

Proposta di formalizzazione linguistica nella descrizione delle risorse elettroniche

di

Gianfranco Crupi

L'area o il campo meno strutturato in una registrazione catalografica di risorse elettroniche è quello relativo alla descrizione del suo contenuto, quello, per intenderci, indicato:

- nell'Area 7.10 di ISBD(ER) come *Note di sommario* («Il riassunto fornisce un resoconto fattuale e non valutativo dell'argomento trattato»);
- nel Blocco 3 di UNIMARC, relativo alle Note, come l'etichetta 327 *Nota di contenuto* e soprattutto come l'etichetta 330 *Sommario o abstract* («Il campo contiene un sommario o abstract della registrazione»), che ha come corrispettivi le etichette 505 e soprattutto 520 di MARC21 («Il campo contiene un'informazione non strutturata che descrive lo scopo e i contenuti generali dei materiali descritti. Può essere un sommario, un *abstract*, un'annotazione, una recensione, o soltanto una frase che descrive il materiale»);
- nell'Elemento di Dublin Core denominato *Descrizione* e definito come «una spiegazione del contenuto della risorsa»;
- infine, tra gli *Attributi di un'espressione* di FRBR, come l'«Indice del contenuto di un'espressione: riassunto, sintesi, sinossi ecc.» (4.3.9).

Al riguardo, risulta assai significativo un sondaggio svolto presso 100 utenti (50 studenti, 30 docenti, 20 dottorandi o dottori di ricerca) della Biblioteca "Angelo Monteverdi" dell'Università di Roma "La Sapienza", sulle modalità di ricezione e uso di record catalografici di risorse elettroniche. Il sondaggio ha utilizzato come campo di indagine l'opac della Biblioteca, dove figurano circa 1700 risorse remote, e il catalogo di CORC, relativamente agli ambiti disciplinari delle lingue e delle letterature romanze. Esso ha rilevato che il 91% degli intervistati, all'apertura di un record catalografico, focalizza la sua attenzione sull'area descrizione, in quanto ritenuta più ricca di informazioni specifiche rispetto alle proprie aspettative; il 68% degli intervistati ha poi dichiarato la sua insoddisfazione circa la qualità ed esaustività informativa, la pertinenza semantica, la precisione terminologica, la chiarezza formale con cui il campo è stato compilato.

Su questo secondo aspetto un esempio può essere illuminante:

la medesima tipologia di risorsa, un sito accademico che pubblica *on-line* una collezione di classici, viene descritta in differenti contesti catalografici come (cito solo l'*incipit* della descrizione o nota di contenuto): Archivio digitale di testi; Archivio di testi; Biblioteca digitale; Biblioteca elettronica; Contiene la raccolta completa di testi; Pubblica il testo integrale informatizzato; Sito che raccoglie una collezione di testi elettronici; Base di dati testuale; Il sito mette a disposizione un *corpus* di classici.

Dall'esempio si può notare: l'aleatorietà terminologica (la medesima tipologia documentaria è variamente denotata come archivio, biblioteca, collezione, raccolta, *corpus*, base di dati), l'oscillazione connotativa (si usa indifferentemente l'aggettivo elettronico, digitale, informatizzato), l'imprecisione della denotazione verbale (contiene, pubblica, raccoglie, mette a disposizione), con prevedibili effetti di disorientamento dell'utente.

A questi limiti, si possono aggiungere, più in generale: una frequente ridondanza informativa rispetto a quanto indicato in altri campi della registrazione bibliografica; uno stile talvolta involuto, che va a discapito della chiarezza e dell'efficacia dei contenuti; una struttura descrittiva incongruente rispetto alla struttura reticolare dei contenuti della risorsa; un lessico spesso poco pertinente e impreciso; infine, un punto di vista limitato, per cui in alcuni casi le descrizioni sembrano rivolte non alla comunità della rete ma ad un bacino di utenza circoscritto, non dal punto di vista delle competenze scientifiche - che è tutt'altro discorso che affronteremo poco oltre - ma dell'area geografica e culturale di appartenenza del compilatore.

Sono limiti che penalizzano, e talvolta notevolmente, la coerenza dell'intero record e l'equilibrio dei suoi diversi componenti rispetto al potenziale di informazione e conoscenza che esso veicola.

