

Granularità: un percorso di analisi

Maurizio Zani

Biblioteca Centrale della Facoltà di Ingegneria Alma Mater Studiorum dell'Università di Bologna

Negli ultimi anni il termine granularità si è diffuso in maniera crescente nell'ambito delle discipline del libro e del documento, con usi spesso difformi. L'intento principale della ricerca è stato quello di "mappare" i significati assunti dal vocabolo. Granulare può essere una risorsa costituita da blocchi di testo (lessie) uniti da link ipertestuali; granulare può essere un metadato relativo a una risorsa; granulare è intesa la descrizione di o l'accesso a un articolo di una rivista piuttosto che al fascicolo al cui interno è stampato. Esigenze commerciali e sviluppo delle tecnologie si contendono anche queste definizioni. Il censimento effettuato evidenzia la diffusione del termine ma anche la profondità storica dei concetti e dei significati associati. La ricerca dell'accesso alle entità più minute di una risorsa (ricerca della parola, della frase, della citazione, dell'articolo, ecc.), favorito dallo sviluppo delle tecnologie digitali, non può d'altra parte ignorare la necessità che all'utente finale siano comunque sempre fornite anche le necessarie relazioni, in particolare quelle gerarchiche, che esistono tra le diverse parti componenti delle risorse e che possono consentirgli una più agevole navigazione.

Negli ultimi anni il termine *granularità* si è diffuso in maniera crescente nell'ambito delle discipline del libro e del documento. I significati attribuitigli sono stati spesso difformi. Granulare può essere una risorsa costituita da blocchi di testo (*lessie*) uniti da link ipertestuali; granulare può essere un metadato relativo a una risorsa; granulare può essere la descrizione di un articolo di una rivista piuttosto che quella del fascicolo al cui interno è stampato. L'intento principale della ricerca è quello di "mappare" i significati assunti dal termine, per renderne più consapevole l'uso e contemporaneamente cercare di capire quali siano state le ragioni che hanno convinto tanti studiosi a utilizzarlo¹.

¹ Modello di ricerca sul senso e il significato di un termine nel tempo rimane, per la sua completezza analitica e per l'ampiezza interdisciplinare, il classico studio, solo recentemente pubblicato, di Robert Merton e Elinor Barber sulla *seredipity* (Merton e Barber 2002).

La ricerca che qui si presenta nasce dalla tesi presentata per il conseguimento del Diploma di bibliotecario presso la Scuola Speciale per Archivistici e Bibliotecari dell'Università di Roma - La Sapienza. La tesi ha avuto come relatrice la professoressa Paola Castellucci, che ringrazio per i molteplici suggerimenti e il costante incoraggiamento. Ringrazio inoltre il professore Giovanni Solimine per i suoi preziosi consigli. La responsabilità di ogni errore è ovviamente da attribuirsi a chi scrive.

Alcuni testi che utilizzano il termine in questione possono darci una prima idea delle problematiche in discussione. Nel recente saggio sulla definizione e catalogazione delle risorse elettroniche, Stefano Gambari e Mauro Guerrini indicano nella *granularità* una caratteristica specifica di quelle risorse, definendola come

«la caratteristica della risorsa digitale di articolarsi in parti componenti (per esempio, sito Web, sezione di più pagine, singola pagina, sezioni della singola pagina, immagini, *link*); sono le parti atomiche o *lessie* che costituiscono i microtesti, i moduli di un ipertesto: il lettore dispone di varie possibilità di configurazione e di assemblamento delle unità atomiche che compongono il testo elettronico» (Gambari e Guerrini, 2002, p. 49-50).

I vantaggi per i lettori sono potenzialmente rilevanti: verificare facilmente se in un testo appare o meno un termine o una frase; accedere solo a parti specifiche di testo, senza dover procedere a una lettura sequenziale; riutilizzare facilmente il testo recuperato per costruire altri percorsi o altri documenti. Il termine *lessia* era stato introdotto da Roland Barthes per denotare le “unità di lettura” ritagliate, quasi “scomposte” in senso cinematografico, all’interno del testo, ed è poi entrato stabilmente nella letteratura sull’ipertesto².

Leggermente diversa è la definizione del termine riconosciuta da Maria Guercio:

«L’innovazione tecnologica consente soluzioni progressivamente più raffinate e differenziate per gli utenti e, quindi, sempre meno standardizzabili e riducibili a formati unitari per chi è impegnato nella conservazione. La cosiddetta ‘granularità’ dell’informazione – un termine ricorrente nella letteratura tecnica per indicare la progressiva riduzione di oggetti un tempo unitari a un insieme di elementi informativi sempre più piccoli e tendenzialmente indipendenti – presenta rischi e difficoltà crescenti per chi ha il compito di rendere disponibile l’informazione all’interno di contesti storicamente definiti» (Guercio 2002, p. 165).

Cercando di identificare un nucleo comune di significato in queste due citazioni, si può avanzare l’ipotesi che granulare è un documento digitale articolato in più parti componenti (anche di diversa natura: testo, immagini, suoni, ecc.), che consente al lettore piste autonome di lettura e di ricerca. Gambari e Guerrini sottolineano le possibilità di ricerca e lettura per il lettore, Guercio evidenzia i rischi per una lettura decontestualizzata e per le difficoltà di gestione di quella che può diventare un’enorme mole di documenti.

Una riflessione interessante è offerta dalla recente traduzione in lingua italiana di un intervento tenuto da John D. Byrum –Jr. alla 71– a IFLA Conference di Oslo.

² Per i riferimenti specifici ai testi di Roland Barthes, e in particolare alla sua lettura di *Sarrasine* di Balzac, rinvio al testo di Landow (Landow 1997).

Il testo inglese presenta in due passaggi il termine in questione, tradotto in italiano in maniera non letterale.

Testo in inglese (Byrum 2005)

Increasingly, users do not depend on the OPAC to reach the resources in which they are interested. [...] Around the year 2000, however, new library systems for responding to *more granular information requests* [corsivo mio] began to appear

The Library of Congress recently commissioned a study by Marcia Bates for the purpose of exploring ways to enrich metadata records by focusing on providing additional subject and other access mechanisms (e.g. front-end user thesauri) and *increasing granularity of access and display* [corsivo mio]

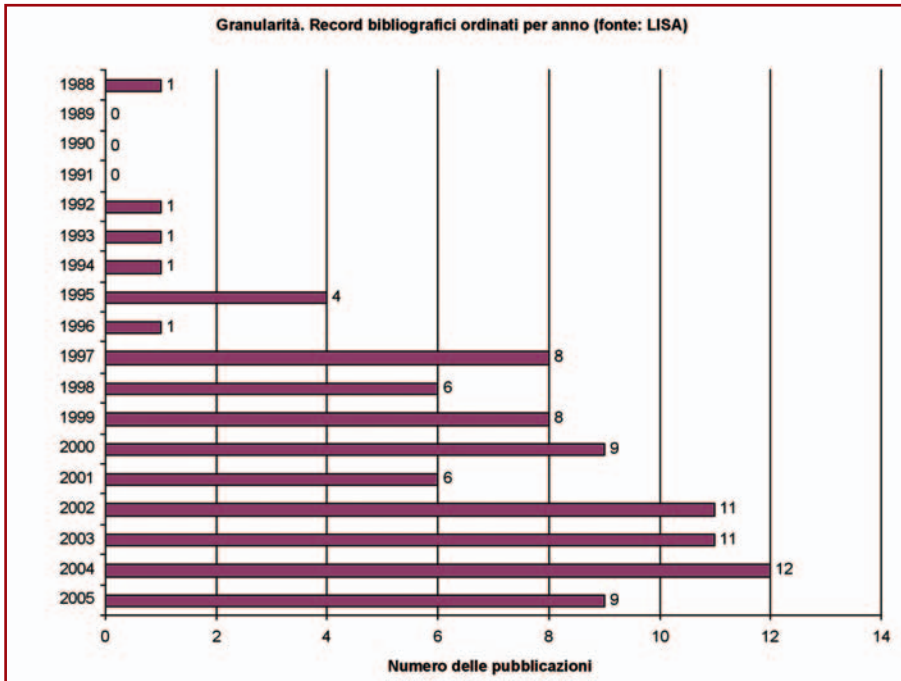
Traduzione italiana

Sempre più di frequente, gli utenti non si affidano agli OPAC per raggiungere le risorse di loro interesse. [...] Intorno al 2000, comunque, sono cominciati a emergere nuovi sistemi di biblioteca pensati per rispondere a *esigenze informative più sottili* [corsivo mio]

Di recente la Library of Congress ha incaricato Marcia Bates di effettuare uno studio avente l'obiettivo di esplorare le modalità di arricchimento dei record di metadati fornendo ulteriori meccanismi di accesso quali quelli per soggetto e altri (ad esempio, i tesauri di consultazione immediata per l'utente) e *maggiore sottigliezza di accesso e visualizzazione* [corsivo mio]

Granular e *granularity* sono stati tradotti con *sottile* e *sottigliezza*. Evidentemente il traduttore ha ritenuto ancora non ben definito e condiviso il significato dei due termini e ha scelto un termine più diffuso. In questo modo però si nasconde l'ampia diffusione del termine nella letteratura angloamericana, e soprattutto il secondo brano non appare del tutto chiaro, e non del tutto legato al passaggio successivo in cui la ricerca di Marcia Bates propone un'integrazione della ricerca nell'opac tramite l'uso di una struttura tesauroale. Come vedremo in seguito, l'uso del termine *granulare* da parte di Byrum, invece, nasce dalle molteplici riflessioni che negli ultimi dieci-quindici anni sono state elaborate dalla comunità internazionale dei bibliotecari e dei documentalisti sulle caratteristiche delle risorse informative e degli strumenti necessari per l'identificazione, la descrizione e il recupero dei documenti elettronici. In questi dibattiti il termine *granularità* è stato utilizzato con una frequenza via via maggiore³.

³ Il grafico illustra l'aumento numerico dei record bibliografici presenti sul database *Library and information science abstracts (LISA)* in cui appaiono i termini *granular* o *granularity* nel titolo o nell'abstract. L'analisi delle parole chiave controllate assegnate a quegli stessi record conferma che i due termini appaiono in ambiti disciplinari non sempre omogenei e con significati diversi (il dato relativo al 2005 è incompleto).



Granularity nel repertorio bibliografico *Library and information science abstracts*, al 10.2.2006. La *query* effettuata in quell'occasione è stata: granular*

Non vorrei anticipare riflessioni che sarà bene far maturare lungo il percorso di questa ricerca. Una breve indagine su dizionari, enciclopedie, basi dati e motori di ricerca dovrebbe chiarire i significati attribuiti al termine nel linguaggio comune e offrire riferimenti più precisi per proseguire la ricerca.

I viaggi della granularità, dal concreto all'astratto

L'analisi effettuata dal *Grande dizionario della lingua italiana* di Salvatore Battaglia è ampia (Battaglia 1961-2002). Granulare è definito come ciò che è

«ridotto in granuli; composto di granuli; costituito da elementi a forma di granuli (una sostanza). Per estensione: ruvido, scabro (una superficie); granuloso».

Sono attestati riferimenti nell'ambito della geologia («struttura granulare o granitica o granitoides»), della biologia («che presenta granulazioni»), dell'elettronica⁴.

⁴ «Effetto granulare: fluttuazioni piccolissime e casuali dell'intensità di corrente anodica di un tubo termoelettronico a circuito chiuso, dovute al fatto che l'arrivo degli elettroni sull'anodo del tubo è discontinuo e irregolare (e si manifesta come un caratteristico rumore di fondo)».

Il termine *granularità* rimanda direttamente invece alla *granulazione*, nell'accezione fotografica. *Granulazione* sta per «ridurre una sostanza o una materia in granuli», e consiste in un

«processo con cui vengono ridotte in piccoli granuli sostanze polverulente (come concimi, medicinali, ecc.) o fuse (come metalli, vetro) allo scopo di renderne più facili il maneggio, la distribuzione, l'utilizzazione, il trasporto, ecc.)».

La granulazione avviene o si produce nell'oreficeria antica, nell'astronomia (granulazione solare), nella fotografia. Nel linguaggio tecnico dei fotografi, si parla di granulazione a proposito dell'

«aspetto eterogeneo della struttura dell'immagine fotografica, dovuta alla sovrapposizione e al modo di dispersione dei granuli d'argento nello strato dell'emulsione».

Sono attestati comunque diversi usi in senso concreto: «insieme di granuli, corpuscoli o tubercoli» in biologia e in medicina.

Il processo della granulazione è rintracciabile anche in testi letterari. L'esperienza visiva dei colori consente quasi di coglierne l'effetto materiale e figurato. Ardengo Soffici, per esempio:

«Chiunque non è sensibile al sapore pittorico di un frego di carbone fatto da un ragazzo sul muro, al vellutato della sua granulazione nera sull'intonaco, sarà per sempre chiuso al godimento della bellezza plastica».

Emilio Cecchi, invece:

«La sua frase accompagna la linea del contorno e risponde alle sue varie tensioni, fermenta nella granulazione d'un impasto coloristico».

A parte questi ultimi, i significati attestati sono comunque tutti fortemente concreti. A sua volta, il termine "granulo" sta per «piccolo grano, granello». In particolare, il granulo è «ciascuno degli elementi che costituiscono alcuni materiali compatti (come il granito) o sciolti (come la sabbia)». Significati specifici sono attestati in astronomia (granuli del sole), in citologia e batteriologia, in botanica, in istologia, in chimica, in farmacia, e in fotografia.

«Fotogr. Ciascuno dei microcristalli d'argento che si formano nell'emulsione fotografica in seguito all'esposizione alla luce e all'azione del bagno di sviluppo e che formano l'immagine fotografica; il loro insieme, grana».

Il fenomeno chimico descritto si trasforma poi in immagine e in effetto visivo. Il significato registrato nelle discipline fisiche, invece, è forse uno dei più interessanti, e lo possiamo considerare un'altra traccia importante per il prosieguo:

«Fis. *Granulo di energia*: nome comune del *quanto* di energia».

La composizione, la tessitura, l'aspetto stesso dei corpi e degli oggetti costituiti da granuli o granelli più o meno piccoli è espressa tramite il sostantivo *grana*. In un'emulsione esposta alla luce e sviluppata, il termine *grana* esprime il numero e la consistenza dei granuli microscopici di sali d'argento, considerati in proporzione alla superficie di una pellicola. La grana di una pietra o di un metallo sono evidenziati soprattutto al momento della frattura, e la finezza o grossolanità di questi materiali dipendono dalla grandezza relativa dei grani in cui la massa si suddivide e dalla maggiore o minore vicinanza, per cui la massa assume un aspetto più o meno compatto.

Il termine *grana* è utilizzato anche per rendere l'aspetto più o meno ruvido di una superficie, di una pelle, di un tessuto o di una superficie metallica trattata da un artigiano, da un artista, da un orafo. La stessa voce umana ha una grana. È Roland Barthes, con una geniale sinestesia, a parlare di *grana della voce*, caratteristica unica di ogni voce individuale, e che ne rende il riconoscimento e l'ascolto un'esperienza unica e irripetibile⁵.

Per riconoscere e apprezzare l'individuale grana o granularità di un oggetto è dunque necessario attivare i propri sensi. Dalle caratteristiche più concrete, che si possono toccare o vedere, si passa a caratteristiche più astratte, che si colgono con l'udito o la vista (anche attraverso lo strumento fotografico), fino addirittura a utilizzare lo strumento della ragione per comprendere il quanto di energia.

Negli ambiti scientifici e tecnici, il termine *granularità* ha acquisito ulteriori significati. Il Dizionario tecnico scientifico McGraw-Hill Zanichelli riporta alla voce *granularity* tre significati, legati a tre ambiti disciplinari diversi:

«GRAF [ARTI GRAFICHE]. The distribution of grains in a portion of photographic material that has been uniformly exposed and processed.

ING SIS [INGEGNERI DEI SISTEMI]. The degree of which a system can be broken down into separate components, making it customizable and flexible.

⁵ Riporto al proposito alcune indicazioni di un recente libro sulla filosofia dell'espressione vocale di Adriana Cavarero: «Secondo Roland Barthes [...] il proprio della voce starebbe invece in quel che egli chiama la sua grana. Piuttosto che la sfera del soffio e del respiro, essa riguarda 'la materialità del corpo che sgorga dalla gola, là dove si forgia il metallo fonico'. L'attenzione va dunque alla cavità orale, luogo erotico per eccellenza. La grana della voce concerne soprattutto il modo in cui, mediante la voluttà dell'emissione sonora, la voce lavora nella lingua. [...] Compito della voce è dunque quello di fare da tramite o, meglio, da snodo, fra corpo e parola» (Cavarero 2003; cito dalla versione online http://www.feltrinelli.it/SchedaLibroEstratto?id_estratto=1011).

PETROGR [PETROGRAFIA]. The feature of rock texture relating to the size of the constituent grains or crystals» (McGraw-Hill Zanichelli 2004).

Il significato del termine nell'ambito delle arti grafiche e della fotografia o nell'ambito della petrografia è quello attestato anche nei dizionari della lingua italiana, ma emerge ora anche la disciplina dell'ingegneria dei sistemi. La granularità indica il grado in cui un sistema può essere articolato in componenti separate, così da renderlo flessibile e adeguabile alle esigenze degli utenti. Queste esigenze possono essere soddisfatte tramite l'utilizzo degli strumenti informatici. È proprio l'ambito informatico a fare la parte del leone nei significati più diffusi del termine, e la conferma la si può ottenere anche utilizzando l'archivio di Google come una sorta di corpus, da cui estrarre le definizioni più in uso. La serie dei significati si allunga di giorno in giorno, e la grande maggioranza fa riferimento all'ambito informatico, in particolare alla progettazione dei data base, alla programmazione parallela, al calcolo parallelo, all'elaborazione del testo e all'organizzazione delle memorie. Il *Dictionary of Computing* tenta di sintetizzare il significato del termine:

«A measure of the size of the segments into which memory is divided for purposes of either memory protection or virtual-memory management»⁶.

Anche *Whatis.com*, ottimo strumento per muoversi tra le definizioni relative al mondo dell'information technology e di Internet, tenta di tirare le fila, con una definizione ampia del termine in grado di abbracciare tutti i diversi settori disciplinari.

«Granularity is the relative size, scale, level of detail, or depth of penetration that characterizes an object or activity. It may help to think of it as: which type of 'granule' are we looking at? This term is used in astronomy, photography, physics, linguistics, and fairly often in information technology. It can refer to the level of a hierarchy of objects or actions, to the fineness of detail in a photograph, or to the amount of information that is supplied in describing a person's age. Its meaning is not always immediately clear to those unfamiliar with the context in which it's being used»⁷.

La definizione intende "coprire" diversi significati. In tutte le discipline, la granularità indica il livello di dettaglio, la precisione con la quale si può cogliere un oggetto o un'attività. Non sono dati riferimenti precisi (misure, numeri, ecc.), visto che si può usare in tantissimi contesti e si può adattare alle esigenze di ciascun utente.

⁶ *Granularity*, in: *A Dictionary of Computing*, Oxford: Oxford University Press, 1996, consultato su *Oxford Reference Online*, <http://www.oxfordreference.com>.

⁷ Cfr. <http://whatis.techtarget.com/definition/0,289893,sid9gci212209,00.html>.

Rischiosa appare comunque la scelta di enumerare, quasi confondendoli, i diversi significati di dimensione relativa, livello di dettaglio, scala, profondità, quasi siano termini intercambiabili⁸.

Vale la pena chiedersi finalmente come si sia legato il termine di *grana* o di *granularità* al concetto e al termine *informazione*. Concepire come granulari l'informazione, un supporto documentario o uno strumento di computazione o ricerca appare come un'innovazione molto recente. Una ricerca sull'*Oxford english dictionary*, per esempio, non attesta storicamente significati assimilabili. L'ampia diffusione del termine nel settore informatico segnala che probabilmente è stato questo il canale privilegiato attraverso il quale il termine è entrato nello specifico delle scienze del libro e del documento. Credo però che il percorso non sia stato del tutto lineare, e probabilmente il percorso concettuale non sempre diretto. Le indicazioni raccolte suggeriscono che la diffusione del termine possa essere avvenuta attraverso percorsi molteplici e non sempre lineari. Le direzioni di ricerca sulle quali ritengo opportuno portare l'attenzione sono sostanzialmente tre.

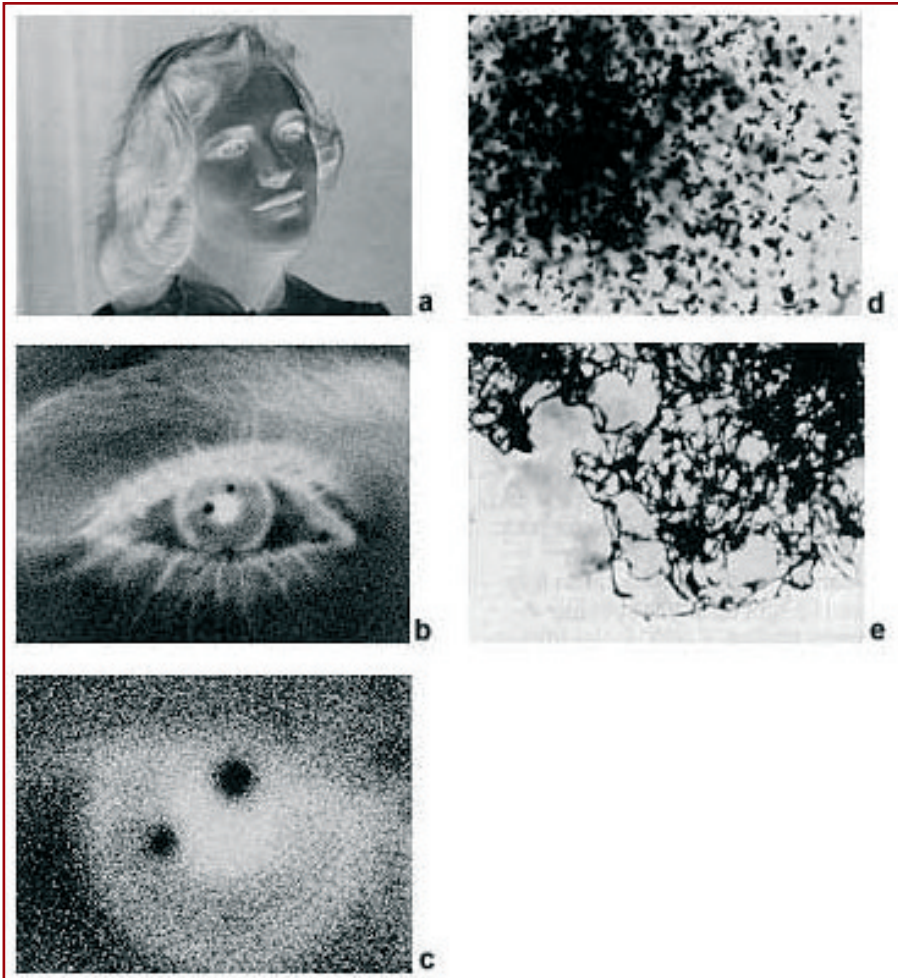
La prima riguarda il significato fotografico del termine granularità. Tra tutti i significati del termine, questo appare forse il più diffuso e comune. L'effetto granulare di un'immagine è basato sulle caratteristiche fisiche della pellicola fotografica, ma produce ben precise conseguenze visive. Si tratta di un effetto ben conosciuto, che si verifica quando, "zoomando" progressivamente, l'occhio percepisce la grana della fotografia, fino al punto in cui tale visione diventa sufficientemente fastidiosa per l'occhio e per le nostre capacità cognitive, l'immagine diventa granulosa e si finisce per perdere il contesto in cui si colloca il particolare che si intendeva cogliere. La sequenza di immagini di seguito riportate – dal titolo *Graininess and granularity area* – mostra come la granularità di una immagine fotografica si possa slabbare e sfrangere fino a perdere di significato per chi guarda.

