterà al server *FTP* (solitamente sempre aperto alla porta 21) e, grazie a questa comunicazione, si aprirà il canale comandi attraverso il quale il client riuscirà a comunicare con il server. Si possono avere, quindi, due tipi di canale, uno attivo, che utilizza un numero di porta casuale per poi spostare il contenuto del trasferimento verso la porta standard, e uno passivo, il quale sfrutta una porta casuale ma superiore a 1023 senza spostare il trasferimento dei dati.

Si tratta, comunque, di un linguaggio di alto livello, quindi facilmente comprensibile all'uomo e quindi relativamente di facile utilizzo.

Sitografia

Poste J. - Reynolds J., *FILE TRANSFERT PROTOCOL (FTP)*, <<http://www.rfc.altervista.org/rfctradot>>

AlessiaMarini

Gold Open Access

Si dice che una sede scientifica pubblica in *gold open access* quando i singoli contributi di ogni autore – una volta editi – sono liberamente e immediatamente disponibili online.

21.1 Sitografia {-}:

**Gold Open Access (s.v.)*, in *Consiglio Nazionale delle Ricerche. Biblioteca d'Area di Bologna*, <<http://biblioteca.bo.cnr.it/index.php/it/open-access/pubblicare-oa/item/272-gold-road>>

AntonelloFabioCaterino

Green Open Access

Si parla di *green open access* quando l'autore contribuisce a diffondere il proprio contributo immettendolo in repository (istituzionali e non) liberamente consultabili. È possibile che alcuni editori vietino espressamente questa pratica, e che dunque siano caricate online versioni pre-print o post-print.

21.2 Sitografia {-}:

**Green Open Access (s.v.)*, in *Consiglio Nazionale delle Ricerche. Biblioteca d'Area di Bologna*, <<http://biblioteca.bo.cnr.it/index.php/it/open-access/pubblicare-oa/item/271-green-open-access>>

AntonelloFabioCaterino

Hidden Web

L'*Hidden Web* (o Invisible Web o Deep Web o Undernet) è l'insieme delle risorse informative del World Wide Web non indicizzate dai normali motori di ricerca, come Google, Bing o Yahoo.

Il *Web Sommerso* nasce a scopo militare, quando nel 1969 il DARPA (*Defense Advanced Research Projects Agency*), un'agenzia governativa del Dipartimento della Difesa degli Stati Uniti incaricata di realizzare nuove tecnologie, dà vita all'ARPANET (*Advanced Research Projects Agency NETwork*), una rete basata su un'architettura client/server destinata all'uso militare. L'uso di questa rete, in declino durante gli anni della guerra fredda, rimane sotto il controllo delle università diventando utile strumento di comunicazione e di scambio di informazioni scientifiche, fino a quando nel 1991 al CERN (*Organizzazione Europea per la Ricerca Nucleare*) di Ginevra si definisce il protocollo HTTP (*HyperText Transfer Protocol*), che permette la nascita del World Wide Web e la diffusione di Internet. Per proteggere le comunicazioni governative, la Marina statunitense inizia nel 2002 a servirsi di TOR (*The Onion Router*), un browser capace di garantire l'anonimato attraverso il continuo rerouting su nodi, che l'Elettronica Frontier Foundation, un'organizzazione no profit promotrice dei diritti e delle libertà digitali, inizia a finanziare nel 2004 perché se ne estenda l'utilizzo al di fuori dell'ambito militare. Molti anni dopo l'uso di questa rete è concesso anche ai civili.

L'enorme vastità del Web e l'incalcolabile quantità di dati di cui è portatore, ha posto gli studiosi di fronte alla difficoltà di individuare e circoscrivere i livelli che lo costituiscono riconoscendo a ciascuno di essi delle peculiarità. I loro tentativi si riassumono nella divisione in tre macroaree corrispondenti a:

1. *Web Surface*, il livello più superficiale che comprende i siti facilmente accessibili e vari circuiti privati;
2. *Deep Web*, a cui si accede solo tramite proxy e in cui sono presenti comunità di hackers, libri di qualsiasi formato e su qualsiasi argomento, ma anche reti internet per compagnie e università;
3. *Dark Web*, un sottoinsieme del Deep Web, a cui è possibile accedere mediante browser speciali, come TOR, I2P o Freenet.

«Deep Web e Dark Web», dice Paganini, tra i massimi esperti di sicurezza italiana a livello internazionale, «sono termini spesso abusati e confusi e di consueto associati ad attività criminali. Con il termine Deep Web si indica l'insieme dei contenuti presenti sul web e non indicizzati dai comuni motori di ricerca, mentre con il

termine Dark Web si indica l'insieme di contenuti che sono ospitati in siti web il cui indirizzo IP è nascosto, ma ai quali chiunque può accedere» tramite browser specifici (Paganini 2012).

Il Deep Web contiene archivi digitali dei dipartimenti pubblici, sistemi di comunicazioni private, sottoreti private come quelle universitarie o governative, contenuti dinamici, pagine ad accesso ristretto, script, dati medici o dei social network, etc., mentre nel Dark Web si è in grado di trovare ogni genere di risorsa e tutti quei contenuti che per vari motivi non posso essere pubblicati in Internet senza conseguenze legali, come merce rubata, traffici d'armi e droga o di bitcoin, la moneta digitale.

Il Web nel quale navighiamo, il Surface Web (o Visible Web o Indexed Web o LightNet) rappresenta solo una minima percentuale dei contenuti reali della rete, mentre sotto la punta dell'Iceberg c'è il Deep Web, di contenuto 500 volte maggiore a quello consultabile in superficie. Secondo i principali esperti di sicurezza mondiale il Surface Web è sotto costante sorveglianza dei governi e rappresenta il 4% dei contenuti del world wide web, mentre il Deep Web ne rappresenta il 96%.

L'*Hidden Web* è oggi utilizzato da un'ampia gamma di persone e non solo per intenti criminali, ma anche per eludere la censura introdotta dai governi in aree critiche del pianeta, per garantire spazi di confronto politico, per organizzare biblioteche di libri proibiti o per la diffusione di informazioni. A gennaio 2016 è uscito il primo numero di *The Torist*, la prima rivista letteraria online pubblicata sul Deep Web, fondata da Gerard Manley Hopkins (G.M.H.) e il Prof. Robert W. Gehl, insieme ad altri utenti. La rivista è lunga 51 pagine e contiene tre racconti, poesie di due autori diversi, due saggi e una prefazione di cinque pagine fatta dagli autori. I temi maggiormente quotati esaminano per lo più il rapporto tra l'uomo e la tecnologia, la privacy e i rischi dell'eccessiva sorveglianza sugli individui e la cultura del Web in generale con un approccio spesso critico e distopico.

Bibliografia consigliata

Echeverri D., *Deep Web: TOR, FreeNET & I2P: privacidad y anonimato*, Madrid, OXWORD, 2016

Florindi E., *Deep web e bitcoin. Vizi privati e pubbliche virtù della navigazione in rete*, Reggio Emilia, Imprimatur, 2016

Frediani C., *Deep Web. La rete oltre Google. Personaggi, storie, luoghi dell'internet profonda*, Genova, Stampa Alternativa, 2014

Meggiato R., *Il lato oscuro della Rete: alla scoperta del Deep Web e del Bitcoin*, Milano, Apogeo, 2014

Paganini P., *The Deep Dark Web: The Hidden World: Volume 1*, Napoli, Apogeo, 2012

Sui D.- Caverlee J.- Rudesill D., *The deep web and the darknet: a look inside the internet's massive black box*, Washington, Wilson Center, 2015, <https://www.wilsoncenter.org/sites/default/files/stip_dark_web.pdf>

Scheeren W. O., *The hidden web. A sourcebook*, Santa Barbara, Libraries Unlimited, 2012

Wojtowicz P., *Darknet and deep web: il lato oscuro del web per la privacy e la protezione dati*, Tesi di Laurea, Università di Bologna, 2013/2014, <https://amslaurea.unibo.it/8456/1/wojtowicz_patryk_tesi.pdf>

Sitografia

AaVv, *Below the Surface: exploring the Deep Web*, 2015, <<https://documents.trendmicro.com/>>

Cybercrime in the Deep Web, <<https://www.blackhat.com/docs/eu-15/materials/eu-15-Balduzzi-Cybercrmine-In-The-Deep-Web-wp.pdf>>

Motori di ricerca scientifici, <<https://www.unipa.it/biblioteche/Cerca-una-risorsa/Motori-di-ricerca-scientifici-/>>

**Web Sommerso (s.v.)*, in *Wikipedia*, <https://it.wikipedia.org/wiki/Web_sommerso>

What is the Deep Web? A first trip into the abyss, <<http://www.aofs.org/wp-content/uploads/2013/04/130925-What-is-the-Deep-Web.pdf>>

Alessandra Di Meglio

(ingl. Acronimo per *HyperText Markup Language*)

È un linguaggio di marcatura che descrive la struttura delle pagine web e definirne i collegamenti ipertestuali, utilizzando una serie di marcatori predefiniti e non modificabili; rappresenta, quindi, uno standard riconosciuto dal *World Wide Web Consortium* (W3C). Come SGML, suo progenitore, e XML, l'*HTML* attua una fondamentale distinzione tra il testo, ovvero il contenuto vero e proprio della pagina web, e i markup che rappresentano, invece, l'ossatura del documento che lo rendono «leggibile» dai browser.

Il linguaggio venne sviluppato da Tim Barners-Lee presso il CERN DI Ginevra (Svizzera) durante i primissimi anni «90 in concomitanza con il protocollo HTTP pensato per il trasferimento dei documenti codificati con tale formato. L'idea, però, risale al 1989, quando lo stesso Lee propose un progetto, denominato World Wide Web, che riguardava la pubblicazione di ipertesti in un ambiente condiviso: i risultati e l'applicazione di tale progetto furono utilizzati solamente all'interno del CERN fino al 1991; generalmente si tende a far risalire anche la nascita di Internet così come lo conosciamo a quella stessa data. Con questa 'liberalizzazione» ebbe inizio una sorta di «guerra dei browser» durante la quale ogni browser esistente utilizzava la propria struttura basata su *HTML* il che, però, rendeva impossibile la corretta visualizzazione di una pagina web su un browser diverso da quello per in cui era stata creata. La questione venne definitivamente risolta nel 1995 quando, un anno dopo la nascita del W3C, venne proposta la versione 3.0 del linguaggio che divenne così uno standard. Nel 1999, con l'*HTML* 4, venne introdotta una novità fondamentale, ovvero la possibilità di creare un foglio di stile, denominato CSS, che fosse esterno al documento *HTML*, in modo da rendere più snella la codifica alleggerendo di conseguenza il documento. L'ultima versione del linguaggio è la 5.0 e risale al 28 ottobre 2014: grazie ad essa ora il linguaggio può avvalersi senza problemi di controlli più sofisticati sulla resa grafica e interazioni dinamiche con l'utente,

avvalendosi di altri linguaggi come il JavaScript, il JSON o altre applicazioni multimediali per l'animazione vettoriale o di streaming audio/video.

Al livello strutturale un documento *HTML* ha bisogno di una dichiarazione iniziale in cui si specificano la versione del linguaggio utilizzata e la conseguente codifica caratteri (che è lo standard UTF-8) e l'uso necessario di determinati marcatori, come `<html>`, `<head>` e `<body>`. A differenza del fratellastro XML, l'*HTML* non dà la possibilità di ampliare la gamma di marcatori utilizzabili, quindi la struttura di qualsiasi file *HTML* risulta grossomodo uguale:

1. Inizia con `<html>` e deve necessariamente chiudersi con `</html>`;
2. successivamente deve essere utilizzato il tag `<head>` nel quale vanno inserite informazioni scontate come il `<title>` o più complesse come quelle necessarie alla visualizzazione del contenuto, definite dal tag `<style>`.
3. Chiuso l'`</head>` viene aperto il `<body>`, che rappresenta il vero «corpo» del documento in cui verrà inserito il testo, le immagini, i link con pagine esterne, eventuali video o animazioni con JavaScript.
4. Il testo viene definito da marcatori prettamente tipografici come `<p>` per i vari paragrafi o `<h1>` `<h2>` `<h3>` per i vari livelli di titolazione. All'interno del testo, poi, sarà possibile inserire tags come `<i>` per il corsivo o `` per il grassetto; notiamo subito come essi siano marcatori prettamente fisici, il che rappresenta una delle fondamentali differenze con l'XML, il quale, invece, si serve di marcatori logici. Ciò non toglie che anche nella sintassi *HTML* è possibile ritrovare marcatori logici come `` (che evidenzia una porzione di testo più estesa del singolo rigo), ma è raro trovarli o utilizzarli.

Bibliografia consigliata

Duckett J., *HTML e CSS. Progettare e costruire siti web. Con Contenuto digitale per download e accesso on line*, Milano, Apogeo, 2017

Tissoni F., *Lineamenti di editoria multimediale*, Milano, Edizioni Unicolpi, 2009

Sitografia

**HTML (s.v.), in w3schools.com, <<http://www.w3schools.com/>>*

AlessiaMarini

L'*Informatica umanistica* o *Digital Humanities* (anche detta *Humanities Computing*) è un settore di ricerca interdisciplinare che incrocia contenuti e metodi propri delle discipline umanistiche con il supporto delle nuove tecnologie informatiche (Catalani 2018); essa coniuga, cioè, il campo tecnologico-informatico, che si occupa del trattamento automatico delle informazioni, con le scienze letterarie, che hanno per oggetto la conoscenza dell'uomo e del suo pensiero.