Negli standard catalografici e descrittivi prima indicati si dice che l'area o il campo "Descrizione" o "Nota di contenuto" possono essere compilati sotto forma di indice, sommario, riassunto o abstract, ascrivendo sostanzialmente la vasta tipologia dell'universo documentario delle risorse elettroniche - bene esemplificata da Riccardo Ridi nel suo intervento - alla forma testo, cioè a una struttura logica del documento, lineare e sequenziale, che non sempre si accorda con la reticolarità ipertestuale, multimediale e interattiva di risorse elettroniche complesse.

Tuttavia, va da sé che nei casi in cui la descrizione riguarda un singolo documento, tradizionalmente inteso, soccorre proficuamente una consolidata e affinata tradizione che si può avvalere sia della specifica letteratura critica sia delle raccomandazioni delle norme ISO 214-1976 e ANSI/NISO Z39.14-1997, dedicate alle istruzioni per la compilazione di *abstract*, intesi come sintetiche e accurate rappresentazioni del contenuto di un documento.

Ma qual è la funzione della nota di contenuto o della descrizione? La sua funzione è quella di aiutare il lettore a decidere se è necessario consultare il documento originale; come per gli *abstract*, una descrizione ben strutturata dovrebbe consentire infatti di identificare, velocemente e sinteticamente, il contenuto fondamentale di una risorsa per determinarne il valore informativo e l'utilità nei confronti dell'interesse del ricercatore. Dovrebbe fornire insomma come valore aggiunto un elemento decisionale che può consigliare o meno l'utente a proseguire la sua ricerca compiendo l'atto decisivo: attivare o meno il *link* che lo collegherà al documento.

Un'area dunque fin troppo critica e sensibile rispetto agli impazienti interrogativi dell'utente per poterne affidare il peso a una forse eccessiva discrezionalità, avente a punti di riferimento opposti il buon senso da una parte o le istruzioni per la compilazione di *abstract* dall'altra.

Rileviamo pertanto la necessità di una formalizzazione linguistica del campo "Descrizione" o "Nota di contenuto" in grado di cogliere dinamicamente e di rappresentare la ricca e mutevole tipologia combinatoria delle risorse elettroniche. Gli strumenti a disposizione sono, come abbiamo visto, poco flessibili per comunicare in una rappresentazione sequenziale chiusa una testualità aperta, che costringe il lettore a una modalità di lettura digressiva, per blocchi verbali e iconici, sfruttando al massimo quei segnali (*link*, parole chiave, titoli, icone) necessari per risospingere la sua navigazione nell'universo documentario: le metafore marine per indicare l'atto della lettura (navigazione, appunto) e la terminologia che definisce il lettore o meglio le diverse tipologie di lettori della rete (nel mondo anglosassone: *surfer*, *clicker*, *user*, *onliner*) ci dicono che quell'universo documentario è uno spazio aperto all'avventura e alla scoperta dell'ignoto.

E' dunque di fondamentale importanza conoscere e saper gestire il sistema reticolare di relazioni semantiche che presiede all'organizzazione di un dominio del sapere o di una disciplina, e la molteplicità dei contesti entro cui definire quelle relazioni, come è altrettanto importante indicare una gerarchia di livelli descrittivi in grado di isolare quegli elementi che possono soddisfare la curiosità o l'interesse di differenti comunità di lettori (ad esempio nella descrizione di un sito accademico, editore di pubblicazioni scientifiche e di strumenti bibliografici, promotore di laboratori di scrittura creativa e di *forum* e liste di discussione su particolari tematiche).

Qualsiasi descrizione non è a costo zero: sia che si ricorra a sistemi esperti, ancora però non in grado di tradurre coerentemente nella logica del linguaggio naturale la struttura e la semantica di un documento o a cogliere la differenza tra un'intenzione e una realizzazione (internet promette molto più di quanto non dia effettivamente); sia che si ricorra a specialisti che assommano competenze professionali a conoscenze di specifici ambiti disciplinari. Ma non tutto deve essere descritto, o comunque non allo stesso modo, e qui condivido l'opinione di coloro che affermano che non tutte le risorse elettroniche debbano essere catalogate o registrate, o comunque non allo stesso modo. Anche perché normalmente l'utente arriva alla nostra descrizione attraverso alcuni punti di accesso (l'autore, il titolo, un soggetto, un codice di classificazione), e dopo ancora