Da un sito specificamente dedicato alla fotografia possiamo cogliere i significati tecnici dei termini in questione.

«GRANA. Osservando un'immagine negativa al microscopio o un forte ingrandimento si possono notare i piccoli ammassi di argento metallico che formano l'immagine dopo lo sviluppo. Questi ammassi prendono il nome di grana. [...]

GRANULARITÀ. Quantificazione oggettiva del concetto di grana. Indica la mancanza di uniformità della Densità di un'emulsione fotografica ovvero lo spostamento dei valori di Densità rispetto ad uno standard. Questo dato viene espresso dai fabbricanti in termini di granularità RMS diffusa. Più alto è il valore maggiore è la grossezza della grana. Un valore 8 è tipico per una pellicola 100 Iso.

⁸ Non stupirà che la diffusione del termine abbia finito per trasformare la *granularità* addirittura in una *buzzword*, un termine di moda, usato per impressionare gli sprovveduti: (<http://www.buzzwhack.com/buzzcomp/indgk.htm>).



A 2.5X enlargement of a negative shows no apparent graininess. (b) At 20X, some graininess shows. (c) When a segment of the negative is inspected at 60X, the individual silver grains start to become distinguishable. (d) With 400X magnification, the discrete grains are easily seen. Note that surface grains are in focus while grains deeper in the emulsion are out of focus. The apparent 'clumping' of silver grains is actually caused by overlap of grains at different depths when viewed in two-dimensional projection. (e) The makeup of individual grains takes different forms. This filamentary silver, enlarged by an electron microscope, appears as a single opaque grain at low magnification. <http://www.kodak.com/US/en/motion/students/handbook/sensitometric6.jhtml>.

GRANULOSITÀ. Impresione soggettiva della grana che appare in un'immagine fotografica. Essa dipende dalla distanza di osservazione, dalle condizioni di visione e dal visus dell'osservatore»⁹.

⁹ http://www.reflex.it/enciclopedia/enciclopedia_g-i.htm.

Se è stata esperienza comune per tanti anni osservare i piccoli puntini che caratterizzavano le fotografie riprodotti sui giornali quotidiani (effetto dovuto a un particolare e più veloce per l'epoca sistema di trasmissione), oggi osserviamo i video dei computer e parliamo della loro risoluzione esprimendoci in pixel¹⁰, come esprimiamo con il numero dei pixel anche la risoluzione di un'immagine. Un procedimento comune è quello di ingrandire anche le immagini digitali cogliendo proprio le unità minime che le compongono.

L'effetto granulare si può associare anche all'uso dello zoom, ovvero dell'obiettivo fotografico a distanza focale variabile che consente di ampliare o ridurre l'inquadratura. Lo zoom dovrebbe mantenere sempre a fuoco l'immagine, ma il movimento di esaltazione dei particolari e l'ansia di ampliare l'inquadratura può portare a perdere il riferimento al contesto, che solo riesce a farci comprendere appieno il significato di quell'immagine.

La diffusione attuale delle immagini fisse e in movimento e dei sistemi di riproduzione digitale — video, televisione, cellulari, ecc. — ha senz'altro giocato fortemente a favore della diffusione del termine granularità. Come si esprimeva un recentissimo articolo apparso su un giornale non scientifico, *The Christian science monitor*, "We're all going dotty" (Walker 2006)¹¹.

La seconda traccia da seguire rimanda al mondo delle scienze fisiche. Il "granulo di energia" citato da Battaglia come «nome comune del quanto di energia» riporta alla ricerca effettuata dai fisici nel settore dell'infinitamente piccolo, in particolare sulla composizione e le caratteristiche del mondo subatomico, e le lunghe, secolari diatribe sulle caratteristiche ondulatorie o corpuscolari della luce. Il tentativo di identificare la realtà fisica più piccola ha portato a teorie sempre più sviluppate e a ipotesi radicali, come quelle della fisica quantistica, per la quale l'uso stesso degli strumenti di misurazione rende impossibile determinare con precisione meccanistica velocità e posizione delle particelle subatomiche. Ogni realtà fisica subatomica diventa sempre più piccola e astratta, e le caratteristiche fondamentali dei "granuli" più piccoli non sono la massa o l'energia, bensì le informazioni trasportate. Su questa base, l'informazione viene concepita da alcuni come linguaggio universale della nuova scienza fisica¹².

Alla metà del secolo scorso, negli anni successivi all'affermazione della fisica quantistica, un fisico prestigioso, Donald Mackay, ipotizzava che, dato che ogni

¹⁰ "Pixel. Un pixel (abbreviazione inglese di picture element) è uno dei molti minuscoli puntini che compongono la rappresentazione di un'immagine nella memoria di un computer. Solitamente i punti sono così piccoli e così numerosi che se vengono stampati su carta o visti su un monitor appaiono fusi in una unica immagine. Colore e intensità di ciascun punto sono usati dal computer per rappresentare una piccola area dell'immagine" (<http://it.wikipedia.org/wiki/Pixel>).

¹¹ L'autore associava i punti (dot) ai pixel e grani di sabbia, approfondendo inoltre il significato di *granularity*.

¹² Sull'argomento, cfr. il recente libro divulgativo del fisico Hans Christian Von Baeyer, dal titolo significativo: *Informazione. Il nuovo linguaggio della scienza* (Baeyer 2004).

realtà fisica presuppone comunque uno scambio di informazioni, si potesse identificare una «unit of structural information [...] for this unite there already exists the suitable name of a *logon*»¹³:

«Mackay suggests that the apparently ‘hard’ science of physics is epiphenomenal to the real science of information: it was because information was inherently quantal that physical reality was discovered to have the same property (Mackay, 1950)» (Rayward 1996).

La strada è diversa da quella battuta da Shannon e Weaver, che misuravano la quantità e la qualità dell’informazione scambiata in una transazione comunicativa, e per questo coniarono il bit¹⁴. Nella prospettiva di Mackay, e di diversi studiosi nei decenni successivi, la ricerca di una unità minima di informazione era sostenuta dall’ipotesi che fosse proprio l’informazione lo strumento di base per l’interpretazione della realtà fisica, in tutti i suoi aspetti.

La terza strada da seguire riguarda la scienza dell’informazione in senso lato, o per meglio dire l’informatica e la scienza dei computer. In queste discipline l’uso è oggi attestato massicciamente¹⁵, e la consultazione dei materiali disponibili in Rete lo può confermare. La consultazione di INSPEC, il repertorio bibliografico prodotto dall’Institution of Electrical Engineers di Londra¹⁶, permette di tentare di cogliere il momento in cui il termine entra nel linguaggio informatico. Se dalla fine del XIX secolo il significato del termine è fortemente ancorato al contesto fotografico, a partire dal 1969-1970 articoli si occupano della granularità dei materiali utilizzati nella registrazione di informazioni in formato digitale, e poi nel 1976 il sostantivo *granularity* appare usato in un senso nuovo, più precisamente nell’ambito della ricerca sulla gestione ottimale dei data base¹⁷. L’uso astratto, e vorrei dire già propriamente metaforico, del termine granularità (o granuli) si conferma negli anni successivi e si estende alla ricerca sui sistemi di gestione dei data base, prima di

¹³ Traggo l’indicazione della citazione da un intervento di W. Boyd Rayward sulla storia e la storiografia dell’*information science* (Rayward 1996, p. 66).

¹⁴ «Unità di misura della quantità di informazione. Non esistono sottomultipli del bit: la minima quantità d’informazione è pari a 1 bit ed equivale alla scelta tra due valori (sì/no, vero/falso, acceso/spento), quando questi hanno la stessa probabilità di essere scelti» ([http://it.wikipedia.org/wiki/Bit_\(informazione\)](http://it.wikipedia.org/wiki/Bit_(informazione))).

¹⁵ L’interrogazione della base dati bibliografica multidisciplinare *Current Contents* conferma per il periodo 2001-2005 come proprio la *computer science* sia nettamente il settore disciplinare in cui il termine è più diffuso, superando le discipline delle scienze dei materiali e registrando un’elevata diffusione anche nei settori affini dell’information technology e dell’intelligenza artificiale e della robotica.

¹⁶ INSPEC indicizza articoli di riviste e atti di convegno nei settori dell’ingegneria informatica ed elettronica, della fisica applicata e dell’ottica.

¹⁷ Gray-JN – Lorie-RA – Putzolu-GR – Traiger-IL, *Granularity of locks and degrees of consistency in a shared data base*, in: Nijsen-GM, *Proceedings of the IFIP Working Conference on Modelling in Data Base Management Systems*, North-Holland: Amsterdam, 1976, p. 365-94.

espandersi ulteriormente in altri settori delle scienze dell'informazione.

I tre percorsi che ho brevemente indicato costituiscono il quadro di riferimento più ampio al cui interno si potrebbe essere svolta questa possibile vicenda di "contrabbando" di significato tra diverse regioni disciplinari confinanti. All'interno di questa cornice, vale la pena indicare almeno altre due aree fortemente interdisciplinari che utilizzano in maniera estremamente interessante il termine *granularità* e che ritengo possono avere concorso alla diffusione del termine nel campo delle discipline del libro e del documento.

Il termine *granularità* è infatti attestato nei manuali e nelle ricerche che si occupano di sistemi di informazione geografici (GIS). Il *GIS Dictionary* della ESRI (un'importante azienda produttrice di software), alla voce *Granularity* riporta due significati:

- «1. The coarseness or resolution of data. Granularity describes the clarity and detail of data during its capture and visualization.
2. The objective measure of the random groupings of silver halide grains into denser and less dense areas in a photographic image»¹⁸.

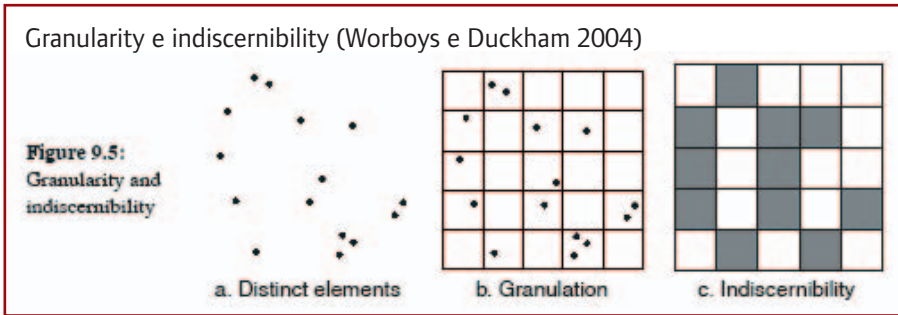
Come si vede il termine ha nel primo caso un'accezione abbastanza generale relativa al settore dell'elaborazione dei dati nei sistemi informativi territoriali, e poi riprende la definizione tecnica di origine fotografica. Una più ampia e interessante definizione la offre un accreditato e aggiornato testo di riferimento dedicato ai GIS, i cui esempi, che utilizzano anche il linguaggio visuale, danno un senso più chiaro al termine.

«Granularity is closely related, but not identical to imprecision. Granularity refers to the existence of clumps or grains in observations or representations, in the sense that individual elements in the grain cannot be distinguished or discerned apart from each other.

For example, imagine a remotely sensed image of a region of the Earth. In this image, any two locations on the ground that fall within the same pixel cannot be distinguished in characteristics by the observation. So, granularity in an observation makes things *indiscernible* that in principle are distinguishable from each other» (Worboys 2004).

Il problema è il trattamento di un numero rilevante di dati, che debbono essere resi intellegibili in fase di osservazione o rappresentazione, selezionando un particolare livello di dettaglio. La scelta di un determinato dettaglio o grado di precisione comporta sempre una perdita di informazioni, per cui il termine *granularità* si associa anche all'imprecisione e all'indiscernibilità dei dati rilevati, registrati ed elaborati.

¹⁸ <http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.Search&search=true&searchTerm=granularity>.



La letteratura sui GIS potrebbe assumere un notevole interesse per questa ricerca, come snodo interdisciplinare dove s'incontrano le scienze del territorio, dell'informazione e le scienze cognitive. In pratica, l'osservazione di dati relativi alla realtà geografica (fenomeni, processi ed entità fisiche, ma anche economiche e sociali) avviene tramite più passaggi, quello fotografico (dove cioè sono implicati i processi chimici della definizione dell'immagine), quello della rappresentazione tramite riproduzioni cartografiche (che usano da secoli lo strumento della rappresentazione del territorio su scale diverse) e quello dell'implementazione di data base numerici.

La rappresentazione e la ricerca di informazioni in questo settore richiede dunque competenze e abilità cognitive diverse, che debbono collaborare strettamente. Non credo possa sorprendere che questo sia un ambito disciplinare di grande interesse per un possibile sviluppo della ricerca sulla granularità. Proprio seguendo il filo di queste problematiche, è comprensibile che il termine granularità appaia in un interessante riflessione offerta da un recente intervento di alcuni studiosi di GIS (Frederico Fonseca, Max Egenhofer, Clodoveu Davis e Gilberto Camara), che hanno sottolineato sì la sempre maggiore disponibilità di dati relativi al nostro pianeta, ma anche come queste sistemi informativi siano sempre più distribuiti ed eterogenei. Ciò costituisce una vera sfida alla ricerca, il cui problema diventa quello di integrare le diverse informazioni, raccolte ed elaborate a differenti livelli di dettaglio o di granularità. Se facciamo riferimento all'immagine del testo di Worboys e Duckham, quando debbono dialogare data base diversi la griglia utilizzata per raccogliere e manipolare gli originali dati granulari deve essere simile, pena la non comunicabilità.

«Frequently, the information exists, but integration is very difficult to achieve in a meaningful way because the available information was collected by different agents and also with diverse purposes. The effective integration of multiple resources and domains is known as interoperation. [...] In the past, exchanging geographic information was as simple as sending paper maps or raw data tapes through the mail. Today, computers throughout the world are con-

nected and the use of GIS has become widespread. The scope of interoperability has changed from static data exchange using flat files to global systems, interconnected using sophisticated protocols to exchange information on-line. In the future, computers are expected to be able to share not only information but also knowledge» (Fonseca e al. 2002).

Il problema di integrare o interoperare archivi diversi è dunque anche un problema di «different levels of detail», in pratica un problema di granularità, che i computer, da soli, non riescono a risolvere.

Per completare queste avventurose incursioni in campi disciplinari così lontani dalle abituali letture di un bibliotecario, vale la pena sottolineare come secondo alcuni uno strumento in grado di coordinarsi velocemente e in maniera flessibile tra più sistemi informativi, spesso a livelli diversi di granularità, esista già: si tratta del cervello umano. Se è consentito pescare nella letteratura dell'intelligenza artificiale – disciplina che mantiene da decenni la leadership tra le scienze della cognizione e dell'informazione – vale la pena leggere almeno la parte introduttiva di un breve ma fondamentale articolo pubblicato da Jerry Hobbs una ventina di anni fa. Il titolo appare decisamente pertinente: *Granularity*.

«We look at the world under various grain sizes and abstract from it only those things that serve our present interests. Thus, when we are planning a trip, it is sufficient to think of a road as a one-dimensional curve. When we are crossing a road, we must think of it as a surface, and when we are digging up the pavement, it becomes a volume for us. [...] Our ability to conceptualize the world at different granularities and to switch among these granularities is fundamental to our intelligence and flexibility. It enables us to map the complexities of the world around us into simple theories that are computationally tractable to reason in» (Hobbs, 1985, p. 432).

Questo approccio suggerisce a Hobbs che quello della granularità è un modo di concepire l'intelligenza tipico degli esseri umani, e che le macchine dovranno per forza avere.

«It is that our knowledge consists of a global theory together with a large number of relatively simple, idealized, grain-dependent, local theories, interrelated by articulation axioms. In a complex situation, we abstract the crucial features from the environment, determining a granularity, and select the corresponding local theory» (Hobbs 1985, p. 435).

Come dotare le macchine di questa capacità di *switchare* tra diversi livelli di dettaglio dei sistemi informativi e delle realtà fisiche è problema che ha interessato non

solo gli studiosi dell'intelligenza artificiale, ma anche gli studiosi delle scienze dell'informazione e i ricercatori di molteplici settori disciplinari, che in questi anni si sono avvicinati al tema delle ontologie. Per quanto riguarda le discipline del libro e del documento, sicuramente i tesauri disciplinari e i sistemi classificatori possono funzionare in questo senso, sebbene non in forma automatica. La riflessione degli studiosi su questo tipo di problemi si è recentemente risvegliata soprattutto da quando si è imposto il tema delle ontologie. Avrò modo di sfiorare di nuovo questo argomento.

Tutti i percorsi intellettuali qui troppo velocemente indicati hanno visto un progressivo abbandono dei significati più concreti del termine, con il passaggio a un uso sempre più astratto e quasi metaforico. In questa progressiva mutazione dalla definizione delle caratteristiche concrete e fisiche di un oggetto granulare all'analisi di una realtà impalpabile e incorporea come quella dei dati, delle memorie, dell'informazione in generale, l'uso metaforico ha forse approfittato dell'abitudine all'uso del termine da parte dei ricercatori di formazione tecnica o scientifica. Il termine granularità al posto del più tradizionale "dettaglio" potrebbe avere dato agli utilizzatori una sensazione di maggiore scientificità. Nello stesso tempo, avere perso il riferimento concreto – le diverse forme e composizioni di minerali e metalli – può avere contribuito a espanderne l'uso e i significati fino quasi a perdere molto della precisione e affidabilità scientifica, e la ricerca informatica potrebbe avere finito per utilizzarlo in senso metaforico. A ogni livello, infatti, l'informatica ha visto un'elevatissima diffusione di ogni tipo di metafora legata all'interazione uomo-macchina. Non c'è praticamente alcun termine di quotidiano uso informatico che non sia fortemente segnato da metafore (windows, mouse, banche dati, rete, ragni, agenti intelligenti, motori di ricerca, data mining, ecc.)¹⁹. L'ambito informatico non può non utilizzare metafore in cui è fortemente implicato il senso della vista. Alla diffusione della metafora della granularità può dunque avere giovato la possibilità di "pescare" via via da esperienze concrete come quella della scienza dei materiali, dell'ottica e della fotografia, per imporsi nel campo informatico – settore disciplinare oggi al centro di numerosissime intersezioni disciplinari – e da qui sbarcare nell'area delle scienze del libro e del documento²⁰.

¹⁹ Gerald Johnson sostiene che l'impiego della metafora nel mondo della *computer science* ha una sua specificità che spiega la sua particolare diffusione, e cioè che oltre a descrivere cose innovative, serve a dare consistenza a realtà talmente astratte da risultare altrimenti incomprensibili e di difficile utilità pratica (Johnson 1994). Ronald Day ricorda come gli stessi Shannon e Weaver usassero termini metaforici in diversi testi collocati alle origini della teoria della comunicazione (Day 2000).

²⁰ L'uso delle metafore – e in particolare di quelle che hanno utilizzato esplicitamente anche il linguaggio delle immagini, attivando gli stili cognitivi visivi di tanti ricercatori – ha sempre avuto anche in queste discipline un'importanza rilevante. In particolare, Werner Bies offre diverse suggestioni e ipotesi di ricerca sull'uso delle metafore nel campo della *knowledge organization*, per esempio quella dell'albero del sapere (Bies 1996).

L'affermazione della granularità come metafora e ponte di transizione tra settori disciplinari originariamente distinti credo possa rientrare in un processo tipico dello sviluppo del sapere. È anzi ampiamente accettato che in ambito scientifico il momento più propriamente creativo è difficilmente formalizzabile, al di là di quanto non appaia nel momento della formulazione teorica e della pubblicazione formale a stampa. «Formulazioni di tipo analogico o metaforico, attraverso invenzioni sperimentali o utilizzazioni eretiche di sistemi precedenti», collaborano in maniera complessa alla costruzione di sperimentazioni e teorie scientifiche.

«La formazione dei concetti [...] può avere anche a che fare con i linguaggi del senso comune. Il modo in cui si usa questo senso comune, così come le terminologie di discipline diverse da quella che è al lavoro, sono dati dalla pratica metaforica; attraverso questa pratica si formano concetti 'pescandoli' dai linguaggi di discipline diverse da quella in oggetto» (Gagliasso 2002, p. 5).

Secondo la teoria di Black²¹, le metafore sono distinguibili come sostitutive o esegetiche (pedagogiche) e metafore interattive o costitutive. Le prime costituiscono un "fragile ponte temporaneo" offerto all'intuizione in mancanza di spiegazioni causative o ricostruttive lineari. Anche le metafore sostitutive possono comunque svolgere una funzione costitutiva, aprendo possibilità di ricerca nuove, attraverso slittamenti progressivi del significato e un continuo scivolamento di significato da una disciplina all'altra. Viene da chiedersi se questa, come altre metafore scientifiche, consentirà di «allargare l'ambito osservativo e ritagliare diversamente, riorganizzandoli, i campi tematici» del nostro specifico disciplinare (Gagliasso, p. 6). Certamente credo che potrebbe essere stimolante rileggere alla luce di questa nuova metafora riflessioni teoriche e soluzioni pratiche prodotte nel corso dei secoli da bibliotecari e documentalisti e, più recentemente, dagli stessi informatici. Se proviamo a pensare al lavoro concreto di chi si occupa di testi e documenti – siano essi bibliotecari e documentalisti o informatici che si cimentano con le problematiche relative alle nuove realtà testuali e documentarie create in ambito elettronico e digitale – è evidente che alcune delle domande di base che si pongono ai cultori delle diverse discipline sono in qualche modo largamente simili, anche per quanto riguarda la granularità e il livello "atomico" dell'informazione e della documentazione. Qual è il livello di dettaglio a cui può giungere il trattamento dell'informazione, o meglio del testo o del documento? Quale è la misura ottimale del "segmento di memoria" (scritta) in cui si può dividere un testo o un documento al fine della sua migliore indicizzazione, circolazione o conservazione? Quali realtà "discrete" o "granulari" si possono riconoscere e individuare in un documento o in

²¹ Riprendo la teoria della metafora di Max Black dal testo di Carmela Morabito sulla metafora nelle scienze cognitive (Morabito 2002).

un testo: una parola, una frase, un concetto, un'immagine?

La metafora della granularità credo abbia riportato alla ribalta temi e problematiche dalle profonde radici storiche. Ciò non per trovare a tutti i costi dei precursori, piuttosto per tentare di rileggere una parte della lunga storia delle discipline bibliografiche e documentarie, e del possibile intrecciarsi con la nascita delle scienze informatiche.

Storicamente, il documento cartaceo nella sua unità fisica ha costituito un supporto indispensabile per la registrazione, la conservazione e il recupero dei dati. L'unità documentaria è una realtà che il bibliotecario e documentalista può studiare, gestire, indicizzare, ma non può modificare. Costituisce anzi una sorta di vincolo al recupero dei dati registrati. La definizione che di unità documentaria danno Anderson e Perez-Carballo è in questo senso molto precisa.

«Documentary unit. The document, document segment, or collection of documents to which entries in an index refer and on which they are based. Examples of verbal documentary units include sentences, paragraphs, pages, complete articles, books, complete serial runs, collections of archival materials, microform sets, and entire library collections» (Anderson e Perez-Carballo 2005).

Individuare un documento all'interno di una biblioteca o di una collezione costituisce una delle tradizionali sfide dei bibliotecari, ma un ulteriore problema è stato per secoli proprio quello di "liberare" i dati contenuti nell'unità documentaria. La produzione di indici relativi a questi due problemi si è evoluta nella storia secondo metodi e con strumenti spesso molto diversi. Negli ultimi decenni, proprio il superamento dell'unità documentaria ha costituito una vera e propria sfida per la ricerca nel settore dell'*information retrieval*.