Tutt'oggi priva di uno statuto epistemologico e di un'identità che la definisca, l'*Informatica umanistica* parte dall'indagine critica dei concetti e dei principi che condizionano l'informazione, le sue dinamiche e l'etica informatica, e risponde a una necessità ben più profonda: conseguire una riflessione consapevole e di alto spessore culturale allo scopo di formare soggetti concretamente – e culturalmente – competenti. Nata dalla Linguistica computazionale e dallo zelo di padre Roberto Busa, che nel 1949 iniziò l'*Index verborum* dell'*opera omnia* di Tommaso d'Aquino, la nuova disciplina ha suscitato fin da subito curiosità e sospetto, finché non ha abbattuto le barriere dei settori «tradizionali» istituzionalizzandosi e moltiplicando i centri, i dipartimenti e i corsi di studio a suo nome. Si è introdotta, in questo modo, la figura accademica dell'umanista informatico, il cui compito, secondo Ramsay, è di «costruire cose», riflettere sull'idea di costruzione e progettare in modo tale che altri possano costruire (Ramsay 2011a) avviando, attraverso il processo di «building», una fase evolutiva che ha trovato negli anni terreno fertile. Tuttavia, la mancata istituzione di un settore scientifico-disciplinare specifico per l'*Informatica umanistica* e la frammentarietà delle iniziative pongono ancora ostacoli al suo consolidamento (Monella 2013). L'opinione comune stabilisce, infatti, che le Digital Humanities non siano una disciplina con un proprio oggetto di indagine, ma semplicemente ricerca umanistica portata avanti con strumenti digitali. Essa, cioè, è considerata, *sic et simpliciter*, mero strumento di supporto dello studio e della ricerca, ottemperando al ruolo di vassallo della disciplina madre. Si ignora invece che l'uso dell'*Informatica umanistica* ha trasformato

la ricerca dall'interno, i suoi obiettivi, l'*iter*, i suoi traguardi qualitativi e l'espansione quantitativa e che, essendo ancora *in fieri* e allo stato sperimentativo, essa chiede all'umanista di riflettere e interrogarsi sul senso dell'uso informatico nelle discipline letterarie, sulle sue possibilità e sui suoi limiti, dando – o togliendo – valore al supporto mediatico. È la pratica quotidiana, infatti, che conduce a una maggiore consapevolezza dei metodi computazionali, degli strumenti digitali e di quelli informatici, con l'impegno di produrre risorse oggi sempre più valide.

Sicuramente lo *status* teorico dell'*Informatica umanistica* pone chiunque tenti di mettere ordine alla sua materia nella scomoda posizione di dover trattare una vastità di dati – e di pregiudizi – impossibili da contenere data la pluridisciplinarietà delle Digital Humanities. Tuttavia riconoscerle un ruolo è oggi importante. Essa infatti rappresenta il futuro delle discipline letterarie, di cui intende potenziare la tradizione culturale senza per questo modificarne l'effettiva natura.

Bibliografia consultata

Flanders J. - Unsworth J., *The Evolution of Humanities Computing Centers*, in *Computers and the Humanities*, XXXVI, 4, 2002, pp. 379-380

Gold M.K., *The Digital Humanities Moment*, in Gold M. W. (a cura di), *Debates in the Digital Humanities*, Minneapolis-London, University of Minnesota Press, 2012, pp. IX-XVI

Bibliografia consigliata

Berry D.M., *Understanding Digital Humanities*, United Kingdom, Palgrave Macmillan, 2012

Bodard G. - Mahony S., *Digital Research in the Arts and Humanities*, Surrey: England, Taylor & Francis Ltd., 2010

Carter B.W., *Digital Humanities: current perspective, practices, and research*, United Kingdom, Emerald Group Publishing Limited, 2013

Gardiner E. - Musto R.G., *The Digital Humanities*, Cambridge, Cambridge University Press, 2015

Numerico T. - Fiormonte D. - Tomasi F., *L'umanista digitale*, Bologna, Il Mulino, 2010

Rydberg-Cox J.A., *Digital libraries and the challenges of digital humanities*, Oxford, Chandos Publishing, 2006

Schreibman S. - Siemens R. - Unsworth J., *A Companion to Digital Humanities*, Oxford, Blackwell Publishing, 2004

Tomasi F., *Metodologie informatiche e discipline umanistiche*, Roma, Carocci, 2008

Tomasi F. - Anselmi G.M., *Informatica e letteratura*, in *Le forme e la storia*, 2011, IV, pp. 163 -182

Tomasi F., *Informatica Umanistica: iniziative, progetti e proposte*, in *La macchina nel tempo. Studi di informatica umanistica in onore di Tito Orlandi*, Firenze, Le Lettere, 2011, pp. 309 - 327

Sitografia

Buzzetti D., *Che cos'è, oggi, l'informatica umanistica? L'impatto della tecnologia*, in Ciotti F. - Crupi G. (a cura di), *Dall'Informatica umanistica alle culture digitali. Atti del convegno di studi (Roma, 27-28 ottobre 2011) in memoria di Giuseppe Gigliozi*, https://www.academia.edu/2305364/Che_cos_e_oggi_l_informatica_umanistica_L_imp

Catalani L., *Informatica Umanistica e Digital Humanities*, <<https://medium.com/@luigicatalani/informatica-umanistica-e-digital-humanities-ff57c44d68be>>

**Informatica* (s.v.), in *Treccani*, <<http://www.treccani.it/enciclopedia/informatica/>>

**Informatica umanistica* (s.v.), in *Wikipedia*, <https://it.wikipedia.org/wiki/Informatica/_umanistica/>

Monella P., *L'informatica umanistica tra istituzionalizzazione e strumentalismo*, http://www1.unipa.it/paolo.monella/lincei/files/where/strumenti_v2.0.pdf

**Umanistico* (s.v.), in *Treccani*, <http://www.treccani.it/vocabolario/umanistico/>

Alessandra Di Meglio

Interfaccia utente (UI)

In termini generali si tratta di un termine generale per ogni sistema, fisico o digitale che permette all'utente l'accesso ad una data tecnologia. È quindi il punto di contatto tra l'uomo e il programma del computer. Ogni *interfaccia utente* (UI) è unica e legata alla funzione che deve svolgere ma tutte hanno in comune certi elementi di base. Una buona UI deve essere user-friendly e quindi permettere un'esperienza d'utilizzo naturale ed intuitiva che non frustri l'utente. Lo sviluppo delle UI ha avuto inizio a partire dalla seconda metà del XX secolo; si può considerare il sistema di schede forate dei primi computer come la prima elaborazione di una UI. Ad oggi, dopo decenni di sviluppo nel campo dell'informatica si è passati dalle *Text-Based UI* alle *Graphical User Interface* (GUI) che permettono una navigazione semplificata all'interno dei software.

Un'interfaccia è composta da un set di comandi, o menù, attraverso i quali si riesce a comunicare con il programma software. Una *UI command-driven* richiede che l'utente inserisca direttamente i comandi; una *menu-driven* permette un sistema di selezione dei comandi attraverso dei menù che compaiono a schermo.

Tra le *menu-driven UI* si trovano le cosiddette GUI. Si tratta infatti di un'elaborazione grafica a schermo che permette all'utente di non interfacciarsi direttamente con il DOS del computer come si era obbligati a fare prima dello sviluppo di questo sistema di elaborazione. Ad oggi questo tipo di interfacce sono considerate lo standard per lo sviluppo di nuovi software dedicati ad un mercato di consumo di massa. Elementi chiave sono: un puntatore, un sistema di puntamento, icone, un desktop, finestre e menù.

Sitografia

**User Interface* (s.v.), in *Tech Terms*, <<https://techterms.com>>

**User Interface* (s.v.), in *webopedia*, <https://www.webopedia.com/TERM/U/user_interface.h>

AntonioMarsonFranchini

Il *Machine Learning* (ML) – in italiano Apprendimento Automatico – è una branca dell'informatica, e in particolare dell'intelligenza artificiale (IA), che realizza algoritmi in grado di imparare autonomamente il compito richiesto – senza essere preventivamente programmati a farlo – e di apprendere informazioni dall'esperienza mediante metodi matematico – computazionali. Tali metodi consentono alla macchina di prendere decisioni liberamente tenendo conto delle probabilità di accadimento di un evento e di modificare gli stessi algoritmi man mano che riceve informazioni. Il primo a coniare il termine *Machine Learning* fu un noto scienziato americano e pioniere dell'IA, Arthur Lee Samuel, nel 1959, anche se la definizione oggi più accreditata è quella di Tom Michael Mitchell, direttore del «Dipartimento Machine Learning» della Carnegie Mellon University, il quale ha affermato che: «si dice che un programma apprende dall'esperienza E rispetto ad una certa classe di compiti T e ad una misura della performance P, se la sua performance nell'eseguire i compiti in T migliora con esperienza E». In pratica, gli algoritmi del *Machine Learning* migliorano le loro prestazioni mediante l'esperienza e gli esempi da cui apprendono.

Le prime sperimentazioni risalgono agli inizi degli anni Cinquanta del Novecento, quando alcuni matematici e statistici pensarono di utilizzare i metodi probabilistici per realizzare macchine intelligenti in grado di prendere decisioni autonomamente. Fu Alan Turing, in particolare, a ipotizzare degli algoritmi specifici che ottemperassero a tale compito indirizzando la ricerca in altra direzione. Infatti, se nella programmazione tradizionale tutta la conoscenza era codificata all'interno del programma e la macchina era in grado di applicare solo le regole preinserite, nell'apprendimento automatico, al contrario, la macchina costruisce da sola le regole – previa introduzione di esempi – e capisce se un nuovo caso risponde o meno alla regola che ha ricavato.

Caratteristica comune ai vari algoritmi di apprendimento automatico è la presenza di dati in input, detti «Training Pattern», che

inducono alla generazione di output, o Istanze dell'esperienza. A seconda dell'algoritmo utilizzato – ossia a seconda delle modalità con cui la macchina impara ed accumula dati e informazioni – si distinguono quattro diversi sistemi di apprendimento automatico identificati dallo stesso Arthur Samuel alla fine degli anni '50: apprendimento supervisionato, non supervisionato, semi-supervisionato e per rinforzo, utilizzati in maniera differente a seconda della macchina su cui si deve operare, per garantire la massima performance e il migliore risultato possibile.

1. L'apprendimento supervisionato (o *Supervised Learning*) consiste nel fornire alla macchina una serie di nozioni specifiche e codificate (cioè input), ossia di modelli ed esempi che le permettono di costruire un vero e proprio database di informazioni e di esperienze, a cui attinge ogniqualvolta si trovi di fronte a un problema, fornendo una risposta (o output) sulla base di esperienze già codificate. In questo tipo di apprendimento la macchina deve solo scegliere quale sia la migliore risposta allo stimolo che le viene dato.

L'apprendimento supervisionato si usa soprattutto per problemi di:

- classificazione, un processo mediante il quale la macchina è in grado di riconoscere e categorizzare oggetti visivi e dimensionali decidendo a quale categoria appartiene un determinato dato;
 - regressione, corrispondente alla capacità della macchina di prevedere il valore futuro di un dato avendo il suo valore attuale (es. la quotazione delle valute o delle azioni di una società).
2. L'apprendimento non supervisionato (o *Unsupervised Learning*) prevede invece che gli input inseriti all'interno della macchina non siano codificati e che sia la stessa macchina a catalogare le informazioni in proprio possesso e a organizzarle, imparandone il significato e soprattutto il risultato a cui portano (output). In questo caso la macchina dovrà organizzare le informazioni in maniera intelligente e dovrà imparare quali sono i risultati migliori per le differenti situazioni che si presentano.

Si ricorre all'apprendimento non-supervisionato per risolvere principalmente due tipologie di problemi:

- raggruppamento (o *clustering*), quando occorre raggruppare dati che presentano caratteristiche simili. In questo caso il

programma non utilizza dati inseriti e categorizzati in precedenza, ma ricava esso stesso la regola raggruppando casi analoghi;

- associazione, che si pone l'obiettivo di trovare schemi frequenti, associazioni o correlazioni tra un insieme di oggetti o item (dati), tentando di formulare regole che ne predicano gli output sulla base di altri item simili. Essa è usata soprattutto nel Data Mining, un processo mediante il quale si estraggono informazioni utili da grandi quantità di dati attraverso metodi automatici o semiautomatici.
3. L'apprendimento semi-supervisionato (o *Semi-Supervised Learning*) prevede di fornire al computer un set di dati (come nell'apprendimento supervisionato), ma incompleti (come in quello non supervisionato), che tocca alla macchina codificare.
 4. L'apprendimento per rinforzo (o *Reinforcement Learning*) rappresenta il sistema di apprendimento più complesso. Esso prevede che la macchina sia dotata di strumenti e procedimenti in grado di migliorare il proprio apprendimento e, soprattutto, di comprendere le caratteristiche dell'ambiente circostante. In questo caso alla macchina vengono forniti una serie di elementi di supporto, quali sensori, telecamere, GPS, che le permettono di rilevare l'ambiente circostante e di effettuare scelte per un migliore adattamento.

Ci sono poi altre sottocategorie di *Machine Learning*, tra cui il Deep Learning (o Apprendimento Profondo), che usa reti neurali – neuroni artificiali tra loro interconnessi che simulano il funzionamento neurale in un sistema informatico – per processare informazioni e comprendere gli schemi presenti nei grandi volumi di dati (o Big Data).

Molti settori riconoscono oggi il valore del *Machine Learning* grazie al quale sono in grado di lavorare con più efficienza e di acquisire un vantaggio competitivo; tra questi: i servizi finanziari (banche e altre aziende), la pubblica amministrazione (enti pubblici di sicurezza o dei servizi), l'assistenza sanitaria, il marketing, i trasporti, etc.