che qualcuno, intendo una biblioteca, ha selezionato quel documento e lo ha inserito tra le proprie raccolte, reali o virtuali poco importa, perché coerente con le proprie collezioni e con gli ambiti disciplinari di propria competenza. L'utente arriva quindi alla nostra descrizione sfruttando al meglio una "organizzazione della conoscenza" che, grazie anche a importanti progetti comunitari e non solo, comincia a trattare sistematicamente l'informazione e l'universo documentario in modo tematico e disciplinare, attivando rigorose procedure di selezione, di valutazione e di controllo di qualità delle risorse (e qui il rinvio al progetto DESIRE è d'obbligo), utilizzando tassonomie e modellizzazioni controllate del sapere, creando sistemi esperti di classificazione e di indicizzazione automatici e adattandone il vocabolario e la terminologia per l'uso di internet.

Questo trattamento ha un duplice obiettivo: la creazione di un web semantico e l'ottimizzazione per l'utente del rapporto tra tempo impiegato nella ricerca e qualità dei risultati; cioè, da una parte, la modellizzazione dell'universo della conoscenza attraverso formalizzazioni logiche e topologiche (è il dominio scientifico di applicazione delle cosiddette "ontologie formali"), e dall'altra l'utilizzo di quelle rappresentazioni concettuali per migliorare la *performance* della ricerca.

Dunque, la nostra descrizione non è orfana; si inserisce in un contesto di sempre maggiore qualità degli strumenti di accesso all'informazione e di varietà delle competenze professionali necessarie al suo trattamento. Proprio per questo motivo risulta pressante, a nostro giudizio, l'esigenza di strutturare la relativa area prospettando due finalità: la descrizione del contenuto della risorsa; la contestualizzazione, sia pure concettuale, di quel contenuto nei domini del sapere.

La prima finalità si raggiunge, come è noto, con la redazione di manuali di stile necessari a dare uniformità linguistica e formale, e coerenza logico-argomentativa alle descrizioni; la seconda, con l'utilizzo di autorevoli strumenti di sussidio (dizionari specialistici, glossari, tesauri, repertori tassonomici) che garantiscono, fuori da ogni ambiguità, imprecisione e aleatorietà terminologica, la correlazione del contenuto descritto in un dominio della conoscenza.

Determinante risulta a tal fine il ruolo che potrebbero svolgere, come in parte già svolgono, le biblioteche accademiche: la disponibilità di specifiche competenze scientifiche e culturali e l'opportunità di cooperazioni a livello tematico e disciplinare, costituiscono di fatto un luogo elettivo di elaborazione, che deve tener conto sia dei nuovi modelli catalografici come FRBR, che dà maggiore enfasi al contenuto e alle relazioni tra entità, sia dei metadati descrittivi, sia ancora di quei modelli di metadati come RDF o di quei linguaggi come XML, necessari per definire una semantica della descrizione e per consentire una ulteriore, straordinaria via di accesso ai documenti.

Un'area descrittiva linguisticamente e logicamente strutturata secondo nuove modalità di rappresentazione del contenuto informativo è non solo più coerente con i dati veicolati dagli altri campi del record bibliografico, ma lo è ancora di più, come si è provato a dire, nel contesto di una diversa modalità di organizzazione del sapere e di accesso alle sue fonti, come può essere la creazione dei *Virtual Reference Desk* o dei *Subject Gateways*. E' coerente, poi, con le politiche di selezione dei documenti, che stanno a monte, e di sviluppo delle collezioni e con l'attività di indicizzazione e classificazione, cogliendo così il valore aggiunto della rete, quella di essere costituita appunto da un sistema reticolare di relazioni, in cui vengono enfatizzati i legami tra i documenti.

A noi tocca appunto il compito di saperli descrivere.

Mi piace infine ricordare, a conclusione di questo intervento, che i concetti su cui più ho insistito si richiamano intenzionalmente alle sei parole chiave che, nelle sue *Lezioni americane*, Italo Calvino ha lasciato in eredità per il terzo millennio: leggerezza, rapidità, esattezza, visibilità, molteplicità, consistenza.