Non esiterei a rintracciare le radici del tema della granularità anche nella storia dell'ideazione e dello sviluppo dei sistemi di indicizzazione dei testi scritti, a partire dalla creazione di note, glosse, indici generali, tematici e alfabetici per l'accesso a punti o soggetti particolari dei libri (Rouse 1989). Tra gli strumenti paratestuali in grado di consentire una produttiva lettura non lineare del testo, Fayet Scribe annovera il riassunto e l'abstract, dai regesti diplomatici delle corti papali ai riassunti degli articoli scientifici, dal «Journal des Savants» alle riviste del nostro secolo. Il periodo finale del Medio Evo, in questa prospettiva, può essere considerato addirittura il periodo del "triomphe de l'indexation".

«Jamais, sauf peut-être à notre époque, l'intérêt ne s'est autant porté vers le contenu sémantique et l'usage qui en est fait. Du point de vue du repérage, la révolution actuelle du texte numérique serait à rapprocher de la 'révolution gothique', plus qu'à celle de l'imprimé de la Renaissance. [...]

De fait, l'apparition de l'imprimerie fait porter l'accent plus sur 'l'objet livre' lui-même que sur son contenu. À partir de la Renaissance jusqu'au milieu du XXe siècle, ce sont les voies d'accès par la référence (bibliographie, catalogue) et leur système de classifications qui seront les plus exploitées» (Fayet Scribe 1997).

L'indicizzazione completa del documento e la liberazione dei concetti e dei dati dai loro confini documentari è al centro anche della riflessione di Paul Otlet, fondatore della scienza della documentazione. L'archivio bibliografico ideato e implementato per anni da Otlet era basato sul *monographic principle*, secondo il quale occorre registrare analiticamente ogni documento o parte di documento dedicata a un determinato argomento. Il lavoro dell'indicizzatore doveva essere mirato non solo al livello dell'intera opera, di un suo capitolo o di una sua sezione, ma anche di ogni dato o "fatto" che fosse possibile estrarre dal testo (Rayward 1994, p. 169). Lo studioso che in fase di ricerca avesse voluto recuperare tutte le unità informative dedicate a un singolo argomento avrebbe potuto utilizzare lo schema costituito dalla classificazione decimale universale. La sicurezza di poter individuare i "fatti" e "ricostruire" la conoscenza, evitando le disfunzioni, i vincoli e le ripetizioni presenti in letteratura, costituisce ai nostri occhi la testimonianza di una precisa epoca storica, quella in cui il sapere scientifico poteva sperare di giovare di un sistema di accrescimento continuo e senza scosse, in cui ogni nuovo libro, ogni nuovo articolo aggiungeva ulteriori informazioni e conoscenze, senza stravolgere il quadro preesistente. La visione di Otlet mantiene comunque a tutt'oggi il suo fascino, soprattutto per le discipline scientifiche, in cui l'apprendimento è fortemente strutturato e l'ideale scientifico è quello di un sapere che cresce costantemente sulle esperienze precedenti.

Gli obiettivi di Otlet non erano molto diversi da quelli che si sono posti negli ultimi decenni i primi sistemi di *information retrieval*. Un'ipotesi di lavoro che appassionava informatici e documentalisti anche degli anni '60 era quella della possibilità di creazione di sistemi di "fact retrieval", distinti dai sistemi di "document retrieval" e di "data retrieval" (intesi come dati numerici). I sistemi di "fact retrieval" avrebbero dovuto essere in grado di rispondere a un utente in maniera diretta, piuttosto che recuperare un testo che può contenere la risposta o, più comunemente, il riferimento a tale testo. Secondo Frederick Wilfrid Lancaster e Amy Warner, quella del *fact retrieval* è stato «the most ambitious project in information retrieval» (Lancaster e Warner 2001, p. 43), ed è ancora all'ordine del giorno. Attualmente i sistemi di *information retrieval* consentono di ipotizzare non solo un efficiente ed efficace recupero delle unità documentarie, ma anche un altrettanto avanzato recupero delle informazioni salvate in "segmenti" più piccoli dei documenti, l'identificazione dei concetti sottesi e la produzione in forma automatica di nuova conoscenza. Particolarmente sviluppati gli approcci nati dall'elaborazione del linguaggio naturale (NLP, *natural language*

processing), e quello che viene chiamato *text extraction*²².

Le scienze dell'informazione hanno cercato dunque di accelerare in ogni modo la circolazione dei documenti scientifici e di superare le barriere "fisiche" che, anche secondo Otlet, frenavano la stessa ricerca scientifica. L'evoluzione delle tecniche ha visto una decisa accelerazione con l'affermazione del documento elettronico e la creazione di un nuovo strumento di recupero dell'informazione, l'ipertesto. In questo modo si sono aperte nuove strade alla realizzazione dell'antico sogno del documento granulare. Prima di parlare dell'ipertesto, è obbligatorio affrontare alcune altre riflessioni di ordine storico.

L'unità minima dell'informazione

Consultando il repertorio LISA, si può ragionevolmente riportare l'entrata della *granularità* nell'ambito biblioteconomico e documentalistico al 1992 quando il termine appare nel titolo dell'articolo di Frans C. Heeman, *Granularity in structured documents* (Heeman 1992). Un'attenta (e paziente) analisi dei molti risultati di una query su Google consente però di risalire a un interessante *abstract* di un intervento che già nel 1993 si poneva l'obiettivo di una possibile storia della granularità dell'informazione. Fu Stephen E. Robertson, uno dei più qualificati ricercatori nel settore dell'*information retrieval*²³, a presentarlo alla sessione di Storia dell'Information Science del Convegno annuale dell'American Society of Information Science.

«One particular theme, that of granularity of information, is followed through developments in information technologies (taken broadly), over the last four thousand years. Some classes of systems in the context of modern information technology are analyzed for their approach to information granularity. Different kinds of systems (e.g. wordprocessors, relational databases, text retrieval systems, knowledge-based systems) assume and use different levels of granularity» (Robertson 1993)²⁴.

²² Riprendo la descrizione della *text extraction* dalla monografia di Lancaster e Warner (Lancaster e Warner 2001, p. 44-46.) Per l'applicazione degli strumenti dell'intelligenza artificiale al settore dell'*information retrieval*, rimando per una breve introduzione a manuali di Heting Chu (Chu 2003, in particolare alle p. 229-236). L'evoluzione dell'analisi automatica del testo ha comunque dovuto fare i conti con la complessità del linguaggio naturale, che rende complessa e problematica la "liberazione" di dati e fatti dai vincoli dei supporti. Qualunque aspirazione in questo senso deve forzatamente tenere conto dei diversi livelli ai quali è necessario elaborare i testi, e cioè il livello fonetico, quello morfologico, lessicale, sintattico, semantico e pragmatico (Chowdhury 2004, p. 373).

²³ Stephen Robertson, tra gli altri, ha contribuito a sviluppare il modello probabilistico nel settore dell'*information retrieval*; cfr. al proposito *il manuale di information retrieval* di Heting Chu (Chu 2003, p. 106).

²⁴ Contattato personalmente per posta elettronica, il prof. Robertson mi ha confermato che non ha mai pubblicato il suo intervento e mi ha inviato gli appunti scritti con i quali aveva preparato l'intervento, indicandomi in particolare che l'oggetto della sua polemica fosse proprio il libro di Keith Devlin. Lo ringrazio per la sua squisita cortesia.

In questo intervento Robertson respingeva le teorie avanzate da Keith Devlin, che nel 1991 aveva ipotizzato l'esistenza logica di un'unità informazionale minima, battezzata *infon* (Devlin 1991). A parere di Robertson tale idea era fuori della realtà e della storia. Il tentativo di Devlin di fondare in senso logico-matematico una nuova scienza dell'informazione non poteva che scontrarsi con l'approccio di un informatico e di uno studioso dell'*information retrieval*, orientato a impostare ogni ricerca sulla concreta – e storicamente determinata – interazione uomo-macchina. Robertson sottolineava che l'informazione è «an human construct», un artefatto umano, nel quale non si possono individuare realtà atomiche reali come quelle che contraddistinguono il mondo della fisica. Tutte le unità che hanno consentito che l'informazione potesse essere elaborata (dai caratteri alfabetici al bit o al pixel), sono creazioni dell'uomo. Ogni tipo di sistema informatico in considerazione ha assunto come unità minime rappresentazioni diverse della realtà, o *different levels of granularity*. Le unità informative che si possono individuare nei documenti e che si possono manipolare grazie ai diversi software a disposizione non sono considerabili realtà fisiche o logiche preesistenti rispetto all'intervento e all'azione dell'uomo. Il quadro concettuale presentato da Robertson è chiarito dalle due tabelle che ho provato a ricostruire sulla base degli appunti gentilmente forniti dallo stesso autore.

Il tentativo di Devlin di identificare e teorizzare l'esistenza di una unità informazionale minima si può forse leggere come il prosieguo di un filone di ricerca quasi cinquantennale, iniziato dal grande successo della teoria di Shannon e Weaver²⁵. L'ipotesi di un *infon* non mi appare molto distante da quella del *logon* di Donald Mackay, almeno nel tentativo di identificare un'unità minima dell'informazione come punto di raccordo tra i molteplici settori disciplinari che negli ultimi decenni hanno utilizzato il termine e il concetto di informazione (la biologia, la fisica, l'economia, ecc.). Mackay aveva parlato di «quantal aspects of information», altri recentemente hanno parlato di «quantum theory of information»²⁶, con riferimento a una realtà ultima dell'informazione strettamente legata al concetto fisico di quanto di energia.

In realtà, diversi tentativi in questo senso non hanno avuto successo. Il bit appare l'unica unità di misura realmente utilizzata, prima dagli ingegneri e poi da noi tutti, come base di tutto il mondo digitale. Il successo della teoria di Shannon e Weaver era basato anche sulla deliberata rinuncia ad affrontare gli aspetti seman-

²⁵ Mi riferisco ovviamente al lavoro alla base della teoria della comunicazione (Shannon e Weaver 1949). Sulle origini e il successo di quella teoria, anche nello specifico delle scienze della documentazione, cfr. almeno il recente lavoro di Ronald Kline (Kline 2004).

²⁶ Per esempio H. Lire, come indicato nella già citata rassegna di Rafael Capurro e Birger Hjørland (Capurro e Hjørland 2003). Si può inoltre seguire sull'argomento la completa ricostruzione storica di Gernot Versig (Versig 1997).

ELEMENTI ATOMICI E TECNOLOGIE DELL'INFORMAZIONE E DELLA COMUNICAZIONE (Robertson, comunicazione personale)	
Tecnologie dell'informazione	Elementi informativi minimi
<i>Comunicazione (tecnica)</i>	
Linguaggio	Messaggio
Scrittura	Parola
Alphabet	Carattere
Telegraph	Bit
<i>Numero</i>	
Scrittura	Numero
Sistema arabo	Digit
Punched card	Bit
<i>Image</i>	
Fotografia	Immagine fissa
Film	Frame
Televisione	Line
Schermo del computer	Pixel ²⁷ /bit
<i>Comunicazione (organizzativa)</i>	
Posta	Messaggio
Biblioteche	Libro
Editoria	Libro
Editoria	Rivista

ti (e pragmatici, verrebbe da aggiungere) della comunicazione, e sull'esclusivo trattamento degli aspetti tecnici. La ricerca relativa agli aspetti semantici e pragmatici ha imboccato altre strade, e almeno per quanto riguarda le scienze dell'informazione non ha più cercato "unità minime"²⁸.

La riflessione presentata nel 1993 da un informatico come Robertson costituisce in qualche modo un punto fermo e sancisce la rottura tra la ormai acquisita specificità delle scienze dell'informazione (nella loro accezione più lata) e le sirene di una "metascienza" dell'informazione, che avrebbe come base le scienze fisiche²⁹.

²⁷ «Pixel. Un pixel (abbreviazione inglese di picture element) è uno dei molti minuscoli puntini che compongono la rappresentazione di un'immagine nella memoria di un computer. Solitamente i punti sono così piccoli e così numerosi che se vengono stampati su carta o visti su un monitor appaiono fusi in una unica immagine. Colore e intensità di ciascun punto sono usati dal computer per rappresentare una piccola area dell'immagine» (<http://it.wikipedia.org/wiki/Pixel>).

²⁸ Forse l'ultimo tentativo in questo senso è stato quello di Anthony Debons, che ha proposto la definizione di realtà come *informs* e *knowgs* sulla base della possibilità di ciascun paragrafo di poter dare o informazioni (rispondere alle classiche domande: *when, where, who, what*) o conoscenza (*why* e *how*) (Debons 1992). Una proposta che non sembra avere avuto seguito.

ELEMENTI ATOMICI E STRUMENTI SOFTWARE (tabella costruita su dati forniti da Robertson, comunicazione personale, i testi in corsivo sono dello scrivente)	
Elementi minimi dell'informazione trattati da strumenti software o strumenti di information retrieval	
Word processing	L'unità minima è il carattere, a metà tra la parola e il pixel Altre unità minime possono comunque essere la parola, la frase e il paragrafo
Fogli elettronici di calcolo	Il numero e la formula di calcolo
Sistemi di gestione dei database	La tabella, il record, il campo, il sottocampo. Si assume che l'analista abbia definito i campi a un tale livello che il sistema non ha bisogno di accedere a un livello più basso <i>Rientrano in questo ambito i database catalografici, qualunque formato di meta dati essi usino.</i>
Text retrieval	Le parole, le frasi, le sentenze, i passaggi
Natural language processing	La struttura della "frase" e la parola e la frase
<i>Iper testo</i>	<i>Il chunk, che esprime un concetto, e può essere di ampiezza variabile; il link o nodo tra più concetti</i>
<i>Linguaggi di marcatura</i>	<i>Il paragrafo, la frase, la formula, il nome, un elemento, un attributo, un concetto</i>
<i>Video</i>	<i>Il frame, lo shot, la scena, il clip, il programma, l'episodio³⁰</i>

²⁹ Questo senza voler criticare chi nelle scienze fisiche prosegue la ricerca di una realtà informazionale minima, in particolare a partire dalla meccanica quantistica, e vede nell'informazione proprio una prospettiva in grado di unificare le diverse visioni della realtà. Cfr. al proposito il recente, divulgativo volume di Hans Christian Von Baeyer, *Information. The new language of science* (Baeyer 2004), che presenta l'ipotesi dell'esistenza del qbit.

³⁰ «A 'piece' of video is variously referred to as a frame, a shot, a scene, a clip, a program, or an episode. To provide anything more than linear navigation through video information, we need to structure video in some way; as has been discussed, this structuring needs to be done automatically. Video is made up of frames grouped together into *shots*, which are defined as the contiguous set of frames taken by a single uninterrupted camera over time. During a shot, the camera may move by zooming in or out, panning to the left or right, tracking, booming up or down, tilting, or by a combination of any of these motions. Shots are often grouped into logical or semantic units called *scenes*, which will have some interpretation in terms of the overall story being related in the program or episode that constitutes the complete video file. A clip of video is any unit; it may be as large as a set of multiple shots or scenes, or as small as a shot fragment» (Smeaton 2004).

In questo senso, l'intervento di Robertson conferma da un altro punto di vista quanto da tempo sostenuto da Alfredo Serrai e lucidamente confermato in un intervento pubblicato nel 2001:

«è necessario ripetere, con forza, come sia indispensabile in merito liberarsi di due pregiudizi assai comuni, e tuttavia entrambi esiziali: il primo consiste nell'erroneo valore che generalmente si attribuisce alla informazione, che viene considerata alla stregua di elemento o particella di conoscenza, nucleo di sapere, unità di giudizio discriminativo o almeno di perizia orientativa; il secondo nella caratterizzazione dell'era attuale e della nostra società quali era e società della informazione.

Niente di più fallace: l'informazione è unicamente la segnalazione di una differenza, e tale differenza risulta incisiva se, eventualmente, suscita e mette in azione altri effetti. Quel che conta sono le modifiche e gli esiti che l'informazione induce quando viene opportunamente ricevuta da un sistema predisposto.

Occorre sbarazzarsi dell'idea che l'informazione posseda una qualità intrinseca, che possa cioè venir considerata e valutata al di fuori del sistema in cui dovrà o potrà produrre i propri effetti informativi» (Serrai 2001, p. 43).

Non si può negare il fascino di alcune riflessioni interdisciplinari sul concetto di informazione che Serrai giustamente sottopone a critica. Un approccio al tema della granularità dell'informazione penso non possa fare a meno della consapevolezza dei temi affrontati in altre discipline e sulla base di approcci pure non da tutti condivisi. D'altra parte trovo condivisibile e in linea con l'atteggiamento di Serrai il richiamo di Rafael Capurro e Birger Hjørland che, nel ricostruire le origini greche e le radici latine del termine e nel segnalarne i diversi significati assunti negli ultimi decenni, hanno sottolineato la necessità di non focalizzarsi eccessivamente sul concetto di informazione e piuttosto di concentrarsi sui segni, i testi e, aggiungerei, i documenti (Capurro e Hjørland 2003).

L'intervento di Robertson è del 1993, agli albori del fenomeno Internet. Le realtà documentarie e gli strumenti software a disposizione non avevano ancora raggiunto il numero e la raffinatezza che conosciamo oggi e le risorse di Rete non avevano ancora assunto le dimensioni attuali. Gli sviluppi registrati nel settore dell'information retrieval, la creazione dell'ipertesto e del World Wide Web e soprattutto il trattamento sempre più analitico delle diverse entità informative possono a mio parere essere in gran parte inseriti all'interno delle griglie predisposte da Robertson. Nel frattempo, la metafora della granularità dell'informazione ha continuato ad agire e probabilmente ad aprire nuovi orizzonti a coloro che si occupano specificamente di testi, di documenti e di surrogati dei documenti.

Granularità dell'ipertesto e architettura dell'informazione

Seguendo le indicazioni di George Landow, l'ipertesto è

«qualsiasi forma di testualità – parole, immagini, suoni – che si presenti in blocchi o lessie o unità di lettura collegati da link. Si tratta, essenzialmente, di una forma di testo che permette al lettore di abbracciare o di percorrere una grande quantità di informazione in modi scelti dal lettore stesso, e, nel contempo, in modi previsti dall'autore» (Landow 1998).

Questa nuova realtà testuale, realizzata e soprattutto sviluppata grazie all'uso delle nuove tecnologie elettroniche, consente al lettore una libertà molto più grande di quella concessa dal tradizionale documento cartaceo. Attraverso la rete dei legami tra i nodi concettuali dei testi e dei documenti, l'autore dell'ipertesto traccia molteplici percorsi di lettura. Il lettore può approfittare di questi percorsi e talvolta, seguendo le proprie intuizioni e associazioni mentali, crearne di nuovi, in maniera più attiva di quanto non succedesse nella tradizionale lettura lineare e sequenziale.

Nella visione di Landow, l'ipertesto nasce nella seconda metà del secolo scorso dal convergere delle correnti della teoria critica post-strutturalista e dal contemporaneo sviluppo tecnologico. Nel complesso percorso che va da Roland Barthes (in particolare, la sua analisi critica di *Sarrasine* di Balzac e la creazione del termine stesso *lessia*) a Michel Foucault, da Derrida e Deleuze e Guattari, Landow inserisce l'ingegnere Vannevar Bush e il suo futuristico *Memex* e Theodor Nelson, che ha propriamente coniato il termine *ipertesto*.

La frammentazione del testo in lessie e i collegamenti (link) tra queste secondo percorsi di lettura e interpretazione sempre nuovi sono gli elementi centrali di un ipertesto.

«Si potrebbe dire che l'ipertesto è una forma di testo composta da blocchi di 'scrittura' e immagini collegati da link, che permette una lettura multilineare: non una lettura non lineare o non sequenziale, ma una lettura multisequenziale» (Landow 1998).

Landow sostiene che i link avvengono tra «blocchi o lessie o unità di lettura», senza dare ulteriori specifiche sulle dimensioni fisiche di queste unità. Creare ipertesti significa disegnare forme di collegamento unidirezionali o bidirezionali tra lessia e lessia, tra stringa e stringa, tra stringa e lessia. È l'affermazione pieno del desiderio di "mobilizzazione" del testo e delle sue parti componenti, non più prestabilite in forme rigide dall'autore o dall'editore. Nel linguaggio di Landow, il testo appare "frammentato", "disperso" o "atomizzato" (Landow 1997, p. 96-97). In altri interventi relativi alla teoria o alla pratica dell'ipertesto – interventi di autori di formazione più

propriamente scientifica – si usano anche altri termini. Jeff Conklin, per esempio, ancora nel 1987, prima dell'affermazione di massa del Web, affermava che

«Most users of hypertext favor using nodes which express a single concept or idea, and are thus much smaller than traditional files. When nodes are used in this fashion, hypertext introduces an intermediate level of machine support between characters and files, a level which has the vaguely semantic aspect of being oriented to the expression of ideas. But this sizing is completely at the discretion of the hypertext writer, and the process of determining how to *modularize a document* [corsivo mio] into nodes is an art, because its impact on the reader is not well understood» (Conklin 1987, p. 35).

Modularize a document, si esprime Conklin, che prosegue con *modularization of ideas*. Credo si possa essere certi che siamo vicini vicini al concetto espresso con *granularity*, almeno in relazione alla realtà dei documenti ipertestuali e ipermediali.

«Hypertext invites the writer to modularize ideas into units such that (a) an individual idea can be referenced elsewhere (i.e. this is a scope of reference concern) and (b) alternative successors of a unit can be offered to the reader (e.g. more detail, an example, bibliographic references, or the logical successor). But the writer must reckon with the fact that a hypertext node, unlike a textual paragraph, is a strict unit which does not blend seamlessly with its neighbors» (Conklin 1987, p. 35).

Non si tratta di problemi sostanzialmente nuovi. Da secoli il testo è *modularizzato* in paragrafi, così da consentire sia una chiara esposizione di ciascun idea e di ciascun concetto, sia al fine di consentire una più agevole lettura. Il nodo ipertestuale costituisce però una realtà nuova, per la lettura, la ricerca e il recupero.

Geoffrey Nunberg è uno dei primi nel 1993 a utilizzare il termine *modularity* per descrivere alcune caratteristiche tipiche del nuovo documento elettronico, in cui viene messa in crisi la corrispondenza tra unità fisica e contenuto (libro, capitolo, articolo, ecc.).

«And finally, it is the individual user who determines the modularity of the document – that is, the particular chunk of text that it contains [corsivo mio].

All of this affects the way documents are used and reproduced. Let me begin with modularities. A mass-produced book is both bound and bounded in a way that's replicated for all its instances: each copy contains the same text in the same order. But the computational representations of texts can be divided and reassembled in an indefinitely large number of documents, with the final form left to the decision of the individual user» (Nunberg 1993).

La crisi dell'unità cartacea "libro" e la distinzione tra struttura fisica del volume e struttura logica del testo aprono al lettore innumerevoli possibilità di creazione e rielaborazione, a cominciare dallo specifico ritaglio di queste unità minime del testo per passare al numero dei link che collegano queste unità atomiche. Le dimensioni della lessia sono difficili da identificare in termini assoluti, forse è possibile valutare solo quale può essere la dimensione ottimale in grado di facilitare il compito del lettore. Per esempio, attraverso un'analisi del comportamento di alcuni lettori di ipertesti, Erping Zhu ha cercato di valutare se la leggibilità di un documento è influenzata dal numero di link ipertestuali e dalla loro granularità.