Bibliografia consigliata

Alpaydin E., *Introduction to Machine Learning*, Cambridge, Mit Press, 2009

Bishop C.M., *Pattern recognition and Machine Learning*, Singapore, Springer Verlag, 2006

Chapelle O. - Shölkopf B. - Zien A., *Semi-Supervised Learning*, Cambridge (MA), Mit Press, 2006

Goutte G. *et al.*, *Learning machine translation*, Cambridge (MA), Mit Press Ltd., 2009

Harrington P., *Machine Learning in action*, New York, Manning Publications, 2012

Holmes D.E. - Jain L.C., *Innovations in Machine Learning. Theory and applications*, The Netherlands, Springer, 2006

Kononenko I. - Kukar M., *Machine Learning and Data Mining: Introduction to principles and algorithms*, United Kingdom, Horwood Publishing Chichester, 2007

Marsland S., *Machine Learning: an algorithmic perspective*, Cambridge, Chapman and Hall/CRC, 2009

Powers D.M.W. - Turk C.C.R., *Machine Learning of natural language*, London, Springer, 1989

Sammur C. - Webb G.I., *Encyclopedia of Machine Learning*, New York, Sammur, 2011

Sebe N. *et al.*, *Machine Learning in Computer vision*, The Netherlands, Springer, 2005

Sutton R.S. - Barto A.G., *Reinforcement Learning: an introduction*, Cambridge (MA), Bradford Books, 1998

Sitografia

**Apprendimento Automatico* (s.v.), in *Wikipedia*, <https://it.wikipedia.org/wiki/Apprendimento_automatico>

Mitchell T.M., *Machine Learning*, Maidenhead 1997, <<https://www.cs.ubbcluj.ro/~gabis/ml/ml-books/McGrawHill%20-%20Machine%20Learning%20-Tom%20Mitchell.pdf>>

Nilsson N.J., *Introduction to Machine Learning. An early draft of a proposed textbook*, Stanford 1998 <<https://ai.stanford.edu/~nilsson/MLBOOK.pdf>>

Shai S.S.-Shai B.D., *Understanding Machine Learning. From theory to algorithms*, USA 2014 <<https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>>

Stefenelli M., *Apprendimento automatico nei giochi di strategia*, Tesi di Laurea, Università di Bologna Campus di Cesena, 2013/2014 <https://amslaurea.unibo.it/7660/1/TESI_Marco_Stefenelli.pdf>

AlessandraDiMeglio

Si tratta di uno dei due sistemi di elaborazione del testo elettronico: si contrappone al WYSIWYG (*What You See Is What You Get*) e, per via delle sue caratteristiche, non è legato ad un determinato tipo di programma che ne garantisca la lettura. In italiano il termine si può tradurre con linguaggio di «marcatatura» o di «codifica». Permette di descrivere la struttura, composizione, impaginazione del testo (o di qualsiasi altro file di contenuto) attraverso comandi visibili (sequenze di caratteri ASCII) immessi nel file di testo.

Se ne possono distinguere due tipi: «procedurali» (basati su un linguaggio specifico) e «dichiarativi» (basati su un linguaggio generico). Il primo consiste in un insieme di istruzioni operative che indicano localmente la struttura tipografica e compositiva della pagina. Viene così definito perché indica alla macchina le procedure di trattamento a cui sottoporre i file di testo. Ha come obiettivo principale l'aspetto tipografico del documento. Il secondo, invece, ha come obiettivo la descrizione della struttura astratta del documento in cui i simboli al posto che indicare il trattamento tipografico della pagina ne identificano le strutture interne (paragrafi, etc.). Entrambi utilizzano un sistema di tag (marcatori) che identificano le parti del discorso all'interno del file di testo ma solo i linguaggi di tipo «dichiarativo» si occupano di identificarne le qualità logiche e non tipografiche. Questo sistema di marcatori viene sempre utilizzato in una forma gerarchizzata che dall'elemento più grande da definire giunge fino al più piccolo.

I linguaggi di *markup* sono lo strumento standard per lo sviluppo di oggetti digitali. Il capostipite di questi linguaggi è SGML (*Standard Generalized Markup Language*) creato nel 1986 da Charles Goldfarb e risultato tanto funzionale da diventare lo standard ISO (*International Standard Organization*) per la codifica dei testi. Ad oggi, nell'ambito dell'informatica umanistica lo standard è il linguaggio di *markup* TEI (*Text Encoding Initiative*) basato sullo standard XML (*eXtensible Markup Language*).

Antonio Marson Franchini

Con il termine *metadato*, mutuato dall'inglese *metadata* – costituito dal greco μέτα (*meta*: oltre, dopo) e dal plurale neutro latino *data*, informazioni – si indica l'insieme delle informazioni strutturate relative ai dati.

L'origine inglese del termine influenza tenacemente l'uso, tanto che la forma singolare *metadato* (ricavata all'interno del sistema morfologico italiano) è di utilizzazione più infrequente.

I *metadati* sono legati al ciclo di vita di una risorsa digitale nel senso che ogni oggetto/risorsa digitale necessita di un bagaglio di dati che, pur non essendo parte del suo contenuto informativo, è indispensabile a renderlo rappresentabile attraverso specifiche procedure di decodifica. Tali informazioni associate, dati sui dati, prendono appunto il nome di *metadati*.

I *metadati* sono quindi informazioni strutturate che descrivono, spiegano, localizzano o comunque rendono possibile il recupero, l'uso e la gestione di una risorsa informativa. Le relazioni tra risorse informative e *metadati*, oltre ad essere indispensabili, si prestano a svariate declinazioni: i *metadati* possono essere statici o dinamici; «embedded» – ossia inclusi nella risorsa come parte integrante – o collegati ad essa in virtù di meccanismi identificativi stabili e sicuri nel tempo; un medesimo oggetto può prevedere l'associazione a diversi schemi di *metadati* come diversi e più oggetti possono essere associati tra loro attraverso *metadati*. Inoltre, il ciclo di vita di una risorsa digitale deve essere accompagnato dal costante aggiornamento del *corpus* dei relativi metadati, affinché tutto quanto accade resti tracciato e documentato.

La comunità scientifica internazionale, preso atto della necessità di associare alle entità digitali un insieme di *metadati* con i caratteri di completezza, accuratezza e appropriatezza, ha elaborato alcuni schemi e modelli a cui fare riferimento, proponendo una distinzione in tre categorie funzionali:

1. *metadati descrittivi (descriptive metadata)*: finalizzati all'identificazione e al recupero degli oggetti digitali; sono costituiti

da descrizioni normalizzate dei documenti fonte (o dei documenti digitali nativi), risiedono generalmente nelle basi dati dei sistemi di *Information Retrieval* all'esterno degli archivi degli oggetti digitali e sono collegati a questi ultimi con link dedicati;

2. *metadati strutturali (structural metadata)*: volti a descrivere la struttura interna dei documenti e a gestire le relazioni esistenti tra le varie parti componenti degli oggetti digitali;
3. *metadati amministrativo-gestionali*, funzionali alla gestione degli oggetti digitali all'interno dell'archivio e alla cura degli aspetti più propriamente tecnici.

A complemento ed integrazione dell'indicata tripartizione vi sono poi i *metadati* per la gestione dei diritti, i *metadati* per la security, i *metadati* per le informazioni personali, i *metadati* per la conservazione a lungo termine, le cui funzioni compaiono diversamente esplicitate e valorizzate in specifici modelli e schemi.

In letteratura il riferimento più noto e ricorrente in materia di *metadati* di valenza descrittiva è lo standard *ISO 15836*, comunemente noto come *Dublin Core (DC, in realtà Dublin Core Metadata Initiative, DCMI)*.

Il *Dublin Core* è un insieme di *metadati*, poi implementato, progettato per la descrizione di una qualunque risorsa informativa, indipendentemente dal dominio di appartenenza; la sua vocazione è quella di integrarsi con altri standard di *metadati*, anche all'interno di una stessa descrizione, così da ottenere un documento descrittivo complesso ed eterogeneo, che possa essere funzionale alla descrizione, alla catalogazione, alla ricerca e all'individuazione di risorse informative differenti per natura, per tipologia e per contesto d'uso.

Difficile realizzare una rassegna sistematica dei diversi modelli tecnici di metadatazione e individuare tra essi quelli che abbiano caratteristiche tali da assurgere ad effettivi standard; si segnalano quindi, senza alcuna pretesa di esaustività:

1. *Metadata Object Description Schema (MODS)*: schema di *metadati* descrittivi derivato dallo standard bibliografico MARC e indicato per la sua flessibilità a descrivere oggetti digitali nativi, con un livello di granularità ampiamente compatibile con gli standard dei formati bibliografici;

2. *Metadata Encoding and Trasmission Standard* (METS): standard per la codifica dei *metadati* descrittivi, amministrativi e strutturali di un oggetto digitale. Un documento METS è caratterizzato da ampia compatibilità con gli altri schemi di *metadati* e ciò rende questo standard particolarmente flessibile e applicabile in diversi contesti;
3. *Metadati Amministrativi e Gestionali* (MAG): il progetto sotteso è finalizzato all'elaborazione, al mantenimento e all'evoluzione di uno standard italiano di *metadati* amministrativi e gestionali. Il profilo applicativo MAG è stato adottato dall'Istituto centrale per gli archivi (ICAR) in qualità di standard nazionale sperimentato per la gestione, conservazione e diffusione via Web delle riproduzioni digitali dei documenti d'archivio. Il comitato MAG, inoltre, nelle fasi di studio ed evoluzione ha destinato grande attenzione alla compatibilità con il modello METS;
4. *Open Archival Information System* (OAIS, standard ISO 14721): il modello prevede che le informazioni vengano organizzate per componenti funzionali e pacchetti informativi, individuando almeno cinque categorie di elementi informativi finalizzati a documentare e realizzare la conservazione a medio e lungo termine e la relativa accessibilità;
5. *Preservation Metadata: Implementation Strategies* (PREMIS): si tratta dello standard più utilizzato per i *metadati* di conservazione; tra i suoi punti di forza si annoverano la facilitazione nello scambio di oggetti digitali tra strutture di conservazione e l'integrazione con modelli non-PREMIS ottenibile con l'uso dei cosiddetti contenitori di estensione.

Lo stretto, e più volte richiamato, legame esistente tra l'oggetto digitale e i *metadati* che ne rappresentano le caratteristiche, nonché il ruolo che i *metadati* svolgono nelle diverse fasi di vita della risorsa condizionano la qualità complessiva di un prodotto digitale e la sua capacità di rispondere nel tempo alle finalità che ne hanno determinato la creazione. La scelta degli standard di *metadati* da utilizzare diventa poi determinante quando si vogliano integrare collezioni di risorse con sistemi nazionali o internazionali. Pertanto «i metadati sono al contempo il mastice che tiene insieme le informazioni sugli oggetti digitali, garantendone la qualità e l'accesso, e la struttura logica che consente la creazione e produzione di nuove e inedite relazioni tra i dati» (Crupi, 2015).

Bibliografia consigliata

Crupi G., *Biblioteca digitale*, in Solimine G. - Weston P.G. (a cura di), *Biblioteche e biblioteconomia. Principi e questioni*, Roma, Carocci, 2015, pp. 373-417

Feliciati P., *Gestione e conservazione di dati e metadati per gli archivi: quali standard?*, in Pigliapoco S. (a cura di), *Conservare il digitale*, Macerata, EUM, 2010, pp. 191-219

Feliciati P., *I metadati nel ciclo di vita dell'archivio digitale e l'adozione del modello PREMIS nel contesto applicativo nazionale*, in Bonfiglio Dosio G. - Pigliapoco S. (a cura di), *Formazione, gestione e conservazione degli archivi digitali. Il Master FGCAD dell'Università degli Studi di Macerata*, Macerata, EUM, 2015, pp. 189-208

Franzese P., *Manuale di archivistica italiana*, Perugia, Morlacchi Editore U.P., 2014

Guercio M., *Archivistica informatica: i documenti in ambiente digitale*, Roma, Carocci, 2013

Guercio M., *Conservare il digitale. Principi, metodi e procedure per la conservazione a lungo termine di documenti digitali*, Roma-Bari, Laterza, 2013

Guercio M., *La conservazione delle memorie digitali*, in Solimine G. - Weston P.G. (a cura di), *Biblioteche e biblioteconomia. Principi e questioni*, Roma, Carocci, 2015, pp. 545-566

Michetti G., *Gli standard per la gestione documentale*, in Giuva L. - Guercio M., *Archivistica. Teorie, metodi, pratiche*, Roma, Carocci, 2014, pp. 263-286

Pigliapoco S., *Progetto archivio digitale. Metodologia sistemi professionalità*, Lucca, Civita editoriale, 2016

Weston P.G. - Sardo L., *Metadati*, Roma, Associazione Italiana Biblioteche, 2017

Sitografia

Data Dictionary for preservation Metadata, 3, 2015, <www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

Linee guida per la digitalizzazione e metadati, in ICCU – Istituto Centrale per il Catalogo Unico, *delle Biblioteche italiane e per le informazioni*

bibliografiche, <<http://www.iccu.sbn.it/opencms/opencms/it/main/standard/metadati>>

METS – Metadata Encoding & Transmission Standard,

<www.loc.gov/standards/mets/METSita.html>

Pierazzo E. (a cura di), *MAG – Metadati Amministrativi e Gestionali. Manuale Utente*,

<<http://www.iccu.sbn.it/upload/documenti/Manuale.pdf>>

Metadata Object Description Schema, <<http://www.loc.gov/standards/mods/>>

**OAIS Reference Model*, in *Conservazione Digitale. Centro di eccellenza italiano sulla Conservazione Digitale*, <<http://www.conservazionedigitale.org/wp/approfondimenti/depositi-di-conservazione/oais-reference-model>>

ConcettaDamiani

N-gramma è una nozione matematica concettualmente semplice: data una sequenza ordinata di elementi, un *n-gramma* ne rappresenta una sottosequenza di *n* elementi. Ma le applicazioni di questo concetto matematico, da cui è possibile elaborare modelli (*n-gram models*), sono straordinarie ed afferiscono a campi anche molto differenti, come la statistica, la teoria della comunicazione, la linguistica computazionale (settore interdisciplinare prima ancora che «scienza» a sé: cfr. Tamburini 2008) *et alia*. In particolare, in campo linguistico, un *n-gramma* è una sequenza di due (bi-gramma), tre (tri-gramma) o più (n-gramma) parole-chiave presenti all'interno di un dato, specifico contesto; più semplicemente, si potrebbe affermare che gli *n-grammi* sono «gruppi» di parole che compaiono insieme in un testo.