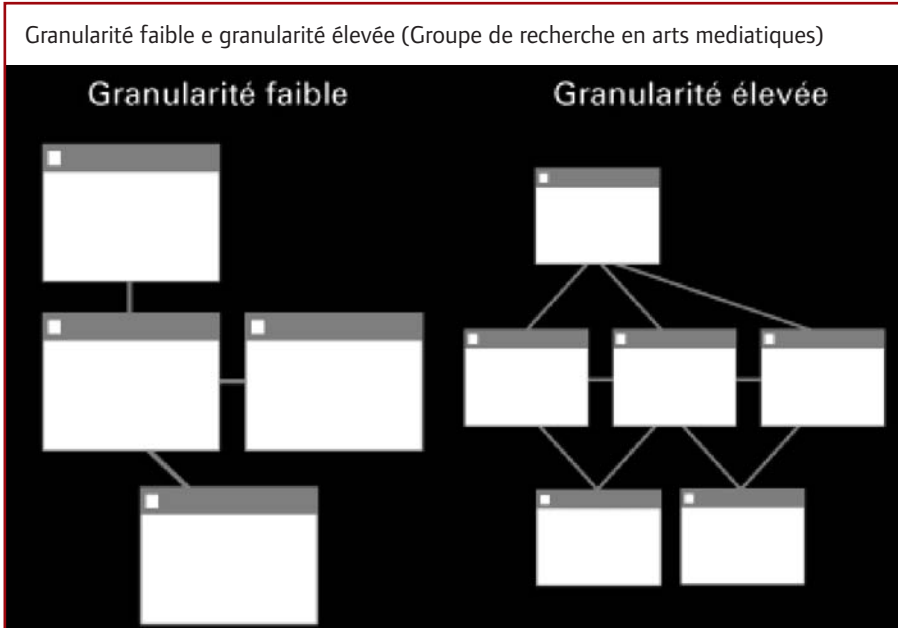
«In hypermedia literature, granularity is defined as the grain size or size of a node. A high or large granularity means a larger node or larger unit of information. A larger node indicates that a node is larger than a few sentences or one screen. A low or small granularity indicates that a node has few sentences and its length is less than one screen, normally explaining a single concepts or idea. This concepts was used in defining the size of a node in this study. Larger nodes in this study consisted of 29-61 lines, roughly 360 to 740 words. Smaller nodes ranged from 5-26 lines, 60 to 312 words» (Zhu 1999, p. 342).

Gli studi di Zhu hanno dato un risultato sostanzialmente aperto, ma impongono la necessità di focalizzare la nostra attenzione sulle capacità e potenzialità cognitive dei lettori. Il problema della *granularità* diventa quello di rafforzare queste potenzialità e non di disorientare chi legge con troppi link. La frammentazione del testo e la moltiplicazione dei percorsi, quindi, non deve eccedere, oppure deve essere sostenuta da molteplici aiuti alla lettura e alla ricerca. In questo senso si esprime un giornalista italiano ben documentato sulla realtà del Web, Franco Carlini, che parla del pericolo di un «sovraccarico cognitivo» e di uno «smarrimento granulare» per un utente alle prese con troppi link, portati a unità o frammenti di testo troppo piccoli (Carlini 1999, p. 56-59).

Anche il *Groupe de recherche en arts médiatiques* dell'University de Quebec de Montreal segue un approccio simile, nella definizione che tenta del concetto di granularità. Nel *Dictionnaire des arts médiatiques*, stilato dal nel 1996, granularità è il

«Degré d'interactivité d'un programme, représenté par la quantité de points décisionnels qu'il comporte. Un grain est en ce sens l'unité fondamentale d'interaction dans un programme interactif»³¹.

³¹ <http://www.comm.uqam.ca/~GRAM/C/term/tcm/tcmt67.html>.



La definizione del *grain* come *unità fondamentale d'interazione* ben evidenzia il ruolo cardine assunto dal lettore. È sulla base delle capacità cognitive di quest'ultimo che deve essere pensata l'organizzazione del sistema interattivo. Cercando di definire le caratteristiche che dovrebbe assumere l'unità minima di un ipertesto, Robert Horn ha stilato una vera e propria 'tabella dimensionale', che va dalla frase, al brano di testo arbitrariamente "ritagliato", al video a cui si affaccia il lettore³².

I problemi connessi alla granularità del link o del nodo di informazione si sono ovviamente ampliati da quando l'ipertesto è diventato strumento di uso quotidiano attraverso Internet. Le dimensioni delle unità informative minime (la pagina, il paragrafo, la videata del computer) sono studiate e valutate nei rapporti con il loro significato semantico, alla luce delle capacità cognitive del lettore e al fine di migliorare l'usabilità del sito web. Se le informazioni sono presentate in maniera corretta, semplice, facile da leggere e memorizzare, l'ipertesto arriverà meglio all'attenzione del lettore. È questa la strada percorsa per esempio da chi si occupa di quella che viene chiamata l'usabilità delle pagine web, cioè del «grado in cui un prodotto può essere usato da particolari utenti per raggiungere certi obiettivi con efficacia, efficienza e soddisfazione in uno specifico contesto d'uso» (UNI EN ISO 9241-11, p. 2).

L'esigenza di ben progettare un sito web così da renderlo maneggevole e funzionale è alla base della nascita di quella che viene talvolta considerata una giovane disciplina, la cosiddetta "architettura dell'informazione".

³² Horn 1989, cit. <http://www.comm.uqam.ca/~GRAM/A/comm/032.html>.

Gli autori di un testo basilare, *Information architecture for the World Wide Web*, Louis Rosenfeld e Peter Morville, definiscono l'architettura dell'informazione come:

«La combinazione di organizzazione, categorizzazione e schemi di navigazione all'interno di un sistema di informazioni.

Il design strutturale di uno spazio informativo atto a facilitare l'esecuzione dei compiti e l'accesso intuitivo ai contenuti.

L'arte e la scienza della strutturazione e classificazione dei siti web e delle intranet per assistere gli utenti a trovare e gestire le informazioni» (Rosenfeld e Morville 2002).

Sono questi "architetti" a usare ripetutamente il concetto di granularità e, in alcuni casi, a tentare la definizione di alcuni parametri. Nel definirne le basi disciplinari, sempre Rosenfeld e Morville indicano come centrali la strutturazione, l'organizzazione e la catalogazione dell'informazione.

«La strutturazione implica la determinazione degli appropriati livelli di granularità degli 'atomi' di informazione del vostro sito e il decidere come correlarli. L'organizzazione riguarda il raggruppamento di quei componenti in categorie significative e distinte. Classificazione significa definire le categorie e la serie di collegamenti ipertestuali che condurranno ad esse. [...] La *granularità* si riferisce alla dimensione relativa o grezza dei singoli elementi informativi di base. I vari livelli di granularità possono includere il numero di un giornale, articoli, paragrafi e frasi» (Rosenfeld e Morville 2002, p. 5).

Non ci sono indicazioni più specifiche o più precise. La granularità rimane come un concetto – o una metafora – estremamente "elastico", che consente di evocare facilmente il problema della dimensione di questi "atomi", e soprattutto della loro gestione, della loro classificazione, dei link da creare, della strutturazione del sito. Evidentemente, più crescono il loro numero e la loro specificità, maggiori saranno l'impegno del sistema dal punto di vista informatico e lo sforzo di raccolta e classificazione attorno al loro nucleo concettuale.

Forme granulari del documento elettronico

La creazione dell'ipertesto ha segnato un momento fondamentale di una più complessiva trasformazione del documento stesso. La sua nuova natura digitale, la possibilità di combinare realtà multimediali diverse, la distribuzione e l'accesso via rete hanno finito per mettere in discussione la natura originaria del documento.

Ancora oggi, il documento cartaceo rimane come modello di riferimento ma, una volta aperto il vaso di Pandora della tradizionale unità documentaria, sia il creatore

che il lettore possono pensare di poter scegliere o creare autonomamente il documento più consono ai propri scopi. Per quanto riguarda l'*information retrieval*, ne sono riuscite esaltate le possibilità dell'indicizzazione automatica e di reperimento di documenti e frammenti di documenti. Inoltre, lo sviluppo del World Wide Web ha portato alla creazione di uno "spazio" informativo assolutamente inimmaginabile solo pochi anni fa. Molti utenti pensano forse oggi alla Rete attraverso il filtro dei motori di ricerca, e sono abituati a raccogliere frammenti di siti, pagine, immagini, la cui realtà ha ormai ben poco a che fare con il tradizionale modello cartaceo. Molto probabilmente la realtà documentaria appare sempre più articolata e sfuggente, come pure sempre più indefinibile e tutto sommato ininfluenza o inutile appare loro la distinzione tra documento e surrogato. Ciò che emerge è uno spazio di memoria informativa di durata indefinibile, in cui ogni documento, ogni testo, ogni immagine, ogni frammento, può essere oggetto di continue ridefinizioni, può mutare di significato, riaggregato in percorsi e contenitori sempre diversi. I motori di ricerca consentono – o obbligano – la ricerca di dati, parole, frasi, pagine, paragrafi, *chunk* o pezzi vari, nel formato più semplice e "nudo", senza alcuna mediazione indicale; le lunghe liste di risultati di ciascuna *query* possono essere ricollocate successivamente all'interno di un nuovo contesto, in un puzzle da riasssemblare e ricomporre vorticosamente.

Non si fatica dunque a comprendere perché la metafora della granularità abbia continuato a operare e il termine a riaffiorare costantemente e sempre più ampiamente. Liberata da ogni vincolo fisico, che le teneva raccolte in un libro o in un altro documento cartaceo, queste ideali realtà informative atomiche che percorrono lo spazio informazionale – un po' onde, un po' granuli – possono essere riaggregate in modi sempre nuovi da parte dei navigatori della rete³³. In questo tipo di realtà, la granularità appare sì una caratteristica fondamentale, ma risulta in qualche modo connotata in modo "orizzontale", quasi piatto. I documenti sono articolati in tante parti, ciascuna di esse facilmente recuperabile per le sue dimensioni, il suo formato, i suoi metadati, il testo o le caratteristiche multimediali. I legami ipertestuali possono legare questi documenti e le loro parti, ma non è obbligatorio, questi ultimi possono anche muoversi liberamente in rete, senza rapporti definiti tra di loro, incrementando lo "smarrimento granulare". Questa dimensione "orizzontale" non è forse sufficiente a connotare come granulare un documento, un surrogato o un'altra realtà informativa. Per approfondire questi aspetti, occorre provare forse a riprendere dalle basi il concetto stesso di documento elettronico.

Descrivere e concettualizzare le modificazioni subite dalle realtà documentarie è uno sforzo che dura oramai da almeno un paio di decenni.

³³ Ammesso che non siano proprio le realtà minime unitarie, i concetti, i *memi*, a costituire l'unica realtà e gli utenti della rete i loro strumenti di navigazione. Ma non vorrei seguire qui ulteriormente l'approccio à *la Matrix* derivato dal *gene egoista* di Dawkins e proposto, per esempio, da Susan Blackmore in *La macchina dei memi. Perché i geni non bastano* (Blackmore 2002).

La rigidità del documento cartaceo, forse più mito che realtà, ha lasciato il campo a realizzazioni molto più flessibili e articolate che consentono modi nuovi di elaborazione, scrittura, stampa, conservazione, pubblicazione, ricerca. La separazione concettuale tra aspetto grafico dei caratteri e della pagina (font, carattere, organizzazione della pagina, ecc.) e il suo contenuto è diventata per taluni divergenza tra il testo in sé e il testo come letto³⁴, ma ha anche consentito di “portare” il testo su diverse piattaforme software, e di aprire nuove possibilità di interrogazione e di analisi a più livelli.

Il documento elettronico appare oggi suddiviso e nello stesso tempo integrato in parti componenti, ciascuna connotata da caratteristiche diverse (testo, suoni, immagini, video, ecc.). A fronte della perdita dei vincoli fisici, e quindi della tradizionale identità o fissità del documento, questa nuova organizzazione composita e i link ipertestuali lo hanno reso aperto all’accesso e all’indicizzazione.

L’unità documentaria, la stessa pagina web, può venire a comporsi di unità più piccole e dinamiche, elementi multimediali spesso linkati da altre realtà e composti in entità informative nuove. Centrale qui non è più il documento, ma quella che viene chiamata *granularity of browsing*, la capacità degli utenti di interagire non solo con la pagina web nella sua interezza, ma anche con unità più piccole, quasi delle ulteriori finestre, al cui interno possono essere richiamate risorse provenienti da altre fonti e dotate di caratteristiche diverse. Il loro accostamento finisce per produrre quella che qualcuno ha chiamato *recombinant information* o *metadocuments*, realtà nuove che sollecitano le capacità cognitive degli utenti. L’esperienza in questione è stata effettuata presso il Department of Computer Science dell’Università del Texas, i cui ricercatori hanno usato espressamente in questo contesto il termine di granularità.

«Granularity refers to the size of the fundamental units that browsers present. combinFormation shifts the granularity of browsing down from documents to smaller units. It makes *information elements* essential. In the current implementation, supported media elements consist of images and chunks of text. [...] users regularly need to deal with these finer grains. combinFormation translates hypermedia documents into *containers* of media elements and references to other containers. Together, the media elements, their containers, and hyperlinks comprise the set of hypermedia *information elements*, on which the program conducts its operations. Documents are automatically downloaded every few seconds; information elements stream into the visual space once per

³⁴ Cfr. in questo senso le riflessioni di Aarseth e il conio delle espressioni “textons” e “scriptons” per sottolineare come il testo elettronico si distingua in realtà differenziate: «A text is any object with the primary function to relay verbal information. Two observations follow from this verbal definition: (1) a text cannot operate independently of a material medium, and this influences its behaviour, and (2) a text is not equal to the information it transmits» (Aarseth 1997, p. 62).

second. Temporally, the granularity of browsing is extended from a single static view that settles and waits for the user to click a hyperlink, to a dynamic one that evolves continuously» (Kerne et al. 2003a)³⁵.

Sempre più questi nuovi documenti sono prodotti dai browser o da veri e propri agenti informatici che navigano in rete alla ricerca dell'informazione, riportando agli esseri umani i risultati di questa ricerca, in forme sempre nuove.

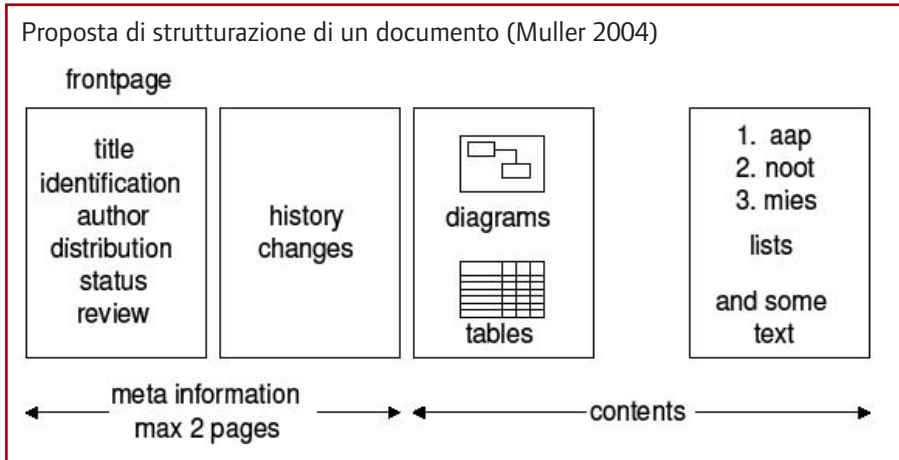


A fronte delle sperimentazioni più ardite, vale comunque la pena di sottolineare come sia riconosciuto da tutti coloro che hanno esperienza di documenti – digitali e cartacei – che le capacità di lettura dell'uomo sono esaltate da una ben articolata organizzazione del documento, preceduta da una ben studiata guida alla lettura. Libri e articoli cartacei si sono storicamente dotati di indici di diverso tipo e sommari, oppure sono stati ripartiti secondo precise strutturazioni. Le tecnologie digitali possono facilitare questo tipo di organizzazione consolidata, che può aiutare la lettura anche quando effettuata sullo schermo di un computer³⁶.

³⁵ Cfr. anche Kerne, Khandelwal e Sundaram 2003.

³⁶ Sull'organizzazione di testi e documenti in grado di facilitare le capacità cognitive dei lettori, cito per esempio il manuale di Information design di Enrica Cavina e Rossella La Piccirella ispirato dalle metodologie di Graziella Tonfoni (Cavina e La Piccirella 2002). Tra i tanti studi sull'argomento, segnalo quello di Heather Silyn-Roberts, che attesta la maggiore leggibilità dei testi tecnici presentati in forma strutturata, con un abstract e una ripartizione in sezioni, ciascuna seguita da una preve presentazione (Silyn-Roberts 2002).

L'organizzazione strutturata del documento cartaceo viene anch'essa associata al termine di granularità. Il titolo di un documento facilmente recuperabile tramite un'interrogazione di Google è *Granularity of documentation*. L'autore, Gerrit Muller, enumera quali dovrebbero essere le caratteristiche principali e l'organizzazione di un buon documento, e presenta graficamente un modello di documento "granulare", organizzato in maniera per l'appunto da facilitarne la lettura e la gestione³⁷.



La strutturazione "granulare" è ancora più essenziale quando l'obiettivo è quello di gestire un intero data base di documenti, di provenienza diversa, ma tutti orientati a un unico scopo. In questi casi, un accordo preventivo e una standardizzazione del documento dovrebbe facilitare la soluzione del problema. Qualcosa del genere è stato tentato da qualche anno a questa parte dall'*International conference on harmonization of technical requirements for registration of pharmaceuticals for human use*, che ha promosso la redazione di apposite linee guide per la registrazione di farmaceutici.

«The 'Granularity Document' of the M4 Organisation defines a series of standards for the physical construction of CTD (*Common technical document*) submissions. [...] The appendix defines the granularity of the documents, their pagination, the use of section numbers, the use of tab dividers (for the paper CTD) and what to include in the Tables of Content for each module (for the paper CTD). The granularity aspects describe what constitutes a document, i.e., divided by tabs and numbered separately in the paper submission and, in

³⁷ Gerrit Muller, *Granularity of documentation, draft*, 3 luglio 2006, <http://www.gaudisite.nl/DocumentationGranularitySlides.pdf>.

an eCTD, provided as separate files. It also defines at what level in the table of contents documents/files should be placed and where multiple documents/files may be provided for a particular section»³⁸.

Nel documento è indicata con precisione la necessità di suddivisione gerarchica delle specifiche sezioni di questo *Common technical document*. La sua granularità è costantemente definita come l'organizzazione in parti strutturate gerarchicamente.

Organizzato e strutturato è anche l'articolo di interesse scientifico che appare sulle principali riviste internazionali. Essenziali per renderne più rapida la lettura e la circolazione tra gli studiosi di tutto il mondo sono il ridotto numero di pagine e la struttura formale. Quella che viene chiamata *topical structure* o "struttura canonica" è il prodotto di una lunga elaborazione storica.

«Il secolo XIX portò il cambiamento che condusse al principio organizzativo, oggi universalmente accettato, noto come *struttura canonica*:

- 1) Intestazione (titolo, lista e credenziali degli autori)
- 2) Riassunto
- 3) Introduzione
- 4) Corpo (metodi, teoria, esperimento, risultati)
- 5) Discussione
- 6) Conclusione
- 7) Ringraziamenti
- 8) Riferimenti bibliografici (bibliografia)» (Matriccioni 2003, p. 8)³⁹.

Al di là delle critiche epistemologiche relative alla scarsa corrispondenza che correla questa organizzazione formale al reale procedimento di ricerca⁴⁰, la struttura canonica ha il pregio di consentire ai lettori una lettura standardizzata e veloce, mirata ai "fatti", ai dati e alle caratteristiche delle esperienze di laboratorio. Veloce risulta anche la valutazione relativa all'affidabilità dei risultati e a un'eventuale possibile riproducibilità dell'esperimento stesso (Harmon 1989).

Recentemente alcuni lavori di un gruppo di ricerca olandese, tra i quali sono da citare quelli di Joost Kircs e Frédérique Harmsze, hanno avanzato la proposta di uti-

³⁸ *International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, ICH harmonised tripartite guideline, Organisation of the Common technical document for the registration of pharmaceuticals for human use, M4*, ultima correzione del 13 gennaio 2004, <http://www.ich.org/cache/compo/1325-272-1.html>.

³⁹ Sulla struttura canonica dell'articolo scientifico, cfr. soprattutto gli interventi di Joseph E. Harmon (Harmon 1989 e 1992).

⁴⁰ Mi limito a rimandare agli studi di Robert K. Merton, e in particolare tra le tante alle sue riflessioni sullo «standard scientific article e le obliterated scientific serendipities» (Merton e Barber 2002, p. 407-430).

lizzare in maniera più efficiente le possibilità ipertestuali, favorendo una maggiore modularità o granularità degli articoli scientifici. La proposta è basata sul riconoscimento che ogni articolo non nasce dal nulla, ma si inserisce all'interno di una tradizione scientifica, riutilizzando e richiamandosi a teorie, modelli, esperienze, dati e strumentazioni preesistenti e già attestati in letteratura. Le ipotesi e i dati che vengono pubblicati in ciascun articolo vanno a integrarsi in un panorama e in un contesto scientifico che in questo modo finisce per consolidarsi sempre di più, almeno fino a quando nuovi dati non falsificano le ipotesi precedenti e si assiste a una rottura del paradigma dominante. La ricerca e la lettura degli interventi più nuovi e originali sarebbero sostanzialmente migliorate qualora si desse la possibilità di recuperare facilmente solo ed esclusivamente quelle parti degli articoli specificamente dedicate ai metodi seguiti, o agli strumenti utilizzati, o alle ipotesi avanzate. In questo modo sarebbero evitate le ripetizioni delle ipotesi di partenza, ed evidenziati solo quei dati e quelle ipotesi che per la loro novità o originalità giustificano la "pubblicazione" in rete al cospetto del "collegio invisibile" dei ricercatori di tutto il mondo. Le potenzialità dell'ipertesto potrebbero in questo modo essere preziose, consentendo di articolare il testo dei documenti scientifici e consentendo una navigazione che faciliti il lettore a diversi livelli (Kircz 1998 e Harmsze 1999).

Sia che si vada verso un documento digitale in grado di sfruttare appieno le potenzialità ipertestuali e ipermediali, sia che ci si richiami alle tradizionali virtù del documento cartaceo strutturato in sezioni, paragrafi, ecc., è ormai ampiamente accettata la visione del documento elettronico come di una realtà in cui l'aspetto grafico e il contenuto sono separati, il testo è strutturato da linguaggi di marcatura e organizzato in parti componenti, correlate tra di loro all'interno di una vera e propria gerarchia. È stato l'influente paper di Steven DeRose, David Durand, Elli Mylonas e Allen Rehear che nel 1990 ha segnato l'abbandono dei precedenti modelli statici e lineari, rappresentati in tabella.

Modelli alternativi del documento secondo DeRose et al. 1990
Testo come bitmap: aspetto grafico
Testo come stream di caratteri: ogni carattere è codificato, l'unico livello di marcatura è costituito dallo spazio bianco e dal <i>return</i>
Testo come carattere e istruzioni per la stampa
Testo come layout di pagina
Testo come corrente di oggetti, in senso lineare
Testo come "ordered hierarchy of content objects" o "OHCO"

Non si può dire che questa concezione del documento abbia risolto tutti i problemi e sia stata accettata senza riserva, e nello stesso *paper* non erano ignorati gli aspetti problematici. Quello che preme sottolineare è che il modello OHCO ha

portato a conseguenze rilevanti per quanto riguarda i temi della granularità. Il documento digitale è dinamico e flessibile, accogliendo altre realtà non testuali (video, immagini, ecc.), e soprattutto non solo statiche, sebbene di non sempre facile gestione in fase per esempio di conservazione e recupero. I linguaggi di marcatura si sono affermati come base indispensabile per la portabilità dei testi su altre piattaforme, per marcare le unità logiche del testo (il paragrafo, per esempio) e per annotarle dal punto di vista semantico. Ogni paragrafo, ogni parola, ogni segno grafico può oggi essere annotato, a livelli di dettaglio estremamente analitici, così da rendere più potente la lettura e la ricerca, e individuare nel testo ulteriori livelli “logici” e concettuali. Queste realtà (testuali e non testuali) che compongono il documento possono essere indagate e indicizzate attraverso l’uso di sistemi di gestione della conoscenza (tesauri e ontologie) che possono sanare la tradizionale distinzione tra indicizzazione dei documenti (tradizionalmente effettuata grazie a classificazione e soggetti) e indicizzazione all’interno dei documenti (come si esprimeva negli indici analitici dei libri)⁴¹.

L’evoluzione dei linguaggi di marcatura assume un ruolo fondamentale nell’ottica dell’affermazione del documento granulare. A partire dalla creazione nel 1974 di SGML – linguaggio nato per consentire la lettura nel tempo dei numerosi testi manualistici prodotti dalle grandi aziende commerciali dell’area della ricerca tecnologica e militare⁴² – il cosiddetto *markup* descrittivo assunse un notevole interesse anche per lo sviluppo del documento di interesse scientifico (Coombs, Renear e DeRose 1987). La complessità di SGML obbligò i creatori del Web alla creazione di un linguaggio di marcatura più semplice, HTML, dall’inaspettato, straordinario successo al momento della diffusione delle pagine web. Proprio la diffusione del Web ha fatto toccare con mano da una parte i possibili vantaggi dei documenti marcati in maniera logica, dall’altra il bassissimo valore semantico di quello specifico linguaggio. Presa coscienza dei limiti, la comunità internazionale ha creato e accettato un linguaggio più maturo, derivato da SGML, ma non così pesante e complesso, designato come XML.

XML è oggi utilizzato in applicazioni estremamente diverse. In un manuale di informatica molto diffuso⁴³, si afferma la natura “ibrida” di XML, a cavallo tra il modello di rappresentazione dei dati e lo standard per la definizione di documenti. Questa doppia natura ha comunque contribuito ad assicurargli il ruolo di anello di congiunzione tra il Web (formato da documenti non strutturati) e le basi di dati (contenenti informazioni strutturate). Come HTML, XML è utilizzato per la rappre-

⁴¹ A questo proposito, cfr. come XML sia oggi presentato anche agli indicatori di professione da un recente articolo di Bill Kasdorf come lo strumento per sanare la storica distinzione tra *open-system indexing* (quella appunto operata su monografie e articoli di rivista attraverso vocabolari controllati), *closed-system indexing* (l’indicizzazione analitica tipica delle monografie e presente al termine del testo) e *word lists* (la ricerca automatica della parola nel testo) (Kasdorf 2004).

⁴² Su SGML cfr. almeno il lavoro introduttivo di Joan M. Smith (Smith 1992).

⁴³ Riprendo qui e nel paragrafo successivo le indicazioni del manuale a più mani sulle basi dati, il cui primo curatore è Paolo Atzeni (Atzeni 2003, p. 222-224).

sentazione di documenti, sia per il web che tradizionali, e per la costruzione di motori di ricerca. L'uso di questo linguaggio consente una migliore e più facile elaborazione di dati presenti in quei contenitori "semi-strutturati" che sarebbero le pagine in formato XML. Una delle caratteristiche essenziali e innovative di XML è poi quella di garantire, molto più di HTML, una maggiore attenzione all'aspetto semantico dell'informazione e, tra le altre cose, una sempre maggiore granularità. Secondo Terrence Brooks,

«The revolutionary aspect of XML is the *modularization* [corsivo mio] of information. Information presents itself as a self-describing unit that can does not inhibit processing, storing or display. Topical subject qualifiers (e.g., attributes) are placed at the appropriate level of granularity. [...] Meaning resides in the semantic structure of information. Meaning can also reside in the meaning-bearing terms and phrases placed at the appropriate level of granularity that serve to qualify a specific element of information» (Brooks 2001).

In questa prospettiva, grazie ai linguaggi di marcatura, l'informazione è sminuzzabile fino a dimensioni realmente minime. A ogni livello, dal più elevato (l'intero documento) al minore (il paragrafo, la frase, la parola, un qualunque "segno") è possibile riconoscere e assegnare un valore semantico, anche senza la necessità di una intermediazione umana, a un dettaglio al quale non era mai stato possibile arrivare. Un esempio offerto dal testo *XML companion* di Neil Bradley⁴⁴ consente di illustrare come si possa marcare un documento XML secondo diversi livelli di granularità. Si può dire che un documento possiede una granularità fine (*fine*) se è possibile isolare e identificare ogni possibile unità di informazione. Per esempio, il nome dell'autore viene suddiviso nel massimo numero di sottounità.

```
<name>
  <salutation>Dr.</salutation>
  <initial>J</initial>
  <last-name>Smith<last-name>
</name>
```

Quando invece un documento ha un numero di elementi inferiore o insufficiente, si dice che possiede una granularità grossolana (*coarse*):

```
<name>Dr. J. Smith</name>
```

⁴⁴ Riporto la definizione *granularity*: «The degree to which an element is divided into child elements. A complex hierarchy denotes a 'fine' granularity. Simple structures with few levels indicate a 'coarse' granularity» (Bradley 2001, <http://proquest.safaribooksonline.com/0201770598/d1e91826>).

Più fine o maggiore è la granularità, più largo e pesante il documento, e più costoso quando è marcato manualmente. D'altra parte, il documento più ricco è anche quello potenzialmente suscettibile di un'analisi e di un'elaborazione più approfondita, con un recupero (un'"estrazione") più avanzato dei dati presenti⁴⁵.

Un testo marcato con XML consente di individuare facilmente le già indicate unità minime del pensiero (paragrafi, concetti, citazioni, formule chimiche, ecc.) e rende direttamente disponibili i "dati" di prima mano della ricerca. XML è oggi uno strumento indispensabile non solo per la gestione di dati in database strutturati, ma anche per quanto riguarda i documenti scientifici, giuridici e letterari. La codifica e marcatura del testo è effettuata anche in realtà complesse come quella biblica, che pure ha richiesto soluzioni articolate. Grazie a schemi di marcatura come il TEI (Text Encoding Initiative, basato originariamente sulla sintassi SGML e poi allineato alla sintassi XML) è possibile rappresentare la struttura astratta di varie tipologie testuali (testo in prosa, testo poetico, testo drammaturgico, fonte manoscritta, etc.), e le caratteristiche testuali rilevanti per diverse aree di ricerca (filologia, analisi linguistica, tematica, narratologica, etc.) (DeRose 1999)⁴⁶.

In ambito scientifico, una marcatura molto attenta e granulare del testo può raggiungere non solo il paragrafo o la citazione, ma anche i dati presenti nei grafici e nelle tabelle, le formule chimiche e quelle matematiche. Negli ultimi anni sono nati "dialetti" XML pensati specificamente per marcare i documenti scientifici, così da consentire un più facile recupero con interrogazione mirate di dati e concetti⁴⁷. Il settore scientifico forse più avanzato è quello della ricerca chimica, in particolare grazie al lavoro di Peter Murray Rust, che ha proposto uno specifico linguaggio (CML). Secondo Peter Murray-Rust e Henry S. Rzepa, grazie all'utilizzazione dei linguaggi di marcatura si potrebbe assistere alla nascita di un nuovo tipo di documento scientifico, il "datumento", realtà strutturalmente e funzionalmente progettata per "salvare" e pubblicare i dati scientifici, prodotti in maniera massiccia dai laboratori di tutto il mondo.

⁴⁵ Usa l'esempio della distinzione del nome e cognome di un autore in campi o sottocampi diversi anche il dizionario online ODLIS (*Online Dictionary for Library and Information Science*, a cura di Joan M. Reitz): «Granularity. The level of descriptive detail in a record created to represent a document or information resource for the purpose of retrieval, for example, whether the record structure in a bibliographic database allows the author's name to be parsed into given name and surname» (http://lu.com/odlis/odlis_g.cfm).

⁴⁶ Non mi soffermo in questa sede sui problemi aperti. Terye Hillesund, per esempio, paventa una vera e propria "single source ideology", e cioè la presunzione che contenuto e presentazione possano essere trattati in maniera completamente indipendente, e che ogni testo possa essere offerto su qualunque supporto e in qualunque occasione (Hillesund 2005). Un'altra critica è quella avanzata, per esempio, da Paul Caton che sottolinea come ogni testo realizzi un originale atto comunicativo, che i linguaggi di marcatura tendono a ignorare, offrendo una visione statica del testo e del documento (Caton 2001).

⁴⁷ Un elenco dei "dialetti" XML di marcatura dei testi scientifici, ciascuno creato secondo le esigenze della comunità scientifica di riferimento, è offerto da Zhongyu Lu 2005.

«Datument: a combination of data and document created using formal markup to allow the processing of individual data components» (Murray Rust 2003).

La marcatura dei testi è comunque operazione estremamente onerosa qualora effettuata da operatori umani⁴⁸. I vantaggi che si possono prevedere da una marcatura dei testi, e pure da una loro organizzazione granulare e funzionale (come ipotizzata da Kircz e Harmsze) necessitano di un investimento massiccio, dal ritorno non garantito, e comunque da spalmare su tempi lunghi se non lunghissimi. Questo è il problema a cui si cerca di ovviare con l'elaborazione automatica del linguaggio naturale (*natural language processing*) e che ha avuto sviluppi di rilievo a livello scientifico e commerciale.

Per tornare al tema della granularità, credo non sia inutile rileggere il primo articolo nel cui titolo appare il termine *granularity*, e cioè il saggio di Frans C. Heeman pubblicato nel 1992, *Granularity in structured documents*. Heeman metteva a confronto tre diversi linguaggi di marcatura dell'epoca (Grif, ODA e SGML) e due diversi approcci alla strutturazione di un documento, l'approccio top-down e l'approccio bottom-up. Tutti e tre i linguaggi ed entrambi gli approcci non riuscivano secondo l'autore a eliminare tutte le *grey area* che venivano a crearsi nel momento in cui la marcatura e la strutturazione logica dovevano confrontarsi con il contenuto stesso del documento. Ogni paragrafo di testo, ogni formula matematica poteva essere marcata in maniera differenziata, secondo visioni diverse ed egualmente ammissibili. In altre parole, il documento – sia dal punto di vista logico, che dal punto di vista del contenuto – offriva per la complessità di alcuni suoi contenuti (tabelle e formule matematiche, per esempio) e per la complessità dei rapporti tra struttura e contenuto, soluzioni molteplici di organizzazione e marcatura.

Rapporti tra struttura e contenuto del documento: il problema della granularità (Heeman 1992)

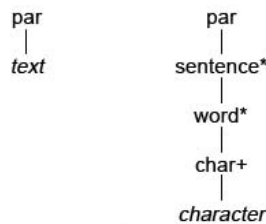


Figure 5. The granularity problem has its effects on content types. At the left, the logical structure stops at the level of a paragraph, and a content type 'text' takes over. At the right, the logical structure continues down to the level of a single character, which is defined as a content type

⁴⁸ L'alternativa è costituita dall'estrazione automatizzata di nomi, formule, ecc., oppure la marcatura in forma automatica. Cfr. su questi temi anche l'intervento di Jon Bosak sui costi della marcatura dei testi scientifici (Bosak 2001): «The principle here is quite simple: the more information we add to a document, the better use we can make of it. But the information has to come from us; it is not going to come from the computer».

Secondo Heeman non può esistere un'unica partizione logica di un documento e il rapporto tra struttura logica e contenuto può essere risolto ragionevolmente secondo modalità diverse, tutte ugualmente compatibili.

«There is no clear distinction between structure and contents. The granularity, the level at which the structure part stops and the content part takes over, differs among systems. On some occasions, the primary structure is (mis)used to express the logical structure of a content type, on other occasions, logical structure is enclosed within contents» (Heeman 1992, p. 153).

Heeman riscontrava inoltre un difficile trattamento di tavole, formule chimiche e matematiche, suoni, ecc, auspicando soluzioni uniformi e integrate per gestire strutture e contenuti così eterogenei rispetto al semplice testo. Le riflessioni di Heeman appaiono interessanti ancora oggi, sia per la messa in guardia verso un eccessivo entusiasmo nell'uso dei linguaggi di marcatura, sia per l'uso del termine granularità, del resto ben poco definito. L'ipotesi che oggi XML lascia intravedere – navigare tra i diversi livelli di dettaglio, quasi zoomando in maniera fotografica – non può che essere considerata un ideale punto di arrivo, considerato il difficile, mai del tutto risolto rapporto tra struttura logica e contenuto del documento stesso.

La manipolazione in formato elettronico del testo e la sua strutturazione logica hanno comunque aperto un'approfondita riflessione sulla possibilità di organizzare in forma strutturata e gerarchica ogni tipo di documento e sugli effetti che tale strutturazione poteva avere ai fini della leggibilità e della ricerca nel testo stesso. L'uso di XML è alla base dell'ambizioso progetto ideato da Tim Berners-Lee per la creazione di un Web semantico in grado di superare i limiti del Web attuale e consentire un recupero dell'informazione maggiormente pertinente agli interessi degli esseri umani attraverso l'uso degli strumenti automatizzati. Riprenderò in seguito questo argomento.

La gestione e il recupero di informazioni da documenti non più costituiti da una sequenza di testo, ma organizzati in una struttura gerarchica in cui sono inserite parti componenti, ciascuna con una propria autonomia, apre inedite possibilità di lettura e di recupero. L'elaborazione automatica di suoni e immagini (fisse e in movimento) è oggi settore di ricerca sempre più sviluppato e che già offre possibilità di ricerca tramite strumenti di *content-based retrieval*, strumenti cioè che recuperano attributi e caratteristiche direttamente dal contenuto dell'immagine o del suono, piuttosto che da parole chiave derivanti da indicizzazione umana (Raieli e Innocenti 2004). L'obiettivo più ambizioso è quello di integrare non solo indicizzazione umana e indicizzazione automatica, ma anche le diverse abilità cognitive degli utenti e le potenzialità delle macchine automatiche. Attualmente, i documenti ipertestuali e ipermediali favoriscono soprattutto l'attività umana del *browsing*, mentre i testi facilitano l'attività del *searching*. Queste due attività cognitive

debbono collaborare in maniera più armonica di quanto non facciano adesso. Utilizzando l'esempio fornito da Yves Chiamarella,

«This integrated view of a hypermedia system allows users to issue queries combining content and structure as parts of the needed information. For example, one can issue a query aimed to retrieve all 'medical reports' that are about 'lung cancer' and which are 'illustrated by images of a given type' (X-ray etc.). In this query 'lung cancer' is the topic assigned to retrieve medical reports (i.e. textual information) while 'X-ray images' refers to a particular type of images. The expression 'illustrated by' refers to links that relate both kinds of data in the hyperbase. As a consequence the answers to such a query will be structured objects extracted from the hyperbase, the structure and the content of which having to match the requirements stated by the query» (Chiamarella 2000).

Modello di iperdocumento (Chiamarella 2000)

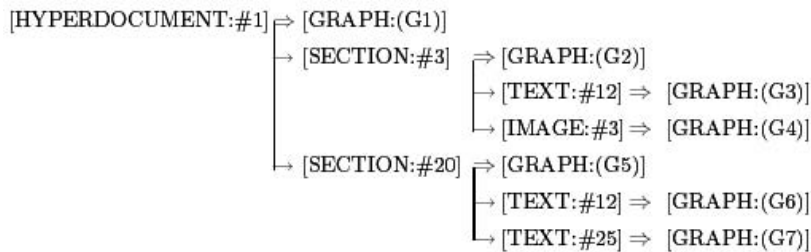


Figure 4. An Example of Hyperdocument.

Il riconoscimento e l'apprezzamento della segmentazione del documento strutturato in piccole unità di base (*granularity choices*, secondo Yves Chiamarella), inserite all'interno di una struttura gerarchica, è ovviamente essenziale per un corretto funzionamento di questo ancora ipotetico strumento di ricerca.

Leggere, sfogliare e cercare un testo sfruttandone appieno le caratteristiche ipertestuali, multimediali di documento strutturato è dunque problema ancora non risolto pienamente, sia per quanto riguarda le capacità degli esseri umani che per quelle delle macchine.

Un altro modo di rafforzare la cognizione umana attraverso lo sfruttamento delle molteplici possibilità di manipolazione offerte dal documento elettronico gerarchizzato, marcato e ipertestuale è offerto da diverse sperimentazioni verso una visualizzazione flessibile. Vale la pena citare come esempio un prototipo che può lavorare a diversi livelli di granularità, nella cui presentazione è esplicitamente usata la metafora dello zoom.

Questo strumento:

«allows readers to view an electronic text in multiple simultaneous views, providing insight at several different levels of granularity, including a reading view. This prospect display is combined with a number of tools for manipulating the text, for example by highlighting sections of interest for a particular task. [...] This project extends research in hypertext and kinetic text theory to provide readers with a text document display that combines simultaneous prospect – an overview of the entire text – and detail views, with related tools» (Ruecker, Homich e Sinclair 2005).

Ma se la visualizzazione di un documento tutto sommato statico può essere favorita da uno strumento di questo genere, la natura composita del documento ipertestuale e ipermediale suggerisce anche la nascita di realtà documentarie nuove e dinamiche. Un documento può essere o apparire granulare non solo perché marcato con uno specifico linguaggio, o perché organizzato in modo tale da consentire una più agevole lettura, ma anche perché può consentire un'agevole riutilizzazione dei suoi contenuti e delle sue parti componenti per creare nuovi documenti. La segmentazione è in questi casi cercata per consentire non tanto e non solo una trasmissione di contenuti, bensì un vero e proprio processo di apprendimento. Nel settore dell'e-learning, in particolare, i termini della famiglia *granularità* sono utilizzati in maniera massiccia.

«Granularité: Niveau de découpage d'un contenu pédagogique en une série d'items élémentaires, appelés grains, que l'on peut re-combiner dans le déroulement de parcours pour répondre aux besoins individuels de formation. [...] Le granule ou le grain est un objet pédagogique. Les objets pédagogiques microscopiques, unités élémentaires d'apprentissage sont associés pour constituer les parcours individuels de formation»⁴⁹.

La partizione granulare di un corso di e-learning è finalizzata all'obiettivo di consentire una formazione del tutto autonoma del discente, che possa ritagliarsi dei moduli formativi su misura. Lester Gilbert, dell'Università di Southampton, richiama comunque una certa dose di cautela. I granuli o i chunk di formazione non devono essere connotati semplicemente da una identica dimensione fisica, bensì debbono correttamente essere definiti all'interno di una vera e propria *learning transaction*, che coinvolge docenti e discenti in un processo comunicativo in più fasi (Gilbert parla delle seguenti fasi: *tell, show, ask, response, feedback*), che non possono essere così nettamente segmentate (Gilbert, Sim, Wang 2005).

⁴⁹ <http://www.educnet.education.fr/dossier/eformation/modularite3.htm>.

Per offrire indicazioni omogenee al fine della creazione di quelli che vengono chiamati *learning object*, è stato elaborato dall'associazione professionale degli ingegneri elettronici americani lo standard IEEE 1484.12.1-2002, *IEEE Standard for learning object metadata*. La definizione di *learning object* rimane abbastanza elastica («any entity, digital or non-digital, that may be used for learning, education, or training»), ma è stata definita con maggiore precisione quale possano i diversi livelli di aggregazione e granularità, anzi di granularità funzionale. Li presento nella tabella successiva, riportando a fianco di ciascun livello l'esempio tratto dallo stesso standard.

Granularità funzionale (IEEE 1484.12.1-2002)	
Aggregation Level	
The functional granularity of this learning object	
1: the smallest level of aggregation, e.g. raw media data or fragments	Una fotografia digitale di Monna Lisa
2: a collection of level 1 learning objects, e.g., a lesson	Una lezione con la fotografia digitale di Monna Lisa
3: a collection of level 2 learning objects, e.g., a course.	Un corso o una collezione di lezioni su Monna Lisa
4: the largest level of granularity, e.g., a set of courses that lead to a Certificate	Un set di corsi comprensivi dell'intera storia, della descrizione, dell'interpretazione su Monna Lisa

Al di là della possibile ulteriore e pragmatica definizione della durata e della composizione di ciascun livello⁵⁰, mi interessa sottolineare come ancora una volta il termine “granularità” venga utilizzato per definire una realtà strutturata e gerarchizzata. Il discente o il docente possono così meglio muoversi tra i diversi livelli di dettaglio, senza perdere i riferimenti al contesto. Si tratta di una visione “verticale” del concetto di granularità, che integra a mio parere in maniera importante il senso “orizzontale” che emerge dalla letteratura sull'ipertesto.

Descrivere e identificare i documenti e le loro parti

L'uso del termine granularità costituisce come detto un interessante segnale delle novità emerse da quando la comunità internazionale dei bibliotecari ha cominciato

⁵⁰ In alcuni casi è stato tentato di definire con ancora maggiore precisione la consistenza di queste unità. Una proposta pragmatica in merito è quella reperibile in rete di Pete Richardson, che parla di 1 ora per il livello 2, 30 ore per il livello 3, un anno di corso per il livello 4. Pete Richardson, *Learning Objects & Granularity - Does Size Matter? (A Pragmatic Approach to LO design)*.

a gestire, descrivere e rendere accessibili le nuove risorse elettroniche. Ho già avuto occasione di riportare come Stefano Gambari e Mauro Guerrini abbiano definito la granularità, ma anche altri bibliotecari hanno utilizzato il termine proprio discutendo i problemi relativi all'identificazione, descrizione e recupero dei documenti elettronici.

Il progetto olandese che nel 1997-1998 prese il nome di DONOR (Directory of Netherlands Online Resources) è stato una delle iniziative che, a livello nazionale, sono state intraprese nel decennio scorso per censire e catalogare le risorse di qualità presenti in Rete e offrire agli utenti percorsi disciplinari guidati (Van der Werf-Davelaar 1999). In un rapporto stilato in quell'occasione, il concetto di granularità è oggetto di un'analisi completa.

«The concept of granularity refers to the fact that in the Web environment, where information entities are hyperlinked, complex hierarchical and/or associative information structures exist. A user may want to identify a network service, a document on that service, a chapter in a document, a collection of images, an individual image, a video or an individual scene or even one frame in that video».

Nell'ambiente di Rete, i confini fisici del documento tradizionale si dileguano e i link, i legami tra i file, o gli oggetti, definiscono realtà documentarie nuove, correlate logicamente ma di difficile gestione da parte dell'utente. Questo ha la possibilità di cercare e trovare le realtà più minute (l'immagine, il video, la scena o il singolo frame di un video) ma corre il rischio di perdere ogni cognizione relativa ai legami – talora molteplici – che legano le diverse entità informative e che danno vita non più ai singoli documenti fisici ma a veri e propri "documenti logici", di lettura e interpretazione non sempre immediata. È quindi necessario identificare e censire non solo le minute realtà e entità, ma soprattutto dare conto dei rapporti che li legano.

«Several 'formal entities' (e.g. files) may be part of the same logical document, e.g. an 'entity of information'. It is possible for a 'formal entity' to be part of more than one 'logical document'. Granularity has to be taken into account when:

you want to assign unique identifiers to information entities

you want to create metadata for information entities

you want to index information and reflect relations between information entities in the presentation of search results.

Retrieval of Web information can be improved when it is made clear of which formal entities a logical document consists. The provider of the information should indicate the relation of an entity with other entities. Indexing services

should be able to link to the logical document when the keyword is found in part of this document in a search» (Peereboom 1998).

Il rapporto elaborato in occasione del progetto DONOR si concentrava sul terzo aspetto, quello della definizione delle relazioni tra le diverse entità documentarie. In via sperimentale, si suggeriva di compilare il campo *Relation* dei metadati Dublin Core dopo avere individuato con precisione le relazioni tra i diversi documenti⁵¹. Si proponeva conseguentemente di sperimentare quelle soluzioni in grado di consentire all'utente, nel momento in cui riceveva una risposta alle proprie query, di apprezzare proprio i rapporti eventualmente esistenti tra i file diversi file reperiti in fase di ricerca.