L'uso degli *n-grammi* si rivela molto utile, ad esempio, nel calcolo delle co-occorrenze, il che – in altri termini – consente di lavorare sulla intertestualità. Il calcolo delle co-occorrenze tramite *n-grammi* non tiene conto dei seguenti elementi: 1. presenza di segni di interpunzione; *stop-words* (cioè parole semanticamente vuote e o parole che, essendo comuni, se isolate, non restituiscono un particolare significato); 3. *ordo verborum*. Per fornire un esempio concreto, che coniughi il concetto di *n-grams* e le digital humanities, basterà chiamare in causa, ancora una volta, quello straordinario strumento che è il TLG, che dispone della specifica, interessante funzione *N-grams* (cfr. Battaglini 2019): si tratta di una nuova funzione (un update del nuovo sito del TLG), grazie alla quale il TLG, «lavorando» sostanzialmente sui trigrammi (eccezion fatta per le opere di cui possediamo solo frammenti, per i quali il TLG lavora sui bigrammi, in considerazione della loro *brevitas*), consente di individuare co-occorrenze. Le co-occorrenze possono essere individuate (con evidenti, significative ricadute per gli studi letterari, filologici, lessicografici): 1. tra due testi di un medesimo autore; 2. tra due edizioni critiche di uno stesso testo (ove disponibili e digitalizzate, naturalmente); 3. tra testi di due autori diversi, con diverse possibilità di interrogazione: a. un testo per ciascuno dei

due autori; b. un testo per il primo autore, tutti i testi (digitalizzati) per il secondo autore; c. tutti i testi (digitalizzati) per il primo autore, un testo per il secondo autore; d. tutti i testi (digitalizzati) di entrambi gli autori in questione. Oltre alla specifica funzione *N-Grams*, il TLG consente di «sfruttare» gli *N-Grams* anche nell'ambito della funzione «Browse»: ciò si rivela ancor più interessante sul piano dello studio dell'intertestualità, giacché consente, a partire da un dato testo di individuare i paralleli tra esso e l'intero *corpus* del TLG (nel caso di accesso al *full corpus*).

Tra le numerose applicazioni linguistico-filologiche degli *n-grammi*, va ricordata anche la possibilità di adoperarli per l'attribuzione della paternità testuale di uno specifico testo, ad esempio studiando la presenza, la frequenza e la distribuzione di specifici *n-grammi* accuratamente individuati e selezionati (cfr., a tal proposito, per la letteratura italiana: Basile-Lana 2008; Basile *et alii* 2008; Boschetti 2018; per la letteratura greca: Gorman-Gorman 2016; per una panoramica generale e concisa: Battaglini 2018). Gli *n-grammi* possono essere utilizzati anche per determinare il «peso» culturale di uno specifico periodo su un dato autore: ciò è reso possibile dal fatto che le idee sono espresse da sequenze di parole, che possono essere trattate come *n-grammi*. A tal proposito, si rimanda all'interessante e recente contributo di Knight-Tabrizi (2016), i quali, tra l'altro, fanno notare che, a seguito della vasta opera di digitalizzazione (cfr. Caterino 2013), «Google has created a database of *n-grams* extracted from these digitized books and has made the database available to researchers online. This is the first time ever that such an extensive repository of cultural data has been made available».

Bibliografia consigliata

Basile C. - Benedetto D. - Caglioti E. - Degli Esposti M., *An example of mathematical authorship attribution*, in *Journal of Mathematical Physics*, 49, 2008, pp. 2-20, <https://www.researchgate.net/publication/228663468_An_example>

Basile C. - Lana M., *L'attribuzione di testi con metodi quantitativi: riconoscimento di testi gramsciani*, in *AIDA Informazioni*, 1-2, 2008, pp. 165-183, <https://www.academia.edu/6604207/Lattribuzione_di_testi_con_metodi_quantitativi>

Battaglini G., *La tessitura matematica dei testi. Filologia e metodi matematico-statistici: l'auspicabile dialogo tra metodi qualitativi*

e metodi quantitativi, in *FRI – Filologia Risorse Informatiche (Carnet de recherche and online journal – Italian Digital Humanities)*, <<https://fri.hypotheses.org/915>>

Battaglino G., *Note minime sul TLG: brevi cenni sulle «origini» del TLG e piccolo vademecum*, in *FRI – Filologia Risorse Informatiche (Carnet de recherche and online journal – Italian Digital Humanities)*, <<https://fri.hypotheses.org/1391>>

Boschetti F., *Copisti digitali e filologi computazionali*, Roma, CNR Edizioni, 2018, <<http://eprints.bice.rm.cnr.it/17545/1/bookBoschetti2018.pdf>>

Caterino A.F., *Note minime su Google Books*, in *FRI – Filologia Risorse Informatiche (Carnet de recherche and online journal – Italian Digital Humanities)*, <<https://fri.hypotheses.org/128>>

Gorman V. B. - Gorman R. J., *Approaching Questions of the Text Reuse in Ancient Greek using Computational Syntactic Stylometry*, in *Open Linguistic*, 2, 2016, pp. 500-510, <<https://www.degruyter.com/downloadpdf/j/opli.2016.2.issue-1/opli-2016-0026/opli-2016-0026.pdf>>

Knight G. P. - Tabrizi N., *Using n-Grams to Identify Time Periods of Cultural Influence*, in *Journal on Computing and Cultural Heritage*, 9, 3, 2016, art. n. 15, pp. 1-19, <<https://dl.acm.org/citation.cfm?id=2940332>>

Tamburini F., *La linguistica computazionale: un crogiolo di esperienze multidisciplinari*, in *GRISELDAONLINE*, 2008, pp. 1-11, <<http://www.griseldaonline.it/informatica/la-linguistica-computazionale-tamburini.html>>

Sitografia

**Google Ngram Viewer* (s.v.), in *Wikipedia* (EN), <https://en.wikipedia.org/wiki/Google_Ngram>

**Intertestualità* (s.v.), in *Treccani*, <<http://www.treccani.it/vocabolario/intertestualita/>>

**N-gram* (s.v.), in *Wikipedia* (EN), <<https://en.wikipedia.org/wiki/N-gram>>

**N-gramma* (s.v.), in *Wikipedia*, <<https://it.wikipedia.org/wiki/N-gramma>>

**Stop words* (s.v.), in *Wikipedia* (EN), <https://en.wikipedia.org/wiki/Stop_words>

**TLG (sub-corpus del) in open access: Canon of Greek Authors and Works (Abridged TLG)*, <<http://stephanus.tlg.uci.edu/Iris/canon/csearch.jsp>>

**TLG* (ultima versione digitale del *TLG: Thesaurus Linguae Graeciae. Digital Library*. Ed. Maria C. Pantelia. University of California, Irvine), <<http://stephanus.tlg.uci.edu/>>

**Thesaurus Linguae Graeciae (s.v.), in Wikipedia, <https://en.wikipedia.org/wiki/Thesaurus_Linguae_Graeciae>*

Giovanna Battaglino

Si definisce *oggetto digitale* qualsiasi testo, documento, immagine, audio/video, ipertesto o base di dati che venga archiviato in una memoria di massa attraverso una codifica che lo renda comprensibile alla macchina e restituibile all'utente umano. La codifica avviene, a causa della sua natura informatica, attraverso il codice binario e un linguaggio (Unicode o ASCII tra i più usati) che rendano l'informazione intellegibile alla macchina. Sua caratteristica peculiare è di essere un'entità composta in modo inseparabile da uno o più file di contenuti e i loro metadati uniti, fisicamente e/o logicamente tramite l'uso di un «digital wrapper», un connettore digitale. Si tratta quindi, in definitiva, dell'associazione del dato e del suo metadato che concorrono a rendere quella sequenza di bit, comprensibile dalla macchina, significativa, individuabile e accessibile per la fruizione, l'archiviazione, la conservazione, la disseminazione e le altre operazioni gestionali. L'*oggetto digitale* può nascere direttamente digitale o aver subito una trasformazione in formato digitale. A sua volta questa entità composita può essere oggetto di un ulteriore livello di specificazione che la distingue in «oggetti digitali semplici» e «complessi». Secondo il glossario della California Digital Library, i primi sono composti da un unico file di contenuto (e dalle sue varianti di formato e forme derivate) e dai suoi metadati mentre secondi sono composti da due o più file di contenuto (e dalle loro varianti di formato e forme derivate) e dai metadati corrispondenti.

L'*oggetto digitale* possiede inoltre come caratteristica la difficoltà di conservazione per diversi motivi, due in particolare: il degrado del supporto fisico (floppy disc e CD ad esempio) e l'obsolescenza degli standard di codifica utilizzati all'atto della creazione dell'oggetto digitale. È per questi due motivi che l'*oggetto digitale* deve essere affrontato come: «oggetto fisico», «oggetto logico», «oggetto concettuale», «collezione di elementi». Tutti questi aspetti devono essere tenuti in considerazione all'atto di creazione dell'*oggetto digitale* perché tutti questi devono essere oggetto della conservazione sia singolarmente che in relazione tra loro.

Da questa sommaria descrizione si comprende che, parlando di *oggetto digitale*, non ci si riferisce a un elemento definito ma, più precisamente ad una categoria di oggetti che possono presentare caratteristiche molto diverse fra loro ma che hanno un unico punto in comune, la necessità di essere archiviate attraverso l'impiego di memorie di massa.

Bibliografia consigliata

Tomasi F., *Metodologie informatiche e discipline umanistiche*, Roma, Carocci, 2008

Numerico T.- Vespignani A., *Informatica per le scienze umanistiche*, Bologna, Il Mulino, 2003

Fiormonte D. - Numerico T. - Tomasi F., *L'umanista digitale*, Bologna, Il Mulino, 2010

Sebastiani M., *Il «documento digitale»: analisi di un concetto in evoluzione*, in *DigItalia. Rivista del digitale nel beni culturali*, 1, 2008, pp. 9-31

Sitografia

**Digital Wrapper*, in *CDL – California Digital Library's Glossary*,
<<http://www.cdlib.org/gateways/technology/glossary.html?field=glossary&action=search&qu>>

Pierazzo E. (a cura di), *MAG – Metadati Amministrativi e Gestionali. Manuale Utente*,

<<http://www.iccu.sbn.it/upload/documenti/Manuale.pdf>>

Antonio Marson Franchini

(Ingl. *Ontology*, fr. *Ontologie*)

Ontologia è un termine mutuato dalla filosofia, che in informatica assume il significato di «rappresentazione di una forma di conoscenza o parte del mondo» da modellare in un programma (dominio). Nei sistemi di condivisione della conoscenza, l'*ontologia* ha il compito di rappresentare «le cose che esistono» in un determinato dominio concettuale, distinguendo i concetti (concettualizzazione), definendoli e indicando il sistema di relazioni entro le quali interagiscono (specificazione). Le *ontologie* sono importanti per il Web Semantico, in quanto permettono di cercare i concetti piuttosto che stringhe dei caratteri e perché creano una relazione tra il contenuto e altre fonti di conoscenza. Il linguaggio solitamente utilizzato è OWL (*Ontology Web Language*), basato sulla sintassi XML.

Esempio: Clavius (*Istituto di Linguistica Computazionale Zampolli di Pisa*) un insieme di manoscritti digitalizzati conservati presso l'Archivio Storico della Pontificia Università Gregoriana, relativi a Christophorus Clavius (1538-1612), matematico ed astronomo gesuita. I concetti (terminologia ed entità di dominio) formano un'*ontologia* legata a risorse già disponibili in rete.

FlaviaSciolette

Il trasferimento dei cataloghi cartacei in banche dati digitali, a metà degli anni Ottanta, ha comportato la diffusione su larga scala degli *OPAC*, acronimo di *Online public access catalogue* (o catalogo in linea ad accesso pubblico). Se le prime biblioteche digitalizzate erano consultabili soltanto tramite «telnet» – un protocollo di rete molto diffuso prima dell'espansione di Internet –, oggi sono accessibili direttamente dal Web.

La consultazione di un *OPAC* avviene mediante la compilazione di uno o più campi di ricerca – ad esempio l'opera e/o l'autore – o l'inserimento di una o più parole chiavi che consentono l'accesso all'elenco dei testi contenuti nella banca dati della biblioteca, informando l'utente sul tipo di documento, sull'anno di pubblicazione, sulla sua collocazione, etc.

Si distinguono varie tipologie di *OPAC*:

1. *OPAC* di singole biblioteche, che occorre consultare mediante il link della biblioteca;
2. *OPAC* collettivi, ossia cataloghi online che contengono i database di più biblioteche (contemporaneamente consultabili) e offrono l'enorme vantaggio di catalogare una volta sola un documento sì che, mediante un'unica descrizione, si ha accesso alle diverse biblioteche che lo possiedono;
3. multi – e meta – *OPAC*, che interrogano più cataloghi online già esistenti in forma indipendente: i multiopac sono interrogabili uno per volta, anche se attraverso la stessa interfaccia; i metaopac sono in grado di interrogare più *OPAC* mediante una sola richiesta, e di restituire i risultati dei diversi *OPAC* coperti. Un esempio italiano di metaopac è il MAI (*Metaopac Azalai Italiano*), <http://www.aib.it/progetti/opac-italiani/mai-ricerca-globale/>.