Le problematiche discusse in occasione del progetto DONOR sono state successivamente riprese e discusse in diverse sedi. Tra queste, una delle più autorevoli è stata quella della *Bicentennial Conference on bibliographic Control in the New Millennium*, tenutasi alla Library of Congress nel novembre 2000. Tra i diversi interventi, quello di Glenn Patton – Manager of Cataloging Products Department presso OCLC – affrontava il tema della granularità in fase di catalogazione. In pratica, all'interno di siti articolati e complessi come spesso si trovano in rete, quale doveva essere il livello di dettaglio a cui doveva scendere un catalogatore? Il titolo dell'intervento di Patton risultava lievemente provocatorio: *Granularity. Are we seeing a return to the 19th-century mixed catalog?* Il riferimento era alla tradizione del XIX secolo quando – prima che fossero disponibili i servizi di indicizzazione degli articoli delle riviste – si procedeva a una indicizzazione degli articoli delle riviste, dei saggi nelle collezioni o nelle collane, dei capitoli nelle opere di grandi dimensioni. Lo spoglio dei periodici e dei seriali costituiva infatti, e costituisce ancora oggi, un paragone preciso a cui richiamarsi quando si discute su quale debba essere oggi l'entità documentaria minima da descrivere.

Nel tentare una prima, concreta risposta alle sollecitazioni degli intervenuti, Patton ricorreva a un'analisi del lavoro dei catalogatori che partecipavano al progetto CORC – progetto lanciato da OCLC per la catalogazione partecipata delle risorse elettroniche remote. In quell'occasione non erano state date indicazioni cogenti ed era stata lasciata una certa libertà ai catalogatori

«At what level of granularity are CORC participants creating records? [...] Do you describe only at the 'site' level or at a level below that – a subsite that forms some kind of logical unit – or at the individual item level, be that a paper or article, an image, or some other kind of resource?» (Patton 2000).

⁵¹ Cfr. la dettagliata formalizzazione dei rapporti tra documenti nel campo Relation dei metadati Dublin Core in Hillmann 2005.

Se il 60% delle risorse descritte appariva simile a un sito nella sua interezza, il 33% rappresentava qualcosa che era una parte di un insieme più grande e il rimanente 7% ricadeva in una sorta di area grigia di difficile definizione. A fronte della sfida quotidiana nel decidere cosa e fino a che punto catalogare, Patton auspicava la redazione di una guida – un codice si direbbe – alle opzioni relative alla scelta della granularità dei documenti da catalogare. Soluzione sicuramente auspicabile, soprattutto per chi deve gestire un archivio di record bibliografici di enormi dimensioni, frutto della catalogazione partecipata di migliaia di catalogatori, ma che non appariva – e non appare neanche oggi – di immediata realizzazione⁵².

Un taglio più teorico ha assunto la discussione tenutasi al *First IFLA Meeting of Experts on an International Cataloguing Code* del 2003 a Francoforte. L'intervento di Gunilla Jonsson affrontava proprio questi temi:

- «1. What is a bibliographic unit?
2. Is there always a one to one relationship between a bibliographic record and a bibliographic unit?
3. Do we need descriptions at a deeper level of granularity for the digital environment?» (Jonsson 2003).

Riprendendo concetti e terminologia proposte dal rapporto FRBR⁵³, il problema di maggior rilievo appare quello della moltiplicazione degli item relativi a ciascuna manifestazione di ciascuna opera, fenomeno a cui si assiste nell'ambiente digitale e che non è affatto semplice tenere sotto controllo. Creare un record bibliografico per ciascun item o parte di item potrebbe essere un'operazione realizzabile in forme automatizzate, magari utilizzando la strutturazione dei documenti elettronici secondo i linguaggi di marcatura. La soluzione non risolve il problema di creare e mantenere un corretto sistema di correlazioni tra i record degli archivi di autorità, quelli bibliografici e quelli relativi ai singoli item.

⁵² Cito dal sito della *Digital library for Earth system education* (<http://www.dlese.org/Metadata/collections/granularity.htm>) alcune sintetiche indicazioni per i catalogatori: «Granularity of metadata records.

Generally create a single metadata record for resources as a whole
 Create additional metadata records for a whole resources if individual parts of a resource differ substantially in technical requirements, descriptions and educational information
 Create relationships between resources using the *fields of relation* and occasionally, *learning resources type* and *description* [...]

Apply the relationship concept above to strike a balance between supporting resource discovery with reasonable user effort and to avoid potential user frustration with retrieving too many redundant records.

Recognize that community developed collections may have varying levels of granularity».

⁵³ Su FRBR, cfr. almeno come riferimento di partenza il testo introduttivo curato da Carlo Ghilli e Mauro Guerrini (Ghilli e Guerrini 2001).

«The bibliographic unit in the digital context is not different from the bibliographic unit in the analog environment. It comes in different constellations and it has different physical properties, and we must build a new infrastructure to be able to obtain items from the digital stacks, the archive. There may be possibilities to produce bibliographic records for a deeper level of granularity than what we achieve for publications on paper, due to economical constraints. It will, however, require that text be structured according to specified standards to allow automatic extraction, or generation, of metadata» (Jonsson 2003).

Questo tipo di riflessione rapporta in maniera diretta la discussione sul concetto di granularità ai temi sui quali negli ultimi anni la comunità internazionale dei bibliotecari si è a lungo confrontata. Mi riferisco non solo al problema della gestione e delle caratteristiche delle risorse elettroniche, ma in senso più ampio a quello della configurazione di tutto l'universo delle relazioni bibliografiche. Il tema è stato affrontato in seguito all'ideazione e alla pubblicazione del modello concettuale conosciuto come FRBR, ma aveva già trovato importanti anticipazioni nelle riflessioni di Alfredo Serrai sull'ontologia catalografica (Serrai 2001a) e nel lavoro con cui Barbara Tillett aveva compiutamente identificato le relazioni bibliografiche (Tillett 1991). Tra queste ultime, si riconnette a mio parere direttamente alla discussione sulla granularità la relazione tutto-parte che sussiste tra

«una parte componente di un oggetto bibliografico o un'opera e la sua totalità, come tra un racconto breve e l'antologia nel quale è contenuto. Le componenti possono essere parti di alcune particolari manifestazioni di un'opera, cioè parti di un oggetto bibliografico, o possono essere parti di un'opera astratta»⁵⁴.

Il punto di arrivo potrebbe essere quello dell'identificazione e della descrizione di tutte le parti componenti che è possibile isolare in un documento, registrando con opportuni legami relazionali i rapporti che esistono tra di loro e tra i documenti e le collezioni dei documenti. Questo consentirebbe all'utente un accesso più facile e "granulare" tramite OPAC rinnovati, come ipotizzato da John Byrum nell'intervento citato nell'introduzione⁵⁵. Credo che secondo questo approccio, il tema della granularità – fino a questo punto della ricerca affrontato in termini "orizzontali" e "verticali" – verrebbe ad assumere una dimensione più compiutamente relazionale. La disponibilità di risorse e strumenti elettronici dischiude la possibilità di accessi più granulari non solo rispetto ai tradizionali documenti cartacei o ai documenti testuali elettronici. Banco di prova e oggetto di riflessioni di particolare interesse è

⁵⁴ Riprendo la traduzione del passo di Barbara Tillett dal recente lavoro di Carlo Bianchini sulle funzioni, gli oggetti e i modelli della catalogazione per autore e titolo (Bianchini 2005).

⁵⁵ Sebbene l'intervento di Byrum faccia riferimento anche all'uso di strutture tesauroali; su questo, cfr. la definizione di Roy Prytherch citata in seguito.

oggi costituito dai documenti digitali multimediali, soprattutto gli audiovisivi. In un recente intervento James Turner e Abby Goodrun sottolineano opportunamente come da tempo i video siano oggetto di una indicizzazione che si spinge al livello della singola scena, se non addirittura al livello del singolo frame. Ciò avviene negli archivi televisivi, per esigenze commerciali e talvolta anche legali (si pensi al caso eclatante delle sequenze relative all'omicidio del presidente degli Stati Uniti John F. Kennedy, analizzate per anni frame per frame). L'esigenza di offrire agli utenti una sempre maggiore granularità in fase di identificazione, descrizione e recupero delle parti componenti dei documenti visivi appare oggi sempre più pressante. Questo richiede uno sforzo preliminare di analisi delle immagini in movimento prodotte nelle diverse fasi di registrazione, preparazione, utilizzazione e riutilizzazione dei documenti visivi, e la preventiva creazione di una sorta di tassonomia dei diversi tipi di materiali che possono trovarsi in una collezione di video (Turner 2002).

La disponibilità dei documenti digitali in ambiente di Rete e di strumenti automatici per il loro reperimento ha suggerito nuove soluzioni anche per quanto riguarda i formati dei record bibliografici. I sostenitori dei nuovi formati di metadati si sono ritrovati ad affrontare diverse problematiche – magari già affrontate e risolte in ambito MARC – relative alla creazione e alla descrizione delle risorse digitali. Soprattutto quando si procede alla progettazione di nuove banche dati o biblioteche digitali, è necessario riflettere approfonditamente sulle specifiche caratteristiche dei metadati associati alla risorsa, sull'architettura dei record, dei campi e dei sottocampi. Interpreto in questo senso i ripetuti interventi di Roy Tennant sul termine "granularità". In un articolo del 2002 sulla sua rubrica periodica che appare *Library journal*, Tennant ha indicato la granularità come caratteristica fondamentale di qualunque tipologia di metadati, anche di quelli creati in fase di marcatura dei testi con il linguaggio TEI (*Text encoding initiative*). Tennant sottolineava:

«Our libraries are increasingly dependent on metadata [...] Granularity refers to how finely you chop your metadata. For example, in the standard for encoding the full text of books using the Text Encoding Initiative (TEI) schema, a book author may be recorded as: <docAuthor> William Shakespeare(</docAuthor>. That's all well and good, if you never need to know which string of text comprises the author's last name and which the first. If you do and most library catalogs should have this capability, you're not going to get very far with information extracted from a book encoded using the TEI tag set».

Chi ha pratica del formato Marc sa bene dell'estremo dettaglio, e quindi dell'estrema "granularità" raggiunta dai record bibliografici creati secondo quello standard. Il richiamo di Tennant – semplice, ma efficace – richiamava comunque gli utilizzatori di questi nuovi standard a progettare con attenzione i nuovi archivi digitali e i

loro metadati. Un campo autore, insomma, deve prevedere un dettaglio adeguato e la possibilità di creare sottocampi che consentano, per esempio, di ordinare i surrogati e i documenti secondo il cognome dell'autore. La soluzione è già stata da tempo trovata nei cataloghi di biblioteca, ma in ambiente di rete i diversi formati di metadati sono creati e implementati in realtà diverse. Se coloro che gestiscono gli archivi digitali non descrivono risorse in maniera dettagliata e per l'appunto granulare, i risultati finali saranno di cattiva qualità. Ne risulterebbero danneggiati gli utenti, sarebbe impossibile un dialogo gestito in forme automatiche tra le diverse biblioteche digitali e sarebbe eventualmente necessario procedere a difficili correzioni di basi dati di notevoli dimensioni.

«Granularity is good. It makes it possible to distinguish one bit of metadata from another and can lead to all kinds of additional user services. [...] Remember: select (or create) and use metadata containers that are granular enough for any purpose to which you can imagine putting them. If you do this, not only can you serve your own purposes, but you can also share your metadata with anyone you wish. If this is not practical, then you must decide which needs will remain unfulfilled. Highly granular metadata doesn't come cheap. There is a trade-off between all possible uses that you may wish to support and the staff time required to capture the metadata required to do so. In some cases, the benefit will not warrant the cost; in others, it will be worth it» (Tennant 2002).

Uno schema di metadati deve quindi essere sufficientemente granulare per rispondere alle esigenze ipotizzabili nel prossimo futuro. I costi necessari per raggiungere questo livello di dettaglio devono d'altra parte essere valutati con attenzione, per evitare uno sforzo troppo oneroso in fase di catalogazione – o di predisposizione della banca dati – in vista di benefici solo ipotetici da parte degli utenti. Verrebbe da commentare che, per esempio, i record in formato MARC sono connotati da una granularità – almeno nell'organizzazione dei campi e dei sottocampi – che certamente non è propria del formato Dublin Core. D'altra parte, anche il formato MARC non è in grado di trattare in maniera soddisfacente né le relazioni tra documenti in ambiente di rete né le relazioni tra le diverse entità dell'universo bibliografico. Solo da alcuni a questa parte, come già detto a proposito di FRBR, partendo dall'insoddisfacente trattamento riservato a queste relazioni dai tradizionali formati delle descrizioni bibliografiche, sono state ipotizzate delle notizie bibliografiche strutturate in modo da tenere conto delle relazioni tra diverse entità documentarie (l'opera, l'espressione, la manifestazione e l'item).

Una visione che mirava a identificare le specifiche caratteristiche "granulari" dei metadati e a integrarle in un quadro complessivo, è stata offerta qualche anno fa in un intervento di Carl Lagoze, che presentava la complessità dei processi di ricer-

ca e scoperta delle risorse informative tra le diverse prospettive e dimensioni della “granularità”.

«Another way to view the process is from the perspective of granularity. Over the term of the discovery process, information seekers move along an information granularity spectrum that impacts their location and examination criteria. Generally, early in the process of information seekers have relatively coarse granularity criteria. They may compose and submit relatively vague queries that, by their nature, lead to large results. In many cases, the examination of the result set may lead to successive refinement of the discovery criteria to finer granularity, as the seekers seeks to decrease the size of the result set. Again, however, the process is not linear and over the long-term the information seeker may move back and forth over the granularity spectrum» (Lagoze 1997).

L'attività di ricerca dell'informazione poteva essere favorita secondo Lagoze dai diversi tipi di metadati allora disponibili, che si componevano secondo lui all'interno del *Warwick Framework*, un meccanismo per associare molteplici set di metadati.

«For example, using the Warwick Framework, we can associate content objects with general descriptive forms such as the Dublin Core and MARC, and domain-specific descriptive forms such as that encoded in the Content Standard for Digital Geospatial Metadata (CSDGM). Each descriptive form can provide data appropriate for a relatively specific niche in the resource discovery process. The Dublin Core is appropriate for coarse-granularity, domain-independent resource discovery. The MARC record is more appropriate for the finer granularity stages of resource discovery. Finally, the CSDGM package enables fine-granularity, domain-specific resource discovery» (Lagoze 1997).

Secondo Lagoze, dunque, la maggiore o minore granularità dei processi di ricerca dell'utente, può essere relativamente assecondata da ciascuna tipologia di metadati. Il rilievo dell'intervento sta nell'impostazione complessiva, attenta sia alle caratteristiche intrinseche dei metadati che alla necessità di venire incontro e accogliere i diversi comportamenti di ricerca degli utenti.

Al di là delle relazioni tra documenti e della gestione dei surrogati descrittivi, un'altra tematica che ha assunto grande importanza nell'età di Internet è quello dell'identificazione delle entità informative, anche e soprattutto ai fini di una corretta attribuzione dei diritti economici. In un catalogo di biblioteca, la descrizione bibliografica è strumento sufficiente per un'identificazione del documento cartaceo. Fino a circa un decennio fa, l'identificazione univoca del libro, per esempio, era garantita a fini commerciali dal numero ISBD, che si integrava nel record bibliografico stilato secondo gli standard bibliografici internazionali. Questa soluzio-

ne non risolve ovviamente i problemi dell'identificazione delle copie, delle ristampe, delle edizioni. Problemi ampliati a dismisura con la moltiplicazione dei supporti e l'avvento delle realtà digitali in ambiente di rete. Identificare e localizzare un documento, gestirne le diverse versioni, consentire l'identificazione delle opere dell'ingegno a fini di tutela dei diritti intellettuali ed economici, autorizzare l'accesso a utenti con caratteristiche e diritti differenziati; si tratta di sfide che richiedono un impegno sempre maggiore.

Crescente appare poi l'esigenza di identificare le parti componenti delle diverse entità informative, come gli articoli dei periodici e soprattutto le diversificate realtà multimediali (testi, immagini, suoni) che compongono i nuovi documenti elettronici diffusi in rete. I collegamenti telematici lasciano intravedere scenari nuovi per la diffusione delle opere dell'intelletto. È stato più volte scritto e affermato che in rete i documenti non possono che essere granulari, così da consentire transazioni intellettuali e commerciali estremamente fini. In pratica, ogni singolo ricercatore potrebbe accedere a un solo articolo senza acquistare l'intera annata o anche l'issue della rivista scientifica. Questo commercio selettivo o granulare potrebbe tutelare la paternità intellettuale e i relativi diritti economici addirittura di una singola frase o immagine, controllandone efficacemente la riproduzione o distribuzione nei più diversi contesti.

Riflessioni attente in questo senso sono state effettuate in occasione della creazione di un nuovo sistema di identificazione dei documenti digitali, il DOI⁵⁶. In diversi articoli e interventi, e nello stesso *DOI Handbook*, Norman Paskin, presidente dell'International DOI Foundation, ha avuto modo di tornare più volte sul tema della granularità, spiegandone il significato in questo contesto e avanzando il concetto di "granularità funzionale".

«Granularity denotes the degree of resolution or granule size of the information items identified. For example, an identifier could be applied to a complete article; for some purposes it is useful to identify individual figures or tables within that article, a finer level of granularity. Similar issues arise in the context of: identification of components and collections; a series or group; an item and parts of an item; a high-resolution, uncompressed image and a thumbnail image; a database; a record in the database; and an image of an item described by a record» (Paskin 1999).

Da una parte, come detto, un documento può essere costituito da parti diverse con caratteristiche differenziate (in questo caso, lascio i termini nella loro vaghez-

⁵⁶ Per una breve presentazione del DOI, all'interno di un più generale inquadramento dei sistemi di identificazione, rinvio agli interventi di Juha Hakala (Hakala 2002) e Giuseppe Vitiello (Vitiello 2004); il *Doi Handbook* è disponibile su <http://www.doi.org/hb.html>.

za), sulle quali potrebbero essere esercitati diritti diversi nel tempo e nello spazio. Identificare preliminarmente ogni possibile elemento “atomico” è però irrealizzabile. Qualunque parte di un documento potrebbe essere oggetto di interesse. La soluzione attualmente ipotizzata è quella di pensare il DOI come un sistema di identificazione aperto, in grado di accogliere gli identificativi standard attualmente esistenti (ISBD, ISSN, SICL, ecc.), consentendo eventualmente integrazioni e specificazioni, sulla base delle esigenze che via via si manifesteranno concretamente (Paskin 1999)⁵⁷. In questo senso, il principio della “granularità funzionale” non assume caratteristiche rigidamente normative, ma rimane aperto: «it should be possible to identify an entity whenever it needs to be distinguished». È l’ambiente digitale a consentire, secondo Paskin, questa elasticità.

«In the digital environment, things can be relatively easily managed at extreme levels of granularity as minute as a single bit. Each of these processes will apply identifiers of different types at different levels of (functional) granularity in different ‘dimensions’, these may need to be reconciled to one another at a point of higher granularity» (Paskin 1999).

Le entità di cui parla Paskin non sono solo quelle delle parti componenti di un documento, ma le collezioni di documenti, le diverse opere presenti in un unico documento e tutti quei documenti che hanno strette relazioni tra loro. Ciascun documento derivato o correlato e ciascuna sua parte potrebbero all’occorrenza essere dotati con un DOI per la sua identificazione in rete, magari utilizzando la partizione creata con i linguaggi di marcatura (Paskin 2003).

La locuzione “granularità funzionale” è stata utilizzata anche da altri studiosi proprio a proposito del tema dei rapporti tra opere, documenti e collezioni e dei modi con i quali questi rapporti possono essere resi evidenti agli utenti. Di granularità funzionale parla Patrick Le Boeuf, alla ricerca di una (non facile) integrazione tra le esigenze che hanno portato alla creazione del DOI e quelle sottese al modello concettuale FRBR. L’obiettivo è quello della redazione di uno standard internazionale relativo a un possibile identificativo di un’opera, o meglio, con un termine FRBR e con ogni possibile cautela, dell’espressione di un’opera. Lo standard ISTC (International Standard Text Code) potrebbe identificare le parti componenti di un volume, le diverse opere all’interno di un volume in commercio, le diverse traduzioni di un’opera. Questo sforzo potrebbe avere interessanti ricadute per i bibliotecari (sia in fase di catalogazione sia al momento del recupero negli OPAC delle espressioni e delle manifestazioni di un’opera) e ovviamente per gli editori, a fini commerciali e a fini di tutela dei diritti economici di un’opera.

⁵⁷ Ciò non implica secondo Paskin nessuna corrispondenza predeterminata tra questi granuli atomici e brani di testo marcati per esempio in formato SGML o XML, a meno che non si reputi che questi stessi frammenti possano essere commercializzati separatamente.

Granularità funzionale nell'interpretazione di Le Boeuf (Le Boeuf 2005)

"Functional Granularity"



(print, online CD-ROM, etc.)

ISSN 1234-567X



Publisher A



ISSN 5674-7727

(online)



ISSN 7654-3216

(CD-ROM)



ISSN 3265-442X

(print)



Publisher B

Diverse proposte di applicazione della "granularità funzionale" per l'assegnazione del numero standard ISSN

La proposta di Le Boeuf è basata proprio sul concetto di *functional granularity*:

«Functional granularity means:

If I need it, I can do it

If I don't need it, I don't have to do it

As a consequence, ISTC applies to various things, meets various needs» (Le Boeuf 2005).

La flessibilità di questo concetto ha il suo fascino. Secondo questo principio, sarebbe dunque possibile creare un identificativo (DOI o ISTC) qualora fosse ritenuto necessario. Non sorprende però che le soluzioni proposte a fini commerciali non siano ancora del tutto conciliabili con quelle avanzate dai bibliotecari. La gestione dei numeri ISTC è tutt'altro che risolta e la prospettiva di creare un DOI per qualunque entità si ritenga opportuno non risolve – secondo Kristin Antelman – il problema di mantenere traccia delle relazioni che legano le diverse entità che vanno a comporre un documento (Antelman 2005). Problemi simili sono stati presi in considerazione in fase di revisione dello standard ISO 3297, relativo all'identificazione dei seriali. La moltiplicazione dei supporti ha infatti creato notevoli problemi

nella gestione dei legami che rapportano tra di loro le riviste pubblicate su supporti diversi e tra ciascun seriale e gli articoli pubblicati al suo interno⁵⁸.

La locuzione *functional granularity* viene ampiamente utilizzata dunque da diversi autori e in diversi contesti, con significati parzialmente diversi. L'individuazione dello specifico livello di dettaglio sembra comunque sempre accompagnarsi a una (non facile) discussione dei legami, tra le diverse entità bibliografiche in gioco.

Un'altra importante traccia di lavoro è offerta al proposito da alcuni interventi e alcune esperienze che si sono registrate in particolare in Scozia. George MacGregor, ricercatore presso il Centre for digital library research, dell'Università di Strathclyde, Glasgow, ha fissato nel 2003 alcune di quelle riflessioni. MacGregor evidenziava innanzitutto alcuni presupposti – se non pregiudizi – dell'attività dei bibliotecari e degli studiosi di *information science*, e cioè quella che lui valuta un'eccessiva enfasi sulle attività (la catalogazione, per esempio) relate al livello dell'item, al contrario di quelle portate sul livello delle collezioni.

A fronte delle gigantesche dimensioni assunte dalle biblioteche digitali negli ultimi anni, è inevitabile secondo MacGregor una diminuzione dell'importanza della descrizione dei singoli documenti e la crescita di valore per l'utente delle informazioni relative al livello della collezione (CLD, collection-level description). L'attenzione si sposta dunque alla definizione di collezione.

«We can consider a CLD to be a structured, open, standardised and machine-readable form of metadata providing a high-level description of an aggregation of individual items. Such descriptions disclose information about their existence characteristics and availability, and employ the use of implicit item-level metadata and, more particularly, contextualise that aggregation of item-level descriptions» (MacGregor 2003).

Nell'ambiente di rete, è probabile che sia più urgente dare e ricevere informazioni sulle collezioni digitali di documenti, piuttosto che sulla singola entità. In realtà, la dimensione e il livello della collezione sono volutamente lasciate imprecise, come indefinite rimangono le parti componenti.

«An institution that agrees upon the particular aggregations that form its collections will invariably discover that these collections are related on a variety of levels. Thus, relationships could be applied to collections of varying sizes and granularity so that, for instance, a 'collection' may contain numerous 'sub-collections', and vice versa ('super-collections')» (MacGregor 2003).

⁵⁸ Cfr. ISO/TC46/SC9/WG 5 N 16, *Revision of the ISSN Standard (ISO 3297), Report from WG5 Adhoc group 1 as follow up to the May 2004 meeting in Amsterdam*, 14 ottobre 2004, <http://www.collectionscanada.ca/iso/tc46sc9/wg5/wg5n16.pdf>; cfr. l'immagine relativa.

Solo l'approccio della granularità funzionale (a "functional granularity approach", nella definizione che MacGregor recupera da Heaney) potrebbe consentire alle istituzioni di adottare in maniera flessibile un diverso livello di dettaglio e gestire così al meglio le parti componenti. La determinazione e descrizione delle collezioni e delle relative sottocollezioni sarebbero lasciate alle conoscenze dei bibliotecari e dei documentalisti e al loro rapporto con gli utenti. In particolare, il concetto di collezione elaborato nell'occasione dell'esperienza scozzese di SCONE richiama direttamente esperienze simili negli ambiti delle collezioni museali e archivistiche, proponendo riflessioni e soluzioni interessanti per ciascuno di questi ambiti, per tanti versi ancora oggi distanti⁵⁹. Il valore di questa esperienza in termini di flessibilità rispetto alle esigenze "locali" degli utenti e alle conoscenze "locali" dei documentalisti trova probabilmente un limite nella difficoltà di rendere comunque disponibili i metadati di queste collezioni e sotto-collezioni locali agli utenti "globali". Se la definizione della collezione avviene a livello "locale", è evidente che la possibilità di riutilizzare i metadati relativi in ambiente di rete globale potrebbe risultarne limitata.

L'attenzione scozzese al livello della collezione è parallela all'attenzione americana al livello dell'item. La crescita delle biblioteche digitali ha trovato un momento di ulteriore sviluppo con il successo del protocollo *Open Archives Initiative*. I metadati relativi ai documenti di interesse scientifico stoccati nei *repository* che rispettano questo protocollo sono raccolti in fase di *harvesting* e le banche dati così create sono messe a disposizione degli utenti. Il limite riscontrato, per esempio, da Muriel Foulonneau, Timothy Cole, Thomas Habing e Sarah Shreeves è che una messe di metadati a livello di documento (il singolo item) ha il difetto di cancellare i legami con la collezione e il contesto specifico in cui gli stessi documenti sono stati creati. L'esempio scelto è quello di una raccolta di documenti su Theodore Roosevelt. Il soggetto assegnato a uno specifico documento di questa collezione digitale può essere, per esempio, «On a horse», che in quello specifico contesto chiunque non può che leggere «Theodore Roosevelt on a horse». Chiaramente, quando i metadati relativi a questo specifico documento finiscono in una base dati di dimensioni mondiali, si può perdere il contesto originale e le relazioni che lo legavano agli altri documenti della collezione.

⁵⁹ «SCONE is based on An Analytical model of collections and their catalogues, a study carried out by Michael Heaney on behalf of the UK Office for Library and Information Networking with support from OCLC. Heaney says: The preliminary work done by UKOLN with respect to collection description identified 'collection' as encompassing the following types of entity: Internet catalogues (e.g. Yahoo); Subject gateways (e.g. SOSIG, OMNI, ADAM, EEVL, etc.); Library, museum and archival catalogues; Web indexes (e.g. Alta Vista); Collections of text, images, sounds, datasets, software, other material or combinations of these (this includes databases, CD-ROMs and collections of Web resources); Collections of events (e.g. the Follett Lecture Series); Library and museum collections; Archives; Other collections of physical items; Digital archives» (<http://scone.strath.ac.uk/service/Guide/gdefinition.cfm>). Cfr. inoltre Powell 2000.

L'utente finale, in fase di interrogazione e lettura dei risultati della ricerca, non ha elementi sufficientemente chiari per comprendere il significato e il valore di quel documento. Per evitare questo effetto – che è poi il tipico effetto prodotto da qualunque motore di ricerca – occorre valorizzare in qualche modo le notizie relative al livello della collezione, che pure esistono, o dovrebbero esistere, in ogni record relativo a ciascun item. In senso inverso a quanto proposto da George MacGregor, occorrerà recuperare queste informazioni, proponendo soluzioni particolari e innovative a livello di interfaccia utente.

«Collection-level description can be a useful adjunct when searching item-level metadata aggregations. Collection-level description can be used as a way to preserve or restore context otherwise lost when item-level metadata are harvested from disparate and heterogeneous repositories and can also provide an additional level of descriptive granularity that may be better suited for some user queries» (Foulonneau 2005).

Questa rapida presentazione di quelle riflessioni biblioteconomiche che hanno utilizzato il termine "granularità" ha consentito di evidenziare alcuni temi rilevanti. Esigenze di identificazione a fini commerciali (DOI, per esempio) e a fini intellettuali (FRBR e ISTC, per esempio) non sembrano ancora comporsi in un quadro preciso, bensì si intrecciano e si scontrano, non riuscendo talvolta a pervenire neanche a linguaggi comuni, mentre altre volte gli stessi termini (*functional granularity*, per esempio) sono usati per sostenere tesi divergenti. Vale comunque la pena di sottolineare come la discussione sulla granularità abbia finito per investire alcuni concetti fondamentali delle scienze del libro: quello del documento, della sua unità e dei suoi confini; quello della collezione, della sua articolazione e del rapporto con il singolo item; e quello del surrogato bibliografico. Il problema più rilevante appare oggi non tanto quello di descrivere, bensì quello di identificare le entità bibliografiche e di rendere chiare e trasparenti – "tracciare", come dicono in gergo gli informatici – le relazioni che esistono tra queste entità e le loro diverse parti componenti. Questo consentirebbe agli utenti non solo un accesso immediato alla singola parte o alla singola entità, ma soprattutto la possibilità di non perdere quei legami che soli possono rendere conto del contesto effettivo in cui si pone un'entità bibliografica o informativa.

Un tentativo di sintesi delle discussioni in corso sulla granularità è stato compiuto nel 2000 da Ray Prytherch in vista della compilazione della nona edizione del prestigioso glossario di biblioteconomia curato originariamente da Leonard Montague Harrod⁶⁰. Credo che quella definizione costituisca a tutt'oggi il più riuscito tentati-

⁶⁰ Ringrazio il prof. Prytherch per la cortesia con la quale, contattato per posta elettronica, mi ha confermato come l'inclusione del termine sia avvenuta nella 9ª edizione (e mantenuta nella 10ª) soprattutto sulla base delle riflessioni svolte all'interno dell'UK Office for Library Networking.

vo di arrivare a una sintesi delle riflessioni sul tema nell'ambito delle discipline della documentazione:

«Granularity. The level of detail expressed in a search term and in an information resource that influences the relevance of retrieved data to the end user» (Harrod 2000).

In fase di indicizzazione il dettaglio con cui è descritto un documento deve potere consentire all'utente un ampio margine di manovra, e la possibilità di avere risultati molto rilevanti per le proprie richieste. Per evitare che il livello di dettaglio sia troppo ampio o troppo ristretto, e ampliare o restringere i risultati della ricerca sulla base della necessità, occorre che lo stesso utente possa di muoversi da una parte all'altra dello spettro della granularità, dal dettaglio più minuto a una visione più ampia e complessiva.

«The end user can move along an information granularity spectrum, beginning with coarse granularity criteria (broad terms) that lead to large results, subsequently narrowing these to a finer granularity that reduces the size of the result set» (Harrod 2000).

Questa definizione sintetizza correttamente gran parte dei significati del termine che abbiamo rintracciato nei diversi contesti. Il focus di questa definizione è correttamente identificato nelle esigenze di ricerca dell'utente, che vanno agevolate dalla strutturazione degli archivi, dei database, dei documenti o dei loro surrogati. Essenziale è dunque il dettaglio con cui viene descritto un documento, ma altrettanto necessaria appare la rete dei legami tra le diverse parte componenti e le diverse versioni dei documenti. Strumenti in grado di consentire una navigazione di questo genere sono i sistemi classificatori e i tesauri, utilizzati da diversi decenni per l'organizzazione dei documenti e dei surrogati bibliografici, sia in biblioteca che nei repertori bibliografici. La navigazione gerarchica consentita dai tesauri e dai sistemi di classificazione gerarchico-enumerativi come la Classificazione Decimale Dewey possono essere concepiti come strumenti granulari, secondo la definizione di Pritchard. Tradizionalmente la granularità dei tesauri e sistemi classificatori consentiva al lettore una navigazione tra i libri e/o tra gli articoli scientifici. La liberazione dal supporto cartaceo e la creazione di un documento elettronico organizzato in maniera articolata e dal testo finemente marcato e semanticamente annotato hanno aperto la possibilità di riconciliare l'indicizzazione esterna e quella interna ai documenti⁶¹.

⁶¹ A questo proposito, cfr. come XML sia oggi presentato anche agli indicizzatori di professione da un recente articolo di Bill Kasdorf come lo strumento per sanare la storica distinzione tra *open-system indexing* (quella appunto operata su monografie e articoli di rivista attraverso vocabolari controllati), *closed-system indexing* (l'indicizzazione analitica tipica delle monografie e presente al termine del testo) e *word lists* (la ricerca automatica della parola nel testo) (Kasdorf 2004).

Navigare tra concetti e relazioni è la prospettiva più ambiziosa che informa attualmente le sperimentazioni portate avanti in tutto il mondo dai ricercatori che si dedicano alla creazione del Web semantico. La prospettiva indicata da Tim Berners-Lee è quella di un Web evoluto in grado di consentire agli strumenti automatizzati di leggere e “comprendere” – dove “semantico” sta per “leggibile dalla macchina” – le informazioni, anzi le metainformazioni disponibili in Rete. Alla base di questo sistema, la marcatura XML dei documenti e la creazione da parte delle diverse comunità di utilizzatori della rete di nuove mappe concettuali, conosciute come ontologie (Berners-Lee e Hendler 2001)⁶², in grado di guidare gli automi alla ricerca dell’informazione.

Il Web semantico costituisce il principale sforzo teorico e sperimentale a livello internazionale che si propone di arrivare a un più efficiente ed efficace recupero dell’informazione in rete. Si tratta di un’esperienza che potrebbe integrare i risultati e le riflessioni migliori di settori diversi come quelli dell’informatica, delle scienze dell’informazione e dell’intelligenza artificiale, portando a evoluzione strumenti già conosciuti come quelli dei tesauri e dei sistemi di classificazione, e procedure informatiche come l’elaborazione del linguaggio naturale e l’estrazione automatica dell’informazione dai testi. Essenziale in questa prospettiva risulta la creazione da parte delle diverse comunità di scienziati e ricercatori di ontologie in grado di consentire la navigazione tra i documenti. Un’ontologia è un’esplicita specificazione della concettualizzazione di un dominio, di una realtà, ed è effettuata da una comunità di ricerca, sulla base della letteratura esistente e della consultazione degli esperti di quel settore, generalmente con un obiettivo specifico.

«Domain ontologies are constructed by capturing a set of concepts and their links according to a given context. A context can be viewed as various criteria such as the abstraction paradigm, the granularity scale, interest of user communities, and the perception of ontology developer» (Vickery 1997).

Questi strumenti andrebbero a innestarsi sulla preesistente esperienza dei tesauri disciplinari, considerati non più in grado di rendere in modo esplicito tutte le possibili relazioni tra concetti, che non possono essere ristrette esclusivamente alle relazioni gerarchiche.

Il livello di dettaglio raggiunto da una ontologia – la sua granularità, come per l’appunto viene chiamata nella letteratura specializzata – è una caratteristica essenziale per valutare le sue potenzialità in fase di trattamento e recupero dell’informazione. La scelta di decidere quali e quanti livelli di granularità assegnare a un’ontologia può costituire un problema e necessitare di una ridefinizione qualora

⁶² Per una presentazione più completa del Web semantico, rimando al chiaro e affidabile manuale di Grigoris Antoniou e Frank Van Harmelen (Antoniou e Van Harmelen 2004).

si decida di modificare gli obiettivi per i quali si era creata l'ontologia oppure si desideri "comunicare" tra ontologie diverse⁶³.

Le ontologie potrebbero essere collegate ai testi opportunamente marcati, ma soprattutto a quelle unità documentarie fino a oggi non raggiunte dall'indicizzazione umana e collocate al livello inferiore rispetto a quello dell'articolo. Il recupero di paragrafi, dati, fatti, e nomi sarebbe poi consentito agli agenti intelligenti (strumenti software costruiti per ricercare l'informazione sulla base di compiti e indicazioni assegnati loro dai ricercatori umani) in grado di effettuare la navigazione concettuale a livello di ontologie.

Il problema della granularità delle ontologie non si esaurisce in una semplice scelta di "livello di dettaglio". Se ne rende conto uno studioso come Brian Vickery, riportando le opinioni di alcuni studiosi dell'intelligenza artificiale su cosa possa essere la granularità di un'ontologia.

«There has been discussion of what has been called the 'granularity' or 'grain size' of an ontology – to what degree of specificity should the concept hierarchy be continued. Some authors, e.g. Poli, concentrate on the top-level types of concept occurring in the domain – the ontological *categories*. For him, 'an ontology is not a catalogue of the world, a taxonomy, a terminology of a list of objects, things or whatever else: it is the general framework or structure within which catalogues, taxonomies, terminologies may be given suitable organisation'. Others, e.g. Guarino, see the need to include all the specific concepts occurring in the domain, and urge that an ontology should be of high granularity» (Vickery 1997).

In una ontologia, le relazioni tra concetti potranno essere di tipo diverso, per esempio:

- *tassonomico* (IS_A)
- *meronimico* (PART_OF)⁶⁴
- *telico* (PURPOSE_OF)

⁶³ Il tema della granularità è stata oggetto di lunghe ricerche nel campo dell'intelligenza artificiale, a partire dall'articolo di Hobbs del 1985. Un libro di testo di riferimento, compilato da Elaine Rich e Kevin Knight affronta il tema della scelta della granularità di rappresentazione.

«Indipendentemente dal particolare formalismo di rappresentazione scelto, è necessario chiedersi a che livello di dettaglio si debba rappresentare il mondo reale, ovvero quali debbano essere le primitive da utilizzare: ne dobbiamo prevedere poche e di basso livello, o molte in grado di ricoprire un ampio spettro di livelli» (Rich e Knight 1992, p. 126).

⁶⁴ Per definire il significato di meronimico, riprendo un brano di una recente introduzione al concetto di ontologia in ambito filosofico, curata da Achille C. Varzi: «Il capitolo più studiato dell'ontologia formale è senz'altro la mereologia (dalla parola greca *meros*, parte), cioè la teoria delle relazioni parte-tutto: le relazioni che sussistono tra un'entità *x* e le sue parti, e tra le diverse parti di una data entità *x*, indipendentemente dalla natura di *x*. [...] il tema principale su cui negli ultimi decenni si è concentrata l'attenzione riguarda la definizione dei principi fondanti della teoria, intesa appunto come una teoria formale applicabile a *qualsiasi* dominio di entità» (Varzi 2005).

In questo quadro, la granularità andrebbe a inserirsi tra le relazioni parte/tutto, ma ciò non appare sufficiente a descrivere tutte le realtà naturali che si intende descrivere. La riflessione teorico-pratica sull'argomento ha assunto negli ultimi anni un certo rilievo, per l'impatto che l'organizzazione di ontologie finisce per avere in ogni settore di ricerca. Ciascun nel proprio specifico disciplinare, i ricercatori dovranno definire o ridefinire formalmente tutti i concetti, i fenomeni e le relazioni abitualmente utilizzati, proprio per renderli meglio leggibili dagli strumenti automatici. La ricerca sulle relazioni che sostanziano le ontologie disciplinari ha assunto una sua autonomia, che tenta da una parte di abbracciare le problematiche simili a livello generale e interdisciplinare, dall'altra è chiamata a confrontarsi con ciascuno specifico disciplinare. Tra gli studiosi che si occupano di questo settore, da anni Thomas Bittner lavora specificamente sul concetto di granularità, e lo ha affrontato in diversi ambiti, da quello biomedico a quello geografico. In un recentissimo articolo Alan Rector, Jeremy Rogers e Thomas Bittner, tentano di illuminare proprio il concetto di granularità nelle ontologie, districandosi dai diversi significati che sono stati dati nel tempo a questo termine. Gli esempi richiamati sono i seguenti:

Significati del termine <i>granularity</i> riscontrati in letteratura (Rector, Rogers e Bittner 2005)	
Specialisation	Category vs. kind—the usual notion of “is-kind-of,” e.g., that “mammal” is a generalisation including, amongst other things, dogs and elephants. Sometimes also labelled ‘abstraction’
Degree of detail	The amount of information represented about each entity, regardless of its level of specialisation. Crudely in an ontology represented in OWL [linguaggio formale con cui viene rappresentata un'ontologia], the number of axioms and restrictions concerning each entity
Density	The number of semantically ‘similar’ concepts in a particular conceptual region. How “bushy” the subsumption graph is. High local density in an ontology usually co-occurs with high levels of specialisation and degree of detail, but in two different ontologies of the same overall depth, in a particular section one may find the same two categories separated by different numbers of intervening categories or possessing very different numbers of sibling categories
Connectivity	The number of entities connected directly and indirectly to a given entity either through generalisation/specialisation or by other properties

Volendo sintetizzare al massimo il senso della loro ricerca – che appare ancora in fieri – i tre autori sostengono che due sono sostanzialmente le nozioni confuse sotto il termine di granularità: la scala dimensionale, che si riferisce sostanzialmente alle dimensioni fisiche, e la collettività, che si collega all'aggregazione di entità individuali in entità collettive che hanno proprietà e caratteristiche diverse da quelle delle entità poste al livello inferiore. Situazioni di questo tipo sono costituite dalle molecole nell'acqua, dalle cellule rispetto ai tessuti, dalle persone componenti folle, dalle stelle rispetto alle galassie. Affrontando su queste basi le ulteriori riflessioni, gli autori giungono a individuare le caratteristiche specifiche di questi due modi di affrontare il concetto di granularità, riservando maggiore attenzione proprio al concetto di "collectivity", che si rivela il più complesso e interessante, proprio al fine di "ricucire" le differenze di scala che, soprattutto nell'ambito biologico, distinguono a vari livelli le molecole, le cellule, gli organi, gli organismi, fino agli ambiente ecologici.

La strada da percorrere su questa tematica è ancora lunga, ma valeva la pena portare l'attenzione su di un approccio e un metodo che appare in prospettiva estremamente fecondo.

Un quadro riassuntivo

Il sondaggio effettuato in questo articolo ha consentito di individuare una serie di usi e significati non sempre convergenti ma ciascuno di grande interesse. Le incursioni effettuate in altre aree disciplinari hanno ampliato le prospettive della ricerca e credo abbiano aperto a ulteriori approfondimenti di indagini in settori non ancora sfiorati. Per offrire una sorta di cartografia orientativa in grado di riepilogare i principali risultati dell'indagine, ritengo opportuno sintetizzare in un unico strumento di riferimento tutte quelle esperienze in cui il termine "granularità" ha assunto un particolare significato o è stato utilizzato in maniera degna di nota.

Dal quadro sommariamente presentato nelle pagine seguenti, credo emerga evidente come nella metafora della granularità appaiano leggibili approcci diversi, via via più elaborati e raffinati: uno più "orizzontale" (il documento e le sue plurime parti componenti in formato digitale), uno più "verticale-gerarchico" (documento organizzato in maniera da consentire la lettura e il recupero in maniera efficiente), uno più compiutamente "relazionale-contestuale", che si pone l'obiettivo di dare conto al lettore delle relazioni che legano i documenti, le loro parti componenti, i concetti stessi. Quest'ultimo è l'approccio sicuramente più complesso, che richiede un impegno costante da parte di chi è chiamato a costruire le ontologie concettuali e di chi sarà deputato a curare i legami relazionali tra le diverse espressioni e manifestazioni documentali.

QUADRO RIASSUNTIVO DELLA GRANULARITÀ NELLE

DOCUMENTO A TESTO PIENO	
Caratteristiche granulari	Strumenti e soluzioni riscontrate in letteratura
Granularità del testo sequenziale	<p>Stream di caratteri, ogni carattere è codificato, l'unico livello di marcatura è lo spazio bianco e il <i>return</i></p> <p>Ricerca della lettera, della parola, delle frasi o dei passaggi</p> <p>Identificazione di importanti concetti in paper scientifici altamente strutturati (Paice & Jones 1993) per la sommarizzazione automatica</p>
Granularità ipertestuale	<p>Ripartizione del testo in lessie (o chunk) e creazione dei link ipertestuali</p> <p>Lettura e ricerca tramite il browsing</p>
Granularità nella marcatura del testo	<p>Strutturazione del documento secondo formati logici: marcatura del testo a livelli sempre più dettagliati (specificazioni e qualificazione a livello di paragrafo o di elemento oppure con la creazione di dettagliati DTD)</p> <p>Dialetti scientifici XML based: SVG, MathML, UnitsML, STXML, CML (Lu 2005)</p> <p>Tensione tra strutturazione logica e contenuto del documento (Heeman 2002)</p> <p>Proposte di marcature sempre più semantiche (utilizzazione della teoria della struttura retorica del testo)</p> <p>Concetto di datumento: Murray-Rust 2003</p>
Composizione granulare del documento	<p>Partizione del documento in paragrafi</p> <p>Paragrafi, sezioni, citazioni, note, (struttura topica o irmdc)</p> <p>Strutturazione gerarchica del documento e segmentazione delle sue parti</p> <p>Strutturazione del testo e dell'abstract</p> <p>Kircs e Harmsze e modularità / granularità del documento scientifico</p> <p>Presenza di realtà documentarie diverse (suoni, testi, animazioni)</p>
Searching granulare	<p>Necessità di integrare searching e browsing per il documento strutturato (Chiaromella)</p> <p>Integrazione dell'indicizzazione con tesauri, sistemi classificatori e ontologie</p> <p>Navigazione per concetti (tesauri, ontologie, topic maps, ecc.) tra documenti e parti componenti del documento</p>

SCIENZE DEL LIBRO E DELLA DOCUMENTAZIONE

	Estrazione dell'informazione e associazione dei testi marcati a ontologie XML e RDF, ontologie > Web semantico
Visualizzazione granulare	Multilevel document visualization
Architettura granulare dei documenti e delle risorse	Granularità a livelli degli strumenti di e-learning Documento organizzato per concetti (topics): DITA, Darwin information typing architecture Architettura dell'informazione (Rosenfeld e Morville) Analisi grafica e interpretazione semantica delle diverse parti del documento (per es., digitalizzazione di documenti cartacei)

SURROGATO DEL DOCUMENTO

Caratteristiche granulari	Strumenti e soluzioni riscontrate in letteratura
Granularità della descrizione	<p>Scelta dell'unità documentaria da descrivere</p> <ul style="list-style-type: none"> • Monografie: capitoli, sezioni, paragrafi • Riviste: articoli, sezioni, paragrafi, ecc. • Video: sezioni, scene, shot, ecc. • RER e REL: siti, pagine, parti componenti • Opposizione tra collezione e item <p>Problema: rapporto 1:1 tra doc. e metadato</p> <p>Granularità del metadato: campi, sottocampi, relazioni, ecc. per una ricerca nei cataloghi e nelle bibliografie</p> <p>Necessità di rendere i metadati il più possibile dettagliati; granularità del formato TEI (Tennant 2002)</p> <p>Diversa capacità dei metadati di favorire la ricerca granulare degli utenti (Lagoze 1997)</p>
Granularità dell'identificazione	<p>Utilizzazione di numeri standard relativi a diverse entità bibliografiche: opere, documenti, ecc.: DOI, ISTC, ISSN</p> <p>Concetto di granularità funzionale (Paskin)</p>
Gestione delle relazioni tra documenti	<p>Necessità di evidenziare i rapporti tra i diversi documenti (fa parte di, altra ed. di, ecc.), (Donor, Tillett, Jonsson)</p> <p>Descrizione della collezione a molteplici livelli</p> <p>Concetto di granularità funzionale (Le Boeuf, ISTC)</p> <p>Equilibrio tra descrizione dell'item e collezione, grazie anche a strumenti di visualizzazione (Foulonneau 2005)</p>

Il quadro complessivo delle caratteristiche “granulari” delle risorse informative – siano esse testi, documenti o surrogati bibliografici – potrebbe rischiare di apparire eccessivamente sbilanciato in senso “datocentrico”. Testi e documenti non possono essere ripartiti, strutturati o marcati in forma “oggettiva” e una volta per tutte, come non è possibile stabilire un unico o ottimale modo di definire una volta per tutte i metadati relativi a una risorsa o a una collezione di risorse. Se è lecito concludere questo censimento con l’indicazione di una prospettiva ulteriore di indagine, indicherei l’obiettivo di rileggere in maniera più esplicita le problematiche della granularità attraverso il focus costituito dagli orientamenti, dalle esigenze, dai comportamenti di chi si trova a cercare informazione.