Per facilitare la ricerca alcune pagine Web raccolgono link di *OPAC*, sia italiani che stranieri (ad esempio il *Gateway to Library Catalogs*:

<https://www.loc.gov/z3950/gateway.html>), nei quali sono contenuti un gran numero di cataloghi. Tra questi vi è l'italiano *OPAC SBN* (*Servizio Bibliotecario Nazionale*), ossia una rete di biblioteche italiane promossa dal MiBAC, dalle Regioni e dalle Università, organizzata dall'ICCU (*Istituto Centrale per il Catalogo Unico*), che interroga più biblioteche contemporaneamente (cfr. OPAC collettivi).

I vantaggi ottenuti dall'introduzione degli *OPAC* sono dunque evidenti. Essi:

1. aiutano a trovare rapidamente i documenti presenti nelle biblioteche;
2. consentono di conoscere la disponibilità di un documento in tempo reale;
3. permettono la prenotazione di un libro;
4. consentono l'accesso a versioni elettroniche di documenti cartacei, nonché a scansioni fotografiche, etc.

È indubbio che tali condizioni semplificano l'indagine e ottimizzano i tempi della ricerca; consentono l'accesso a una banca dati di enormi dimensioni e il reperimento di testi anche di limitata diffusione, conferendo all'*OPAC* un ruolo di imprescindibile utilità e promuovendolo a necessario strumento delle Digital Humanities.

Bibliografia consigliata

Dhanavandan S. - Isabella Mary A., *Online public access catalogue (OPAC)*, New Delhi, Write and Print publications, 2015

IFLA, *Guidelines for online public access catalogue (OPAC) displays*, in *Series on Bibliographic Control*, 27, 2005

Nelson Bonnie R., *Opac Directory 1998: a guide to Internet-Accessible online public access catalogs*, Medford (NJ), Information Today Inc., 1997

Tramullas J. - Garrido P., *Library automation and OPAC 2.0: information access and services in the 2.0 landscape*, Hershey (PA), Information Science Reference, 2012

Sitografia

AIB Web, Il Web dell'Associazione Italiana Biblioteche: MAI, <<http://www.aib.it/progetti/opac-italiani/mai-ricerca-globale/>>

Gateway to Library Catalogs, <<https://www.loc.gov/z3950/gateway.html>>

Gnoli Claudio, *Opac in Italia: una panoramica delle tipologie e delle modalità di consultazione*, <http://www.aib.it/aib/sezioni/emr/bibtime/num-ii-1/gnoli.htm#nota1>

**OPAC* (s.v.), in *Wikipedia*, <<https://it.wikipedia.org/wiki/OPAC>>

OPAC SBN, Catalogo del Servizio Bibliotecario Nazionale, <<https://opac.sbn.it/opacsbn/opac/iccu/informazioni.jsp>>

**Telnet* (s.v.), in *Wikipedia*, <<https://it.wikipedia.org/wiki/Telnet>>

Alessandra Di Meglio

Open Access (OA) è il termine comunemente utilizzato per indicare un movimento che promuove la libera circolazione e l'uso non restrittivo dei risultati della ricerca e del sapere scientifico. SCOPO PRIMARIO DELL'OA è riguadagnare possesso della comunicazione scientifica, offrendo libero accesso ai risultati della ricerca.

La letteratura *Open Access* (OA) è digitale, online, gratuita e libera da alcune restrizioni dettate dalle licenze per i diritti di sfruttamento commerciale. Queste condizioni sono possibili grazie ad Internet e al consenso dell'autore o del titolare dei diritti d'autore.

La definizione emerge dalle risultanze di tre conferenze che hanno avuto luogo tra il 2002 e il 2003:

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities;
- Bethesda Statement on Open Access Publishing;
- Budapest Open Access Initiative.

Nel 2004 la *Conferenza dei Rettori delle Università Italiane* (CRUI) ha formalizzato l'adesione alla *Dichiarazione di Berlino* con la *Dichiarazione di Messina*, Documento italiano a sostegno della Dichiarazione di Berlino sull'accesso aperto alla letteratura accademica.

Come facilmente intuibile, la questione economica non rappresenta l'aspetto di maggior rilevanza per la comunità scientifica internazionale. Il concetto di apertura, infatti, non si limita alla gratuità dell'accesso ma richiama questioni di tipo etico, come la libertà di uso e riuso dei contenuti pubblicati e dei prodotti della ricerca e questioni di effettiva condivisione, auspicando alla realizzazione di «uno spazio globale, multidisciplinare e perfino oltre i confini di università e laboratori, includendo tutti i cittadini».

Bibliografia consigliata

Castellucci P., *Carte del nuovo mondo: banche dati e open access*, Bologna, Il Mulino, 2017

Di Donato F., *La scienza e la rete. L'uso pubblico della ragione nell'età del web*, Firenze, Firenze University Press, 2009

Sitografia

Suber P., *Breve introduzione all'accesso aperto*, traduzione italiana di Susanna Mornati, 2004, <<http://www.iss.it/binary/bibl/cont/Suber.1110468892.pdf>>

ConcettaDamiani

Definizione

Il concetto di *paratesto* appartiene alla teoria della letteratura e fu introdotto dal critico francese Gérard Genette. Secondo il Vocabolario della Lingua Italiana Garzanti il *paratesto* è l'insieme degli elementi accessori e complementari di un testo a stampa, come il titolo, l'introduzione, le note, ed anche l'insieme delle caratteristiche editoriali e tipografiche. L'Enciclopedia Treccani on-line aggiunge che il termine deriva dal francese *paratexte* ed indica l'insieme di produzioni, verbali e non verbali, sia nell'ambito del volume (il nome dell'autore, il titolo, una o più prefazioni, le illustrazioni, i titoli dei capitoli, le note), sia all'esterno del libro (interviste, conversazioni, corrispondenze, diari), che accompagnano il testo vero e proprio e ne guidano il gradimento e, aggiungiamo, la comprensione del pubblico.

Genette teorizza la nozione di *paratesto* in *Seuils*, Paris, Seuil, 1987 (edizione italiana a cura di Camilla Maria Cederna, Soglie. I dintorni del testo, Torino, Einaudi, 1989).

Etimologia

Secondo il Vocabolario on-line Treccani il termine *paratesto* è formato dalla preposizione greca *para-* («vicino», «affine», «contrapposto») e dal sostantivo *testo* (dal latino *textus*, tessuto, da *texere*, tessere, intrecciare) sul modello del francese *paratexte*.

Parti

Il genere *paratesto* viene suddiviso in base alla sua ubicazione in *peritesto* ed *epitesto*.

Peritesto

Il *peritesto* si trova nello spazio dell'opera, con una funzione paratestuale quasi esclusivamente di presentazione, di indirizzo e di commento. È il nucleo del *paratesto*, ha forma e, generalmente,

posizioni fisse: all'inizio (frontespizi, titoli, dediche, epigrafi, prefazioni), in margine (note, chiose) e alla fine del testo (postfazioni, tavole, appendici).

Appartengono al genere *peritesto* anche gli elementi direttamente dipendenti dall'editore: il formato e la composizione grafica, la collana, la copertina, le pagine bianche del volume all'inizio e alla fine del libro (sguardie o risguardi: separano il blocco delle pagine dalla copertina e lo tengono ad essa legato), l'occhiello (pagina che riporta il titolo del libro, se posta prima del frontespizio, o il titolo di un capitolo o simili, se posta all'interno), le pagine 4-5-6 con le ulteriori indicazioni editoriali, il frontespizio (pagina all'inizio del libro, nella quale sono indicati l'autore, il titolo, le note tipografiche), il *colophon* nelle pagine finali (nei libri moderni la formula «finito di stampare» con i dati d'obbligo: la data, il luogo di stampa, il nome dello stampatore e altre notizie simili alla fine dell'opera), la composizione tipografica, la qualità della carta. Vengono classificati come *peritesti* anche il nome dell'autore, il titolo dell'opera, le dediche e le epistole dedicatorie, le epigrafi (le citazioni in margine al testo, per esempio all'inizio del volume o dei capitoli), le prefazioni e le postfazioni, gli intertitoli (i titoli di capitoli o tomi o sezioni), le note. Infine, è considerato *peritesto* il cosiddetto *prière d'insérer*, *paratesto* della consuetudine editoriale francese, un breve prospetto dell'opera, spedito ai direttori dei giornali allo scopo di pubblicarlo, una specie di pubblicità, e poi evolutosi in varie forme, fino all'odierno riassunto nella quarta pagina di copertina.

Epitesto

L'*epitesto* è qualsiasi elemento paratestuale non contiguo, ma comunque in relazione al testo: tutto quello che un autore dice o scrive nella sua vita o sul mondo che lo circonda può avere un carattere paratestuale.

L'*epitesto* si distingue in pubblico e privato. Appartengono al primo genere l'*epitesto* editoriale con funzione promozionale (pubblicità, inserzioni), l'allografo ufficioso (recensioni, interventi critici), l'*epitesto* autonomo (le risposte e i commenti pubblici dell'autore) o mediatizzato (interviste, convegni, occasioni pubbliche di rapporto con l'autore).

L'*epitesto* privato è rivolto ad una persona, che si interpone tra l'autore e il pubblico e può essere un corrispondente, un confidente o anche l'autore stesso. L'*epitesto* privato si suddivide in confidenziale, quando è rivolto a un confidente, ed intimo, se il destinatario è l'autore stesso.

Teoria retorica

In *Der humanistische Geleittext als Paratext – am Beispiel von Brants Beigaben zu Tennglers Layen 'Spiegel'* Joachim Knappe include nel *paratesto* il testo di presentazione umanistico, che aumenta il valore del testo introdotto. A nostro parere confermano questa interpretazione per esempio anche le lettere dedicatorie erasmiane del *De copia verborum ac rerum*.

In *Powerpoint in rhetoriktheoretischer Sicht* Knappe analizza la produzione del testo per Powerpoint in base alla teoria retorica della progettazione testuale: il testo di Powerpoint viene considerato come *paratesto* oppure come testo centrale. In quanto *paratesto*, funge da compendio (Kondensat), da complemento (Korrelat) e da illustrazione (Illustrat) del testo centrale. Come testo centrale il testo di Powerpoint rappresenta anche la ricerca di una soluzione a due canali: acustico, la lettura dell'oratore, ed ottico, la lettura muta del pubblico.

Bibliografia consigliata

Knappe J., *Der humanistische Geleittext als Paratext – am Beispiel von Brants Beigaben zu Tennglers Layen Spiegel*, in Deutsch A. (Hg.), *Ulrich Tenglers Laienspiegel. Ein Rechtsbuch zwischen Humanismus und Hexenwahn*, Heidelberg, Blaukreuz-verlag, 2011, pp. 117-138

Knappe J., *Powerpoint in rhetoriktheoretischer Sicht*, in Bernt Schnetler, Knoblauch H. (Hg.), *Powerpoint-Präsentationen. Neue Formen der gesellschaftlichen Kommunikation von Wissen*, Konstanz, UVK Verlagsgesellschaft mbH, 2007

Zingarelli N., *Lo Zingarelli 1995. Vocabolario della lingua italiana di Nicola Zingarelli, Dodicesima edizione a cura di Mirco Dogliotti e Luigi Rosiello*, Bologna, Zanichelli, 1995

Sitografia

**Paratesto* (s.v.), in *Garzanti Linguistica*, <<http://www.garzantilinguistica.it/ricerca/?q=paratesto>>

**Paratesto* (s.v.), in *Treccani*, <<http://www.treccani.it/vocabolario/paratesto/>>

Cristiano Rocchio

Uno dei linguaggi di programmazione maggiormente diffusi e utilizzati per applicazioni di vario tipo in diversi campi, soprattutto in combinazione con MySQL. Il *PHP* è un linguaggio di scripting, con cui dunque è possibile realizzare sequenze di comandi per risolvere un determinato compito, da richiamare con riga di comando da terminale (vd. Script). Come molti linguaggi di scripting, *PHP* necessita di un interprete, affinché la macchina comprenda le istruzioni contenute nel codice in *PHP*. Nel momento in cui installiamo *PHP* dal sito ufficiale, configuriamo e installiamo anche l'interprete. Un server locale, come ad esempio Xampp, permette l'installazione dell'interprete, assieme ad altri strumenti utili per realizzare applicazioni.

L'uso di *PHP* quindi comprende la realizzazione di script a riga di comando, da terminale (vd. Terminale) di applicazioni lato server (vd. Lato-server), ovvero comprendenti operazioni, ad esempio, legate a gestione, trattamento e immagazzinamento di informazioni in un database; di applicazioni stand-alone, non legate quindi a risorse esterne.

PHP è un linguaggio open-source (vd. Open-source), attualmente alla versione 7.2, con un'ampia quantità di risorse reperibili in rete, come librerie e script già pronti. Il suo punto di forza è sicuramente la possibilità di integrazione con molti DBMS; dal punto di vista della sicurezza, richiede un aggiornamento costante, sebbene le versioni più recenti abbiano raggiunto un livello estremamente soddisfacente, in confronto alle precedenti.

Nelle DH, *PHP* è uno dei linguaggi più comuni per la creazione di interfacce per database di vario tipo. (vd. Database)

Esempio:

Tlion – *Tradizione della Letteratura Italiana Online*:

Il Tlion – assieme ad altri DB che ne condividono l'impianto – è sviluppato in *PHP* e MySQL su server Apache. Si tratta di una banca dati a schede, catalogate per autore e opera, con notizie relati-

ve alla tradizione dei testi della letteratura italiana e liberamente consultabile dal sito: <http://tlion.sns.it/>.