In the recent years the term ‘granularity’ has increasingly come into usage within the librarian and archival field, though it does not seem to be employed in a uniform way. The research here presented was mainly focused on “mapping” the various meanings of the term. ‘Granular’ can be used when referring to a resource structured as blocks of text (lexia) kept together by hyperlinks; to a resource’s metadata; to the description of or access to an article contained within a periodical or within one of the periodical’s files. The research highlighted the large usage of the term, but also the historical profoundness of the associated concepts and meanings. Nevertheless, in a context in which digital technologies have made it possible to search for the smallest entities of a resource (you can search for a word, a line, a citation, for the article itself, etc.), it becomes necessary to always deliver information on the relations –especially the hierarchical ones– that exist between a resource’s various components, so as to facilitate end users’ navigation.

Au cours des dernières années le terme granularité s’est de plus en plus diffusé dans le domaine des disciplines du livre et du document avec des utilisations souvent différentes. Le but principal de cette recherche a été celui de «cartographie» les significations que ce terme a assumé. Granulaire peut indiquer une ressource constituée par des blocs de textes (lexies) unis par des liens hypertextuels ; granulaire peut indiquer une métadonnée relative à une ressource ; granulaire renvoie à la description de ou à l’accès à un article d’un périodique plutôt qu’au fascicule dans lequel il est imprimé. Les exigences commerciales et le développement des technologies se disputent aussi ces définitions. Le recensement effectué souligne non seulement combien ce terme est diffusé mais aussi la profondeur historique des concepts et des significations qui lui sont associés. D’autre part, la recherche de l’accès aux plus petites entités d’une ressource (recherche du mot, de la phrase, de la citation, de l’article, etc.), favorisée par le développement des technologies numériques, ne peut ignorer la nécessité de fournir à l’utilisateur les relations nécessaires, surtout les relations hiérarchiques, qui existent entre les différentes parties composantes des ressources et qui peuvent lui permettre une navigation plus aisée.

RIFERIMENTI BIBLIOGRAFICI

- Aarseth 1997: Espern J. Aarseth. *Cybertext: perspectives on ergodic literature*. Baltimora: John Hopkins University Press, 1997.
- Anderson e Perez-Carballo 2005: James D. Anderson – Jose Perez-Carballo. *Information retrieval design: principles and options for information description, organization, display, and access in information retrieval databases, digital libraries, and indexes*. St. Petersburg (FL): Ometeca Institute. 2005. <http://www.scils.rutgers.edu/~carballo/ird2005.html>.
- Antelman 2005: Kristin Antelman. *Identifying the serial work as a bibliographic entity*. «LRTS», 48 (2005), n. 4, p. 238-255.
- Antoniou e Van Harmelen 2004: Grigoris Antoniou – Frank Van Harmelen. *A semantic web primer*. Cambridge, London: Mit Press, 2004.
- Atzeni 2003: Paolo Atzeni et al. *Basi dati. Architetture e linee di evoluzione*. Milano: McGraw-Hill, 2003.
- Baeyer 2004: Hans Christian Von Baeyer. *Information: the new language of science*. Cambridge (Mass.): Harvard University Press, 2004. Trad. it.: *Informazione. Il nuovo linguaggio della scienza*. Bari: Dedalo, 2005.
- Battaglia 1961-2002: Salvatore Battaglia. *Grande dizionario della lingua italiana*. Torino: UTET, 1961-2002.
- Bellardo Hahn e Buckland, 1998: Trudi Bellardo Hahn – Michael Buckland (a cura di). *Historical studies in information science*. Medford (NJ): Information Today for the American Society for Information Science, 1998.
- Berners-Lee e Hendler 2001: Tim Berners-Lee – James Hendler. *Publishing on the semantic web. The coming Internet revolution will profoundly affect scientific information*. «Nature», 410, 2001, p. 1023-1024. <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>.
- Bianchini 2005: Carlo Bianchini. *Riflessioni sull'universo bibliografico: funzioni, oggetti e modelli della catalogazione per autore e titolo*. Milano: Bonnard, 2005.
- Bies 1996: Werner Bies. *Thinking with the help of images: on the metaphors of knowledge organization*. «Knowledge organization», 23 (1996), n. 1, p. 3-8.
- Blackmore 2002: Susan Blackmore. *La macchina dei memi: perchè i geni non bastano*. Torino: Instar libri, 2002.
- Bowden, Bellardo Hahn e Williams 1999: *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, a cura di Mary Ellen Bowden, Trudi Bellardo Hahn, Robert V. Williams. Medford (NJ): Information Today for American Society for Information Science and Chemical Heritage Foundation, 1999.
- Bradley 2001: Neil Bradley. *Xml companion*. Boston: Addison-Wesley, 2002.
- Brooks 2001: Terrence A. Brooks. *Where is meaning when form is gone?: knowledge representation on the Web*. «Information Research», 6 (2001), n. 2. <http://InformationR.net/ir/6-2/paper93.html>.
- Byrum 2005: John D. Byrum Jr. *Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it*. In: *Libraries: a voyage of discovery. atti del World Library and Information Congress, 71th IFLA General Conference and Council, Oslo, 14-18 agosto 2005*. <http://www.ifla.org/IV/ifla71/papers/124e-Byrum.pdf>. Trad. it.: *Raccomandazioni per miglioramenti urgenti dell'OPAC*. «Biblioteche oggi», 2005, p. 5-14.

- Capurro e Hjørland 2003: Rafael Capurro – Birger Hjørland. *The concept of information*. «Annual review of information science & technology», 37 (2003), p. 343-411.
- Carlini 1999: Franco Carlini. *Lo stile del Web: parole e immagini nella comunicazione di rete*. Torino: Einaudi, 1999.
- Caton 2001: Paul Caton. *Markup's current imbalance*. «Markup Languages: Theory & Practice», 3 (2001), n. 1, p. 1-13.
- Cavarero 2003: Adriana Cavarero. *A più voci. Filosofia dell'espressione vocale*. Milano: Feltrinelli, 2003.
- Cavina e La Piccirella 2002: Enrica Cavina – Rossella La Piccirella. *Information design: progettare la comunicazione su basi cognitive*. Torino: Utet, 2002.
- Chiarabella 2000: Yves Chiarabella. *Information retrieval and structured documents*. In: Maristella Agosti – Fabio Crestani – Gabriella Pasi (a cura di). *Lectures on Information Retrieval: Third European Summer School, Essir 2000 Varenna, Italy, September 11-15, 2000: Revised Lectures (Lecture Notes in Computer-Science)*. Berlin [ecc.]: Springer, 2001. p. 286-309.
- Chowdhury 2004: Gobinda G. Chowdhury. *Introduction to modern information retrieval*. London: Facet, 2004.
- Chu 2003: Heting Chu. *Information representation and retrieval in the digital age*. Medford (NJ): Information today for American Society for Information Science and Technology, 2003.
- Conklin 1987: Jeff Conklin. *Hypertext: an introduction and survey*. «Computer», 20 (1987), n. 9, p. 17-41.
- Coombs, Renear e DeRose 1987: James H. Coombs – Allen H. Renear – Steven J. DeRose. *Markup systems and the future of scholarly text processing*. «Communications of the ACM archive», 30 (1987), n. 11, p. 933-947.
- Day 2000: Ronald E. Day. *The "conduit metaphor" and the nature and politics of information studies*. «Journal of the American Society for Information Science», 51 (2000), n. 9, p. 805-811.
- Debons 1992: Anthony Debons. *The measurement of knowledge*. In: *Proceedings of the Fifth Annual Meeting of the American Society for Information Science, Pittsburgh, 26-29 ottobre 1992*, a cura di Debra Shaw. Medford (NJ): Learned Information for American Society for Information Science, 1992, p.212-15.
- De Mauro 1999-2000: *Grande dizionario italiano dell'uso*, ideato e diretto da Tullio De Mauro. Torino: UTET, 1999.
- DeRose et al. 1990: Steven J. DeRose – David G. Durand – Elli Mylonas – Allen H. Rehear. *What is text, really?*. «Journal of Computing in Higher Education», 2 (1990), n. 1, p. 3-26.
- DeRose 1999: Steven J. DeRose. *XML and the TEI*. «Computers and the Humanities», 33 (1999), p. 11-30.
- Devlin 1991: Keith J. Devlin. *Logic and Information*. Cambridge: Cambridge University Press, 1991.
- Fayet-Scribe 1997 : Sylvie Fayet-Scribe. *Chronologie des supports, des dispositifs spatiaux, des outils de repérage de l'information*. «Solaris», 1997, n. 4. http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d04/4fayet_2com.html.
- Fonseca et al. 2002: Frederico Fonseca – Max Egenhofer – Clodoveu Davis – Gilberto Câmara. *Semantic granularity in ontology-driven geographic information systems*. «Annals of Mathematics and Artificial Intelligence», 36 (2002), p. 121-151.
- Foulonneau et al. 2005: Muriel Foulonneau – Timothy W. Cole – Thomas G. Habing – Sarah L. Shreeves. *Using collection descriptions to enhance an aggregation of harvested item-level metadata*. In: *Proceedings of the 5th ACM/IEEE Joint Conference on Digital Libraries, 7-11 giugno 2005. Denver, Colorado*. New York: ACM Press, 2005, p. 32-41.

- Gagliasso 2002: Elena Gagliasso. *Usi epistemologici della metafora e metafore cognitive*. In: Morabito 2002, p. 1-17.
- Gambari e Guerrini 2002: Stefano Gambari – Mauro Guerrini. *Definire e catalogare le risorse elettroniche: un'introduzione a ISBD(ER), AACR2 e metadati*. Milano: Bibliografica, 2002.
- Ghilli e Guerrini 2001: Carlo Ghilli – Mauro Guerrini. *Introduzione a FRBR*. Milano: Bibliografica, 2001.
- Gilbert, Sim, Wang 2005: Lester Gilbert – Yee-Wai Sim – Chu Wang. *Modelling the learning transaction*. In: *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*. Piscataway: IEEE, 2005, p. 615-616.
- Guercio 2001: Maria Guercio. *Rischi e promesse dell'innovazione tecnologica*. «Bollettino AIB», 41 (2001), n. 2, p. 157-173.
- Hakala 2002: Juha Hakala. *Principi di identificazione: prospettive europee*. In: Mauro Guerrini (a cura di). *Le risorse elettroniche: definizione, selezione e catalogazione: atti del convegno internazionale, Roma, 26-28 novembre 2001*. Milano: Bibliografica, 2002, p. 77-91.
- Harrod 2000: *Harrod's librarians' glossary and reference book*, a cura di Ray Prytherch. Aldershot: Gower, 2000.
- Heeman 1992: Frans C. Heeman. *Granularity in structured documents*. «Electronic publishing», 5 (1992), n. 3, p. 143-155.
- Hillesund 2005: Terye Hillesund. *Digital text cycles: From medieval manuscripts to modern markup*. «Journal of Digital Information», 6 (2005), n.1.
<http://jodi.tamu.edu/Articles/v06/i01/Hillesund/>.
- Hillmann 2005: Diane Hillmann. *Using Dublin Core: The Elements*. 2005.
<http://dublincore.org/documents/2005/11/07/usageguide/elements.shtml>.
- Hobbs 1985: Jerry R. Hobbs. *Granularity*. In: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, 1985, p. 432-435.
- Horn 1989: Robert E. Horn. *Mapping Hypertext*. Lexington (Mass.): The Lexington Institute, 1989.
- Johnson 1994: Gerald J. Johnson. *Of metaphor and the difficulty of computer discourse*. «Communications of the ACM», 37 (1994), n. 12, p. 97-102.
- Jonsson 2003: Gunilla Jonsson. *The bibliographic unit in digital context: how to define it?*, presentato al *First IFLA Meeting of Experts on an International Cataloguing Code*.
http://www.ddb.de/news/ifla_conf_index.htm.
- Kasdorf 2004: Bill Kasdorf. *Indexers and XML: an overview of the opportunities*. «The indexer», 24 (2004), n. 2, p. 75-78.
- Kerne et al. 2003: Andruid Kerne – Vikram Sundaram – Jin Wang – Madhur Khandelwal – J. Michael Mistrot. *Human + agent: creating recombinant information*. In: *International Multimedia Conference archive: proceedings of the eleventh ACM international conference on Multimedia, Berkeley (CA), 2003*. New York: ACM, 2003, p. 454-455.
- Kerne, Khandelwal e Sundaram 2003: Andruid Kerne – Madhur Khandelwal – Vikram Sundaram. *Publishing evolving metadocuments on the web: conference on Hypertext and Hypermedia archive: proceedings of the fourteenth ACM conference on Hypertext and Hypermedia, Nottingham, UK, 2003, Session: Link aggregation*. New York: ACM Press, 2003, p. 104-105.
- Kline 2004: Ronald R. Kline. *What is information theory a theory of?: boundary work among information theorists and information scientists in the United States and Britain during the Cold War*. In: *The history and heritage of scientific and technological information systems: proceedings of the 2002 conference*. Medford (NJ): Information Today for American Society for Information Science and Technology and Chemical Heritage Foundation, 2004.

- <http://www.chemheritage.org/events/asist2002/01-kline.pdf>.
- Lagoze 1997: Carl Lagoze. *From static to dynamic surrogates: resource discovery in the digital age*. «D-Lib magazine», giugno 1997. <http://www.dlib.org/dlib/june97/06lagoze.html>.
- Lancaster e Warner 2001: Frederick Wilfrid Lancaster – Amy Warner. *Intelligent technologies in library an information service applications*. Medford (NJ): Information today for American Society for information science and technology, 2001.
- Landow 1997: George P. Landow. *Hypertext 2.0: the convergence of contemporary critical theory and technology*. Baltimore (MD): Johns Hopkins University Press, 1997. Trad. it.: *L'ipertesto: tecnologie digitali e critica letteraria*. Milano: Bruno Mondadori, 1998.
- Le Bouef 2005: Patrick Le Bouef. *Identifying 'textual works': ISTC: controversy and potential*. In: *FRBR in 21st Century Catalogues: An Invitational Workshop, 2-4 maggio 2005, Dublin (Ohio)*. <http://www.oclc.org/research/events/frbr-workshop/presentations/leboeuf/ISTC.ppt>.
- Lu 2005: Zhongyu Lu. *A survey of Xml applications on science and technology*. «International Journal of Software Engineering and Knowledge Engineering», 15 (2005), n. 1, p. 1-33.
- Macgregor 2003: George Macgregor. *Collection-level descriptions: metadata of the future?*. «Library Review», 52 (2003), n. 6, p. 247-250.
- MacKay 1950: Donald M. Mackay. *Quantal aspects of scientific information*. «Philosophical Magazine», 41 (1950), p. 289-311.
- McGraw-Hill Zanichelli 2004: *Il McGraw-Hill Zanichelli. Dizionario enciclopedico scientifico e tecnico inglese italiano, italiano inglese*. Bologna: Zanichelli, 2004.
- Merton 2002: Robert K. Merton – Elinor G. Barber. *The travels and adventures of Serendipity: a study in historical semantics and the sociology of science*. Trad. it.: *Viaggi e avventure della Serendipity: saggio di semantica sociologica e sociologia della scienza*, Bologna: Il Mulino, 2002.
- Morabito 2002: Carmela Morabito. *La metafora nelle scienze cognitive*. Milano: McGraw-Hill, 2002.
- Murray-Rust e Rzepa 2003: Peter Murray-Rust – Henry S. Rzepa. *The next big thing: from hypermedia to datuments*. «Journal of Digital Information», 5 (2003), n. 1. <http://jodi.tamu.edu/Articles/v05/i01/Murray-Rust/>.
- Nunberg 1993: Geoffrey Nunberg. *The places of books in the age of electronic reproduction*. «Representations», 24 (1993), estate. Rip. in: Howard Bloch – Carla Hesse (a cura di). *Future libraries*. Berkeley: University of California Press, 1994. <http://www-csl.stanford.edu/~nunberg/places3.html>.
- Oxford English Dictionary: *The Oxford English dictionary*, a cura di John A. Simpson, Edmund S. C. Weiner. Oxford: Clarendon Press, 1989. <http://www.oed.com>.
- Paskin 1999: Norman Paskin. *Toward unique identifiers*. «Proceedings of the IEEE», 87 (1999), n. 7, p. 1208-1227.
- Paskin 2003: Norman Paskin. *On making and identifying a "copy"*. «D-Lib Magazine», 9 (2003), n. 1. <http://www.dlib.org/dlib/january03/paskin/01paskin.html>.
- Patton 2000: Glenn Patton. *Crossing a digital divide: AACR2 and unaddressed problems of networked resources*. In: *Library of Congress Bicentennial Conference on Bibliographic Control in the New Millennium, 15-17 novembre 2000*. http://www.loc.gov/catdir/bibcontrol/patton_paper.html.
- Peereboom 1998: Marianne Peereboom. *DONOR WP1 Granularity and the use of DC.Relation: WP1.1. Discussion paper, version 1.0*. 1998. <http://www.kb.nl/coop/donor/project-en-index.html?coop/donor/rapporten/granularity.html>.
- Powell, Heaney e Dempsey 2000: Andy Powell – Michael Heaney – Lorcan Dempsey. *RSLP collection description*. «D-Lib Magazine», 6 (2000), n. 9.

- <http://www.dlib.org/dlib/september00/powell/09powell.html>.
- Raieli e Innocenti 2004: Roberto Raieli – Perla Innocenti (a cura di). *Multimedia information retrieval: metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione*. Roma: AIDA, 2004.
- Rayward 1996: W. Boyd Rayward. *The history and historiography of information science: some reflections*. «Information Processing & Management», 32 (1996), n. 1, p. 3-17; Rip. in: Bellardo Hahn e Buckland, 1998, p. 7-21.
- Rayward 1998: W. Boyd Rayward. *Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext*. In Bellardo Hahn e Buckland, 1998, p. 65-80.
- Rector, Rogers e Bittner 2005: Alan Rector – Jeremy Rogers – Thomas Bittner. *Granularity scale and collectivity: when size does and does not matter*. «Journal of biomedical informatics», 39 (2006), n. 3, p. 333-349.
- Rich e Knight 1991: Elaine Rich – Kevin Knight. *Artificial intelligence*. 2. Ed., New York: McGraw-Hill, 1991. Trad. it.: *Intelligenza artificiale*, Milano: McGraw-Hill, 1992.
- Robertson 1993: Stephen E. Robertson. *Information granularity: a theme in the history of information science and technology*. Intervento (non pubblicato) all'Annual Meeting dell'American Society for Information Science, Session on the History of Information Science; il riassunto è in: <http://www.sims.berkeley.edu/~buckland/asiconf.html>.
- Rosenfeld e Morville 2002: Louis Rosenfeld – Peter Morville. *Information architecture for the World Wide Web*. Cambridge: O'Reilly, 1998. Trad. it.: *Architettura dell'informazione per il World Wide Web*. Milano: Hops, c2002.
- Rouse e Rouse 1989: Richard H. Rouse – Mary A. Rouse. *La naissance des index*. In: Roger Chartier – Henri-Jean Martin (sous la direction de), *Histoire de l'édition française: le livre conquérant, du moyen âge au milieu du XVIIème siècle*. Paris: Fayard, 1989, p. 95-108.
- Ruecher, Homich, Sinclair 2005: Stan Ruecker – Eric Homich – Stefan Sinclair. *Multi-level document visualization*. «Visible language» 39 (2005), n. 1, p. 33-41.
- Serrai 2001a: Alfredo Serrai. *L'ontologia catalografica*. In: Alfredo Serrai. *Il cemento della bibliografia*. Milano: Bonnard, 2001, p. 13-32.
- Serrai 2001b: Alfredo Serrai. *Linguaggio informazione cultura*. In: Alfredo Serrai, *Il cemento della bibliografia*. Milano: Bonnard, 2001, p. 43-52.
- Shannon e Weaver 1949: Claude E. Shannon – Warren Weaver. *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949. Trad. it.: *La teoria matematica delle comunicazioni*. Milano: ETAS Compass, 1971.
- Smeaton 2004: Alan F. Smeaton. *Indexing, browsing, and searching of digital video*. «Annual review of information science and technology», 38 (2004), p. 371-407.
- Smith 1992: Joan M. Smith. *SGML and related standards: document description and processing languages*. Upper Saddle River (NJ): Horwood, 1992. Trad. it.: *SGML e altri standard: linguaggi di descrizione ed elaborazione dei documenti*. Bologna: Clueb, 1997.
- Tennant 2002: Roy Tennant. *The Importance of being granular*. «Library Journal», 127 (2002), n. 9, p. 32-34.
- Tillett 1991: Barbara B. Tillett. *A taxonomy of bibliographic relationships*. «Library resources & technical services», 35 (1991), n. 2, p. 150-158.
- Turner e Goodrum 2002: James M. Turner – Abby A. Goodrum. *Modeling videos as works*. «Cataloging & classification quarterly», 33 (2002), ? , p. 27-38. Rip. in: Richard P. Smiraglia (a cura di), *Works as entities for information retrieval*. New York: Haworth, 2002, p. 27-38.

- UNI EN ISO 9241-11: *Requisiti ergonomici per il lavoro di ufficio con videotermini: guida sull'usabilità*. Norma UNI EN ISO 9241-11, maggio 2002.
- Van der Werf-Davelaar 1999: Titia van der Werf-Davelaar. *Identification and resolution systems for networked access to information*. «Online & CD-Rom Review», 23 (1999), n. 6, p. 325-331.
- Vickery 1997: Brian C. Vickery. *Ontologies*. «Journal of Information Science», 23 (1997), n. 4, p. 277-286.
- Vitiello 2004: Giuseppe Vitello. *L'identificazione degli identificatori*. «Biblioteche oggi», 22 (2004), n. 2, p. 67-80.
- Walker 2006: Ruth Walker, *We're all going dotty*, «The Christian science monitor», 15 marzo 2006, p. 18.
- Worboys e Duckham 2004: Michael F. Worboys – Matt Duckham. *GIS: A Computing Perspective*. Boca Raton: CRC Press, 2004.