FlaviaSciolette

assegnata a Concetta Damiani

Il processo di *Self-archiving* (o autoarchiviazione o Green Road) consiste nel deposito di documenti digitali da parte di un autore in archivi ad accesso libero (Open Access o OA), per consentirne liberamente l'utilizzo e/o la consultazione. Il termine *self-archiving* è stato proposto in ambito universitario da Stevan Harnad nel 1994, in *Subversive Proposal*, pubblicato nel 1995 per massimizzare e condividere i prodotti della ricerca.

L'autoarchiviazione è eseguita in Open Archives (o E-prints Server o Data Provider), ossia archivi aperti, che, attraverso la funzione di harvesting (raccolta), possono essere interrogati da un servizio (il Service Provider) atto ad indicizzare i metadati raccolti. Suddetti archivi rientrano nell'OAI (Open Archive Initiative), un registro preposto al deposito dei documenti scientifici in forma elettronica, nato per rendere facilmente fruibili gli archivi contenenti documenti prodotti in ambito accademico e per incoraggiare la loro produzione in ambito scientifico/universitario.

Gli Open Archives si distinguono in due tipologie:

1. Open Archive istituzionali: l'archivio raccoglie tutti i lavori di un ente (università, ente di ricerca, dipartimento) o una parte dei lavori che l'ente ritiene di conservare nel deposito. In questo caso i materiali raccolti coinvolgono varie discipline.
2. Open Archive disciplinari: come ArXiv (contenente in larga parte pre-prints di più di mezzo milione di articoli di Fisici) e RePEc (Research Papers in Economics, la più grande collezione decentrata di documenti ad Accesso Libero per l'economia) o PubMed, una banca bibliografica relativa alla letteratura scientifica biomedica. L'archivio disciplinare raccoglie i lavori in una determinata disciplina. Può anche trattarsi di un server di un ente che decide di aprire più archivi per discipline differenti. Molto spesso però si tratta di più soggetti (enti o anche soggetti individuali, dipende dall'organizzazione che si vuole adottare) che interagiscono nei Reposi-

tories (depositi) di una stessa disciplina o di un argomento specifico.

L'elenco di tutti i Repository aperti OA si può consultare su:

1. OpenDOAR (*Directory Access Repositories*): un repertorio di oltre 2000 Repository ricercabile per area disciplinare, lingua, nazione, software utilizzato, etc.
2. ROAR (*Registry of Open Access Repositories*): un registro di oltre 2000 Repository ricercabili per nazione, tipo di Repository e software;
3. PLEIADI: portale per la letteratura scientifica elettronica italiana su Archivi Aperti e Depositi Istituzionali.

I documenti digitalizzati e archiviati sono gli e-prints (versione digitale di un documento di ricerca), ma anche poster scientifici, tesi, presentazioni, capitoli di libri, etc. Gli e-prints, a seconda dello stadio editoriale in cui si trovano, si distinguono in:

- Pre-print (o pre-stampa): tipologia di documento, distribuito in modo più o meno limitato, relativa ad un lavoro tecnico spesso in forma preliminare, precedente la sua pubblicazione in un periodico.
- Post-print (o post-stampa): versione modificata del pre-print, che ha passato il comitato editoriale e che è già stata sottoposta a refereeing.
- Camera ready (o definitivo): l'articolo nella sua ultima versione, come viene pubblicato dalla casa editrice.

Secondo la politica di molti editori, un ricercatore può autoarchiviare diverse versioni del proprio documento: la versione pre-print precedente alla peer review, e la versione post-print che è stata rivista e accettata per la pubblicazione.

Per essere certi che l'editore a cui l'autore sottopone l'articolo non vieti l'autoarchiviazione, è bene conoscere la politica editoriale di archiviazione e consultare la banca dati SHERPA/RoMEO *Publisher Copyright Policies and Self-Archiving*, che, a seconda che l'editore consenta o meno l'autorarchiviazione, assegna un colore a seconda dei diritti concessi agli autori:

- bianco: nessun diritto all'autoarchiviazione;
- giallo: diritto di archiviare pre-print;
- blu: diritto di archiviare post-print o una versione PDF dell'editore;

- verde: diritto di archiviare pre-print e post-print e, a volte, la versione elettronica della pubblicazione finale, prodotta dalla casa editrice.

Il colore assegnato da SHERPA costituisce solo una catalogazione preliminare, potendo le case editrici imporre particolari condizioni a loro discrezione.

I ricercatori usano diversi Academic social networks, come ResearchGate (<https://www.researchgate.net/about>) e Academia.edu (<https://www.academia.edu/about>), principalmente utilizzati per condividere documenti, monitorare il loro impatto e seguire i progressi della ricerca, selezionando un'area di interesse ed esplorando i profili di utenti con interessi affini.

Dalla vendita degli articoli pubblicati in riviste scientifiche, i ricercatori non traggono «guadagni di pubblicazione» ma ottengono «guadagni di impatto», se gli articoli sono diffusi in modo adeguato. L'obiettivo, infatti, è di mostrare i loro lavori al maggior numero di persone, indipendentemente dal guadagno economico personale, e la pubblicazione online dei loro contributi consente tale visibilità.

Come dice Stevan Harnad in *Subversive Proposal*: «i costi elevati nell'era cartacea di Gutenberg, dispendiosa e inefficace, erano inevitabili; ma oggi, nell'era post-Gutenberg on line, il funzionamento alla vecchia maniera, con i suoi costi elevati deve essere mantenuto come opzione complementare invece che come strumento indispensabile».

Bibliografia consigliata

Delle Donne R., *Studi e ricerche di scienze umane e sociali*, Napoli, fedOA Press, 2014

De Robbio A., *Archivi aperti e comunicazione scientifica*, Napoli, Clio-Press, 2007

Sitografia

Antelman K., *Self-archiving practice and the influence of publisher policies in the social sciences*, <<https://repository.lib.ncsu.edu/bitstream/handle/1840.2/83/antelman-archiving.pdf?sequence=1&isAllowed=y>>

**Autoarchiviazione* (s.v.), in *Wikipedia* <<https://it.wikipedia.org/wiki/Autoarchiviazione>>

Brody T. *et al.*, *The effect of Open Access on citation impact*, <<http://opcit.eprints.org/feb19oa/brodyimpact.pdf>>

De Robbio A., *Auto-archiviazione per la ricerca: problemi aperti e sviluppi futuri*, <file:///C:/Users/Utente/Downloads/Auto-archiviazione_per_la_ricerca_problemi_aperti_.pdf>

Eberechukwu Eze M. - Chukwuma Okeji C. - Ejiobi Bosah G., *Self-Archiving options on social networks: a review of options*, <file:///C:/Users/Utente/Downloads/Self-archivingonsocialmedia.pdf>

Gadd E. A. – Troll Covey D., *What does «green» open access mean? Tracking twelve years of changes to journal publisher self-archiving policies*, <<https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/21555/5/Article%20-%20What%20does%20green%20mean%20v6%20Submitted%20%20reformatted%20for%2>>

Harnad S., *Self-Archive Unto Others as Ye Would Have them Self-Archive Unto You*, <<https://jcom.sissa.it/sites/default/files/documents/jcom0203%282003%29F03.pdf>>

Jenkins C. *et al.*, *RoMEO Studies 8: Self-archiving – when Yellow and Blue make Green: the logic behind the colour-coding in the Copyright Knowledge Bank*, <https://www.researchgate.net/publication/28693015_RoMEO_Studies_8_Self-archiving_The_logic_behind_the_colour-coding_used_in_the_Copyright_Knowledge_Bank>

Mallikarjun D. - Bulu M., *Driving on the Green Road: Self– Archiving Research for Open Access in India*, <<https://www.questia.com/library/journal/1G1-331807690/driving-on-the-green-road-self-archiving-research>>

**Self-Archiving* (s.v.), in *Wikipedia* <<https://en.wikipedia.org/wiki/Self-archiving>>

Swan A. - Brown S., *Open access self-archiving: an author study*, <<https://eprints.soton.ac.uk/260999/1/jisc2.pdf>>

Alessandra Di Meglio

(ingl. Acronimo per *Standard Generalized Markup Language*)

È il progenitore dei «markup languages» utilizzati attualmente. Charles Goldfarb inventò GML (*Generalized Markup Language*) nel 1969 durante il suo lavoro per la IBM sotto la supervisione di Steve Furth, e ne definì i presupposti teorici: «This analysis of the markup process suggest that it should be possible to design a generalized markup language so that markup would be useful for more than one application or computer system.

...

This could be done, for example, with mnemonic «tags». The designation of a component as being of a particular type would mean only that it will be processed identically to other components of that type».

Goldfarb afferma e struttura, quindi, quelli che saranno le peculiarità di tutti i linguaggi di codifica basati su GML: i markup devono essere, prima di tutto, applicabili a differenti sistemi e piattaforme informatiche e dovranno poter essere inserite all'interno del testo, mentre tutte le istruzioni procedurali verranno inserite in un documento a sé stante. Proprio per queste peculiarità GML diventò standard nel 1986, anche se per l'eccessiva complessità della codifica non viene più utilizzata, in quanto si preferiscono codifiche più snelle come HTML e XML.

SGML venne utilizzato in molti progetti, il primo fu quello sviluppato dall'*American Association of Publishers* (AAP): selezionarono tre tipi di testo dal vasto campo dell'editoria, un libro, una serie di pubblicazioni e un articolo, e ne definirono le rispettive *Document Type Description* (DTD).

All'interno del documento troviamo tre elementi fondamentali:

1. La dichiarazione *SGML*, che va sempre posta come apertura di un nuovo documento, poiché ne definisce letteralmente il

tipo. Dichiarare, senza troppi giri di parole, che quel documento sarà codificato secondo le norme del *SGML*.

2. La DTD, necessaria in ogni documento. Essa è un insieme di stringhe di codice che esplicitano elementi e attributi e sistemanano le relazioni tra di loro: è una sorta di vocabolario per la codifica del documento.
3. L'istanza del documento, nella quale viene marcata ogni parte del documento utilizzando i marcatori contenuti nella DTD.

Con l'utilizzo di questo tipo di linguaggio entriamo nell'ordine di idee di un sistema strutturato secondo una precisa gerarchia degli oggetti testuali; sotto questo punto di vista il testo è visto come «una gerarchia ordinata di oggetti di contenuto» (De Rose). Ciò vuol dire che ogni oggetto testuale ha una relazione gerarchica con gli altri e che essi altro non sono che le strutture editoriali: a partire dal macro contenitore del libro, scendendo via via a livelli più bassi con capitoli, paragrafi, citazioni, note, apparati, ecc.

Questa teoria gerarchica ha i suoi limiti dati principalmente dai diversi livelli di interpretazione che possiamo affidare al testo. Non è possibile, infatti, creare un sistema gerarchico che sia in grado di evidenziare, allo stesso tempo, ogni caratteristica testuale. Si è cercato di dare una nuova interpretazione «pluralista» alla teoria gerarchica, ma quello che qui ci preme sottolineare è l'importanza di questa caratteristica dell'*SGML* che venne come uno dei principi fondamentali del linguaggio in *Gentle Introduction of SGML* di C. M. Sperberg-McQueen e L. Burnard. Oltre ad asserire il principio gerarchico, gli autori parlano del suo essere «descriptive markup», quindi un sistema di codifica che descrive la struttura logica del documento, e della sua «data independence» che poggia, a sua volta, su due caratteristiche: la prima è che si tratta di un metalinguaggio, la seconda è l'uso del concetto di entità con il quale è possibile codificare dei caratteri alfabetici diversi dallo standard inglese.

Bibliografia consigliata

DeRose S.J. - Durand D. - Milonas E. - Renear A.H., *What is Text, Really?*, in *Journal of computing in Higher Education*, 1, 2, 1990, pp. 3-26

Sitografia

Goldfarb C.F., *Design Considerations For Integrated Text Processing System*, in *IBM Cambridge*

Scientific Center – Technical Report No. 320-2094, 1973, <<http://www.sgmlsource.com/history/2094/G320-2094.html>>

SGML Web Page, <<http://www.sil.org/sgml/sgml.html>>

AlessiaMarini

L'esigenza di creare un sistema di classificazione che miri all'immediata e puntuale reperibilità delle informazioni contenute in una pagina web, ha dato vita alla *tassonomia* informatica o informatizzata, che si occupa della classificazione degli argomenti informatici secondo un ordine logico.

L'etimo greco del termine – da *τάξις* («ordine») e *νόμος* («regola») – designa un'idea di ordinamento che, in un sito web, segue sostanzialmente due criteri:

- di vicinanza, qualora si uniscano due o più categorie di contenuto simile;
- di specificità, che corrisponde all'espansione di una categoria i cui contenuti sono circoscritti a un solo argomento.

A questi criteri tassonomici si affiancano due tipologie di tassonomie generali:

- *tassonomia orizzontale*, che si configura come semplice elenco di categorie di primo livello – in cui gli elementi elencati hanno pari importanza –, o come raggruppamenti tematici ottenuti mediante tag. Il tag coincide con una key-words orizzontale e trasversale a un *corpus* di contenuti che unisce diversi documenti, anche non pertinenti alla specifica ricerca dell'utente, ma affini tematicamente;
- *tassonomia verticale*, in cui i contenuti sono pertinenti a un macro-argomento di riferimento e le categorie – che contengono a loro volta sottocategorie – sono disposte gerarchicamente all'interno del sito in ordine di importanza.

Hanno un impianto tassonomico i vocabolari che organizzano in maniera strutturata categorie concettuali a seconda degli ambiti di analisi, o, in termini meno specifici, qualsiasi pagina o sito web organizzati ordinatamente.

I vantaggi ottenuti dall'introduzione della *tassonomia* consistono nella funzionalità e nell'operabilità di un sito atte a migliorare l'esperienza di navigazione dell'utente (User Experience o UX): la *tassonomia*, infatti, mette gli utenti nella condizione di reperire comodamente un contenuto, consentendo al motore di ricerca di associare il sito web a un particolare topic (Search Topic).

Data l'importanza di questa disciplina, certa è anche la sua validità nell'ambito delle Digital Humanities: essa infatti struttura e utilizza le informazioni importanti per le discipline umanistiche rendendole facilmente individuabili, nonché consultabili, e dispone di un enorme potenziale da esplicarsi soprattutto nella raccolta e nella classificazione delle informazioni, degli strumenti, dei metodi e dei progetti di DH necessari a facilitare la ricerca.

Bibliografia consigliata

Isoni L., *Strategie SEO per l'E-COMMERCE*, Milano, Hoepli, 2017, <<https://books.google.it/books?id=XjskDwAAQBAJ&pg=PT99&lpg=PT99&dq=tassonomia+orizzontale+e+verticale&source=bl&ots=9GuQqPRwYg&sig=ACfU3U0S6WNix0VGaq70l1yl21x0jk0RVQ&hl=it&sa=X&ved=2ahUKEwjm3v24r5LgAhWGM-wKHUtdDjkQ6AEwCHoECAQQAQ#v=onepage&q=tassonomia%20orizzontale%20e%20verticale&f=false>>

Sitografia

AaVv, *Dall'informatica umanistica alle culture digitali. In memoria di Giuseppe Gigliozi*, <file:///C:/Users/Utente/Downloads/DallInformatica_umanistica_alle_culture.pdf>

Minini A., *Le tassonomie verticali e orizzontali*, <<http://www.andreaminini.com/seo/la-tassonomia-del-sito-web-nella-seo>>

Ruecker S., *A brief taxonomy of prototypes for the Digital Humanities*, <<file:///C:/Users/Utente/Downloads/222-Article%20Text-2065-2-10-20151015.pdf>>

**Tassonomia* (s.v.), in *Treccani Enciclopedia*, <<http://www.treccani.it/enciclopedia/tassonomia/>>

**Tassonomia* (s.v.), in *Wikipedia*, <<https://it.wikipedia.org/wiki/Tassonomia>>

Wahl Z. - Busch J., *Il valore della tassonomia nella ricerca delle informazioni*, <<http://www.datamanager.it/rivista/searching/il-valore-della-tassonomia-nella-ricerca-delle-informazioni/999/>>

AlessandraDiMeglio

(ingl. Acronimo per *Text Encoding Initiative*)

È un progetto sviluppato a partire dal 1987 che si propone di conservare le linee generali per la codifica di testi letterari e linguistici: i suoi scopi sono stati espressi alla fine di una conferenza tenutasi al Vassar College, N.Y., nel novembre dello stesso anno. Si tratta di una marcatura basata sul linguaggio XML e «il suo uso è ambizioso nella sua complessità e generalità, ma fondamentalmente non c'è differenza tra questo e gli altri schemi di markup XML» (Burnardn, Sperberg-McQueen). Le *TEI Guidelines*, le linee guida su cui si basa il vocabolario della codifica e che gestisce anche le regole gerarchiche, sono state rese pubbliche nel maggio del 1994, dopo sei anni di sviluppo gestito da centinaia di studiosi, provenienti da differenti discipline, sparsi per il mondo. Da questa conferenza nacque anche quella semplificazione del linguaggio, denominata *TEI Lite*, che gli ideatori usarono per fornire una dimostrazione di come lo schema di codifica *TEI* poteva essere adottato per servire le necessità dei testi e per rendere la codifica più comprensibile ed usabile da tutti gli utenti, senza il bisogno di una conoscenza approfondita della *TEI DTD*.

Dal punto di vista pratico, la codifica si divide in due parti:

1. un *<tei header>* in cui sono contenute tutte le informazioni relative alle notazioni bibliografiche e di catalogazione del testo e le specifiche della codifica;
2. la trascrizione del testo vero e proprio contenuto nel marcatore *<text>*.

All'interno del *<text>* che possiamo identificare come il macro contenitore in cui sono contenuti frontespizio, testo vero e proprio, apparati paratestuali, sono presenti altri marcatori studiati appositamente per riferirsi a tutte le componenti presenti in un'edizione, studiati sempre per garantire organizzazione e gerarchia del documento di codifica. Troviamo, quindi, in successione:

1. *<front>* che contiene tutto il materiale paratestuale, come intestazioni, frontespizio, prefazioni, dediche, ecc., che si trova prima del testo vero e proprio;
2. *<group>* mette insieme tutti i testi appartenenti ad un determinato gruppo. È utile, ad esempio, quando ci si trova a dover codificare un'opera omnia di un autore che può, perciò, contenere testi in prosa, in versi o di saggistica;
3. *<body>* è riferito ad un testo unitario ad esclusione di tutte le aggiunte paratestuali;
4. *<back>* raggruppa invece tutti quei materiali che sono successivi al testo. Parliamo, in questo caso, di note, appendici, bibliografie, ecc.

Tra questi sotto contenitori del *<text>* ciò che ci interessa maggiormente è quello del *<body>* nel quale si concentra il grosso degli sforzi del codificatore, poiché il suo obbiettivo primario è quello di rendere fruibile il testo per l'utente.

L'utilizzo delle *<div>* per organizzare le porzioni di testo è molto comune in tutti i progetti che utilizzano la codifica *TEI*, in quanto è uno degli elementi che, a loro volta, possono essere divisi ed identificati da attributi globali inseriti nella DTD:

- *type*: sta ad indicare letteralmente il tipo di categoria a cui appartiene la *<div>*, che sia riferita ad un capitolo, ad un paragrafo, ad un sonetto, e via scorrendo;
- *id*: specifica un unico identificatore per la divisione, che potrebbe essere usato per inserire del «cross references» o altri link;
- *n*: specifica il numero della divisione, che può essere usato per identificare una preferenza gerarchica o la successione delle parti.

Essendo *TEI* basato su XML è molto importante ricordarsi di chiudere ogni marcatore: la mancata chiusura determinerebbe, infatti, un errore nel documento e, conseguentemente, l'impossibilità di una corretta visualizzazione del testo.

Oltre a tutti i marcatori specifici di ogni testo, gli ideatori della *TEI Lite* si sono anche dedicati alla creazione di elementi per mettere in evidenza porzioni di testo:

- *<emph>* usato per espressioni che vengono messe in qualche modo in risalto.

- *<foreign>* atto all'identificazione di parole o frasi in una lingua diversa da quella dell'intero testo.
- *<mentioned>* indica una menzione o una citazione
- *<title>* la quale contiene il titolo di un'opera, sia essa articolo, libro, giornale, o collana, compreso ogni titolo alternativo o sottotitolo. Viene spesso completato da degli attributi come «level» che indica se il titolo è di un articolo, libro, giornale, collana o materiale inedito, o il già citato type.

Il *TEIheader*, invece, è una sezione iniziale, posizionata esattamente tra la dichiarazione XML e il tag *<text>* e contiene informazioni di diverso genere:

- *<fileDesc>* contiene una descrizione bibliografica dell'opera (titolo, autore, data e luogo di stampa, numero di pagine, ecc.). Al suo interno troviamo altri marcatori più specifici quali:

<titleStmt> in cui troviamo i riferimenti a titolo, autore e responsabile della codifica.

<editionStmt> è il gruppo di informazioni riguardanti l'edizione a testo, nelle quali troviamo anche la descrizione delle particolarità dell'edizione stessa.

<extent> descrive quanto pesa il file in *bytes*.

<publicationStmt> è una sezione obbligatoria, in essa sono contenute informazioni circa l'organizzazione responsabile della pubblicazione, l'agenzia responsabile della distribuzione e il responsabile della messa online del documento codificato. Contiene, inoltre, una serie di informazioni relative alla pubblicazione stessa (anno, luogo, identificazione ISBN) che, però, sono completamente facoltativi.

<seriesStmt> è un elemento non obbligatorio circa l'inclusione o meno del volume in una serie.

<noteStmt> contengono vari elementi *<note>* in cui vengono specificate le varie annotazioni al testo.

<sourceDesc> è un elemento obbligatorio che registra tutte le fonti da cui il file digitale è derivato.

- *<encodingDesc>* documenta la relazione tra il testo elettronico e la fonte o le fonti da cui è derivato.

<*projectDesc*> descrive nei dettagli gli scopi e i propositi che si propone il file codificato, insieme ad altre informazioni sul processo stesso.

<*samplingDecl*> contiene una spiegazione su metodi di codifica.

<*editorialDecl*> descrive i principi editoriali e le pratiche utilizzate durante la codifica.

<*tagsDecl*> contiene informazioni circa la marcatura applicata ad un documento SGML.

<*refsDecl*> riassume i riferimenti usati durante la codifica.

<*classDecl*> contiene una o più tassonomie che definiscono ciascun codice di classificazione usato nel testo.

- <*profileDesc*> sottolinea gli aspetti non bibliografici, facendo quindi riferimento alla lingua e alle sottolingue utilizzate, a come è stato prodotto, ai partecipanti alla codifica e alle loro impostazioni.

<*creation*> informa su come è stato creato il testo.

<*langUsage*> registra le lingue, le sottolingue, i dialetti, ecc presentati all'interno del testo.

<*textClass*> è una serie di informazioni che descrivono la natura o l'argomento del testo.

Sitografia

Burnard L. - Sperberg-McQueen C. M., *TEI Lite: an introduction to Text Encoding for Interchange*, 1995 (revisited May 2002), <https://www.tei-c.org/Vault/P4/Lite/teiu5_en.pdf>

TEI: Guidelines, in *Text Encoding Initiative*, <<http://www.tei-c.org/Guidelines/>>

AlessiaMarini

Il termine *thesaurus* è – per così dire – semanticamente figlio dei tempi nei quali è stato utilizzato. Nel periodo medievale, ad esempio, il termine è utilizzato per indicare repertori scientifici, opere di carattere enciclopedico e/o divulgativo (e. g. *Thesaurus pauperum*, ricettario medico *a capite ad pedes* del XIII secolo; il *Trésor* di Brunetto Latini *et alia*). Successivamente, il termine è utilizzato in riferimento a vocabolarî, soprattutto delle lingue classiche: *Thesaurus Linguae Graecae* di Henri Estienne (1572), *Thesaurus Linguae Latinae* di Robert Estienne (1532), *Thesaurus Linguae Latinae Epigraphicae* di G. N. Olcott (1904).

Oggi il termine ha un campo semantico più ampio e più generico; l'antico concetto di tesauroizzazione lessicale, semantica, concettuale ed enciclopedica s'intreccia con la dimensione del Web, sicché il termine *thesaurus* oggi viene utilizzato per indicare un database terminologico relativo ad un determinato campo del sapere. Nella definizione moderna assume particolare rilevanza anche l'utente e la sua necessità di reperire informazioni «strutturate». Pertanto, più precisamente, si potrebbe affermare che il *thesaurus*, nella sua accezione moderna, sia sostanzialmente concepito come uno strumento gnoseologico tassonomico, funzionale alla indicizzazione di documenti (afferenti ad uno specifico settore o campo semantico) ed atto a semplificare e rendere più efficaci le ricerche condotte dagli utenti. Infatti, la Broughton (2006a, 4) definisce il *thesaurus* come «a tool used for the subject indexing of documents. It consists of terms (usually in one particular subject field) than an indexer or records manager may use to describe documents so that end-users can retrieve relevant items when searching for material about a particular subject». A tale definizione è possibile accostare quella elaborata più recentemente dalla *International Organization for Standardization*, che descrive il *thesaurus* come un «controlled and structured vocabulary in which concepts are represented by terms, organized so that relationship between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms» (ISO 2011; cfr. ISO 2013).

L'uso di *thesauri online* va sempre più accrescendosi e ciò è legato ad una serie di concomitanti motivazioni, quali l'ingente mole di informazioni disponibili online, la «migrazione» sul Web di molte informazioni in precedenza fruibili solo su supporto cartaceo, la necessità di garantire un certo livello qualitativo delle informazioni presenti sul Web e la necessità di offrire agli utenti un sapere «strutturato» (cfr. Shiri-Revie 2000, 273-274). In sintesi, i *thesauri* si configurano come un «bi-functional tool» e rivelano la propria utilità sia nell'organizzazione che nel recupero delle informazioni (cfr. Feldvari 2009). I *thesauri* hanno, inoltre, importanza trasversale, giacché potenzialmente l'uso di un *thesaurus* può rivelarsi produttivo in ogni campo del sapere, come è evidente dalla recente definizione proposta da Ryan (2014, 6): «*thesauri* are vital and valuable tools in content discovery, and in information organization and retrieval, activities common to all fields, including cultural heritage and higher education as well as business and enterprise. *Thesauri* allow information professionals to represent content in a consistent manner and enable researchers, employees and the public to find this content easily and quickly».

Davies (1996, 38-39) ha proposto una prima classificazione dei *thesauri online*, suddividendoli – per quanto attiene al formato e al processo di pubblicazione – in «statici» e «dinamici». Shiri-Revie (2000, 274) hanno proposto una classificazione più articolata: «1. thesauri in simple static formats (ASFA Thesaurus); 2. thesauri in HTML format but still static, without effective use of hyperlinks (Infotermis); 3. thesauri in dynamic HTML format with fully navigable hyperlinks MeSH); 4. thesauri with advanced visual and graphical interfaces (Plumb Design Visual Thesaurus); 5. thesauri in XML format (Virtual HyperGlossary)». Per una accurata disamina quantitativa dei principali *thesauri* liberamente interrogabili in varie lingue europee, cfr. Mochó Bezares-Sorli Rojo (2010, 643-663).

Bibliografia consigliata

Aitchinson J. - Gilchrist A. - Bawden D., *Thesaurus construction and use: a practical manual*, London, Fitzroy Dearden, 2005

Broughton V., *Essential Thesaurus Construction*, London, Facet Publishing, 2006

- Broughton, V., *The need for a faceted classification as the basis of all methods of information retrieval*, in *Aslib Proceedings: New Information Perspectives*, 58, 1/2, 2006, pp. 49-72, <<https://www.researchgate.net/publication/328952>>
- Davies R., *Publishing thesauri on the World Wide Web*, in *Proceedings of the 7th ASIS SIG/CR Classification Research Workshop* (1996), pp. 37-48, <<http://journals.lib.washington.edu/index.php/acro/article/view/12688/11192>>
- Feldvari K., *Thesauri usage in Information Retrieval Systems: Example of LISTA and ERIC Database Thesaurus*, in *INFuture 2009-Digital Resources and Knowledge Sharing*, 2009, pp. 279-288 <[https://infoz.ffzg.hr/infuture/2009/papers/4-09%20Feldvari,%20Thesauri%20usage%](https://infoz.ffzg.hr/infuture/2009/papers/4-09%20Feldvari,%20Thesauri%20usage%20)>
- International Organization for Standardization, information and documentation, *Thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval*, Geneva, ISO, 2011
- International Organization for Standardization, information and documentation, *Thesauri and interoperability with other vocabularies. Part 2: interoperability with other vocabularies*, Geneva, ISO, 2013
- Mochón Bezares G. - Sorli Rojo A., *Thesauri en acceso abierto en Internet. Un análisis cuantitativo*, in *Revista Española de Documentación Científica*, 33, 4, 2010, pp. 643-663, <<http://redc.revistas.csic.es/index.php/redc/article/view/>>
- Ryan C., *Thesauri Construction Guidelines: An Introduction to thesauri and guidelines on their construction*, Dublin 2014, <<http://apps.dri.ie/motif/docs/guidelines>>
- Shiri A. A. - Revie C., *Thesauri on the web: current developments and trends*, in *Online Information Review* 24, 4, 2000, pp. 273-279, <<https://strathprints.strath.ac.uk/1896/1/strathprints001896.pdf>>

Sitografia

- **Quick Guide to publishing a Thesaurus on the Semantic Web*, <<https://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050517/>>
- * *Thesaurus* (s. v.), in *Treccani*, <<http://www.treccani.it/enciclopedia/thesaurus/>>
- * *Thesaurus interrogabili* (s. v.), in *Treccani*, <<http://www.treccani.it/vocabolario/thesaurus/>>
- **Thesaurus Linguae Graecae* (online), <<http://stephanus.tlg.uci.edu/>>
- **Thesaurus Linguae Latinae* (online), <<https://www.degruyter.com/view/db/tll>>
- **Thesaurus Linguae Latinae Epigraphicae* (online, scaricabile), I, <<https://archive.org/details/ThesaurusLinguaeLatinaeEpigraphicae>>
- **Thesaurus Musicarum Latinarum* (online), <<http://boethius.music.indiana.edu/tml/>>

Giovanna Battaglino

Il *Web Semantico*, detto anche «Web dei dati» è l'insieme dei servizi e delle strutture capaci di interpretare il significato dei contenuti del Web. In buona sostanza si tratta di un'estensione del Web che implica un nuovo modo di concepirne i documenti, in cui le informazioni assumono un ruolo ben preciso e in cui computer e utenti lavorano in cooperazione, secondo le intenzioni di Tim Berners-Lee che l'ha ipotizzata nel 2001.

Il Web attuale è un insieme di testi collegati tra loro da link, in cui i soli utenti umani sono in grado di leggere e comprendere i contenuti delle pagine che stanno visitando, grazie alla loro esperienza di navigazione e alla capacità di interpretazione.

Da un punto di vista strutturale il *Web Semantico* è stato rappresentato da Berners-Lee come una piramide di sette strati, composta da nove elementi. Tale modello architettonico sorregge tre tipologie di informazioni: documenti autodescrittivi, dati e regole. Il significato della nuova architettura va seguito dal basso verso l'alto: il Web può essere concepito come un insieme di strati con standard, linguaggi o protocolli che agiscono come piattaforme sulle quali possano poggiarsi formalismi nuovi, più ricchi e più espressivi.

Aspetti fondamentali delle innovazioni tecnologiche che sottendono all'architettura del *Web Semantico* sono rappresentati da metadati; identificatori di dati (URI); ontologie, reti di fiducia.

La presenza di sistemi rigorosi e standardizzati di metadati nei contenuti delle pagine e di correlazioni tra essi, invece, consentirebbe l'uso di automi in grado di comprendere il significato dei testi presenti sulla rete e di guidare l'utente direttamente verso l'informazione cercata, oppure di svolgere compiti per suo conto, rispondendo così in maniera efficace anche ai problemi posti dalla crescita esponenziale della quantità di informazione disponibile in rete e dal moltiplicarsi delle sue tipologie. La semantica dei dati consiste nel rappresentare il modello di uno specifico dominio di conoscenza codificando le informazioni mediante ontologie, con

la descrizione formale dei concetti articolata per classi, relazioni e regole, in modo che la macchina sia in grado d'interpretare le informazioni e di utilizzarle correttamente.

Per definire una fondamentale caratteristica del *Web Semantico* si fa riferimento agli open data e in particolare alla disponibilità dei dati e alla conseguente possibilità di identificarli e citarli. Il *Web Semantico* è quindi come già indicato un'estensione del Web tradizionale, nel senso che rappresenta il successivo passaggio del linking ed è pensato per funzionare nel contesto di un modello relazionale di dati, in cui il link - da collegamento generico e cieco tra due documenti - diviene capace di esprimere relazioni concettuali che convogliano significati.

La prospettiva di un *Web Semantico* realmente sviluppato non è ancora matura, ma non vi è dubbio che la tendenza alla standardizzazione e all'interoperabilità degli strumenti e dei parametri descrittivi abbia ricevuto, con la crescita di Internet, un impulso notevole. I primi passi in questa direzione sono stati compiuti garantendo l'interoperabilità dei cataloghi ad accesso pubblico dei sistemi bibliotecari (OPAC, *On-line public access catalogue*).

Chiaramente l'aspirazione della rete semantica non è quella di rappresentare tutti i dati o il sapere in qualche ristretto insieme di formalismi, ma ottenere che la possibilità di linkare i dati a nuovi dati permetta di usarli in maniera sempre più ampia, confidando nell'intelligenza distribuita e interconnessa dei navigatori.

Bibliografia consigliata

Berners-Lee T., *L'architettura del nuovo Web. Dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, Milano, Feltrinelli, 2001

Di Donato F., *La scienza e la rete. L'uso pubblico della ragione nell'età del web*, Firenze, Firenze University Press, 2009

Sitografia

**Web semantico* (s.v.), in *Treccani - Lessico del XXI secolo*, <http://www.treccani.it/enciclopedia/web-semantico_%28Lessico-del-XXI-Secolo%29/>

Concetta Damiani

(ingl. Acronimo per *Exstensible Stylesheet Language*)

Si occupa di definire, all'interno di un document XML, quale sarà la sua visualizzazione sul browser, in quanto è in grado di contenere le informazioni su formattazione e visualizzazione del documento, filtrare i dati, riorganizzarli nelle gerarchie ed eseguire calcoli. Volgarmente definito come foglio di stile, l'*XSL* è più che altro un insieme di linguaggi atti a rendere possibile la trasformazione del documento, in modo da renderlo leggibile e visualizzabile sui browser. Tra tutti i linguaggi, quello maggiormente interessante è l'*XSLT* (*Extensible Stylesheet Languages for Transfomation*), raccomandato ufficialmente dal W3C nel novembre del 1999.

La sua funzione primaria è quella di trasformare letteralmente il documento XML in HTML e, per farlo, utilizza un linguaggio basato su regole di «pattern matching», con le quali si danno delle specifiche di visualizzazione a delle porzioni di testo, in modo tale che ogni volta che quella determinata parte, o una simile, viene rinvenuta nel documento, verrà immediatamente sostituita con l'impostazione data dalle regole di visualizzazione.

Un documento *XSLT* si compone di tre fasi fondamentali: nella prima *<xsl: output>* è il marcatore che specifica il tipo di formato in cui verrà visualizzato il documento finale, ad es. HTML, mentre *<xsl:template>* specifica quali parti del documento XML dovranno essere trasformate. Nella seconda fase si ha una trasformazione strutturale dei dati: si passa quindi dalla struttura dell'input a quella dell'output desiderato. Nella terza ed ultima fase avviene la formattazione dei dati secondo le specifiche dell'output selezionato. In parole più semplici il compito dell'*XSLT* è quello di prendere le informazioni da un documento di partenza, trasformarle secondo il linguaggio del documento finale e poi formattare questa nuova gerarchia di informazioni secondo le specifiche del foglio di stile dell'output (si passerà, quindi, da un file.xml ad un file.html che verrà di visualizzato ed organizzato con CSS).

Bibliografia consigliata

Ausiello G. *et al.*, *Modelli e linguaggi dell'informatica*, Milano, McGraw-Hill Education, 1991

Ciotti F., *Il testo e l'automa. Saggi di teoria e critica computazionale dei testi letterari*, Roma, ARACNE editrice s.r.l., 2007

Moller A. - Schwartzbach M.I. - Gaburri S. (a cura di), *Introduzione a XML*, Milano, Pearson, 2007

Tissoni F., *Lineamenti di editoria multimediale*, Milano, Edizioni Unicopli, 2009

Sitografia

**Learn W3C* (s.v.), in *w3schools.com*, <<http://www.w3schools.com/>>

AlessiaMarini

(ingl. Acronimo per *eXtensible Markup Language*)

Assieme ad HTML (*HiperText Markup Language*), è un linguaggio di marcatura derivato da SGML (*Standard Generalized Markup Language*). Esattamente come il suo genitore è di tipo descrittivo, ciò vuol dire che tutti i marcatori sono creati e pensati per descrivere la funzione delle porzioni di testo che racchiudono.

Il progetto per lo sviluppo del XML ha inizio nel 1996 con il fine di ampliare e potenziare la capacità di gestione ed elaborazione dei documenti sul Web; per questo venne creato un gruppo di lavoro *XML Working Group*, composto dai maggiori esperti mondiali di SGML, che nel 1998 completarono lo sviluppo del linguaggio di marcatura. Con la successiva sottoscrizione di Michael Sunshine, allora presidente del W3C (*Web Consortium*), il linguaggio divenne uno standard internazionale.

XML può essere pensato come un sottoinsieme di SGML da cui eredita sia la sintassi sia la logica di funzionamento. Al contrario dell'HTML, XML non ha dei marcatori fissi e determinati, ma, come ci suggerisce il nome, è estensibile e modificabile. Tutto ciò è possibile grazie all'uso di una DTD (*Document Type Description*): è importante avere ben chiaro, prima di iniziare ad avventurarsi nella pratica di questo linguaggio, che non può esistere un file XML privo di DTD in quanto essa ne rappresenta implicitamente le fondamenta.

Un'altra differenza fondamentale con l'HTML è che esso si preoccupa più di descrivere i particolari fisici del documento (utilizzando `<i>` `</i>` per segnalare un titolo in corsivo), l'XML invece preferisce un tipo di descrizione degli aspetti logici, così un titolo in corsivo viene marcato come `<title>` per differenziarlo logicamente, e non graficamente, dagli altri oggetti testuali. Il linguaggio di marcatura, quindi, non si interessa tanto di espletare il significato del singolo elemento, quanto di definire quale sia la sua relazione con gli altri.

Altra peculiarità dell'*XML* è il foglio di stile necessario per la visualizzazione sui browser: chiamato *XSL (Exstensible Stylesheet Language)*, esso permette non solo l'inserimento di specifiche di visualizzazione come i contenuti grafici ed il posizionamento dei contenuti, ma principalmente di preoccupa rendere leggibile il file *XML* per il browser.

Indispensabile in ogni documento *XML* è il prologo nel quale è contenuta la dichiarazione della versione del linguaggio utilizzata e specifica il set di caratteri usati dalla codifica caratteri.

All'interno della codifica è anche possibile utilizzare dei commenti utili nel caso in cui un certo documento subisca più lavorazioni da parte di mani diverse: l'uso dei commenti serve al codificatore per segnalare delle particolarità del documento o per suggerire metodi di lavorazione.

In ultimo va citato l'uso delle entità che aiuta a codificare tutte quelle parti, non necessariamente testuali, che non possono essere piegate alla logica gerarchica del documento. Possono essere, quindi, di due tipi, quelle generali, già normalmente inserite nella DTD (delimitate da & e ;), e quelle parametriche, più specifiche, che vengono usate solo nelle dichiarazioni di markup (delimitate da % e ;).

Bibliografia consigliata

Ausiello G. *et al.*, *Modelli e linguaggi dell'informatica*, Milano, McGraw-Hill Education, 1991

Ciotti F., *Il testo e l'automa. Saggi di teoria e critica computazionale dei testi letterari*, Roma, ARACNE editrice s.r.l., 2007

Moller A. - Schwartzbach M.I. - Gaburri S. (a cura di), *Introduzione a XML*, Milano, Pearson, 2007

Tissoni F., *Lineamenti di editoria multimediale*, Milano, Edizioni Unicopli, 2009

Sitografia

Sorato A., *Linguaggi per la rete: XML*, <<http://www.dsi.unive.it/~asorato/SlideXML/DTD.pdf>>

**Learn W3C* (s.v.), in *w3schools.com*, <<http://www.w3schools.com/>>

AlessiaMarini

