

L'«Archivio fotografico dell'Arcidiocesi di Gorizia» Presentazione del sistema di recupero informativo

Promosso dall'Istituto di Storia Sociale e Religiosa, in collaborazione con l'Arcidiocesi di Gorizia, il progetto ha l'obiettivo di censire, catalogare e rendere consultabile la documentazione fotografica conservata non solo presso archivi parrocchiali, ma anche istituzioni e privati.

Nella fase iniziale, sono stati ideati e predisposti gli strumenti teorici e operativi: in buona sostanza, un modello di dati con relativo tracciato descrittivo e un software capace di gestire contestualmente l'attività di catalogazione nonché le specifiche esigenze connesse alla divulgazione e alla valorizzazione dei documenti fotografici.

Nella primavera di quest'anno è stato condotto un primo intervento catalografico che ha interessato dapprima il materiale posseduto dall'Istituto, poi alcune raccolte ecclesiastiche e private del Goriziano: sono stati catalogati e scansionati circa 800 documenti; da questo corpus è stato tratto il campione documentale che sostanzia il prototipo di sistema informativo.

I motori di ricerca insegnano

Il Web pare soffrire di un morbo paradossale: più cresce il magma informativo online, maggiore è la probabilità che la documentazione rilevante non venga recuperata: sia in modalità di ricerca testuale, semplice o avanzata (pur con tutti gli operatori di contesto e i filtri), che in modalità simil-classificata; riguardo ai reali interessi di recupero informativo da parte dell'utente il richiamo documentale resta sempre molto alto e la pertinenza molto bassa: tutto questo è dovuto alla scarsa presenza di elementi semantici tanto nella fase d'indicizzazione che in quella di ricerca. Il recupero documentale si basa sulla presenza o meno delle parole chiave utilizzate come criterio di ricerca e sull'individuazione dei documenti più affini alla query formulata; il potere di un algoritmo che valuta la messe di citazioni ricevute da ogni documento è in grado di valutarne la qualità fino a un certo punto, perché non basta la valutazione quantitativa della qualità di chi segnala a determinare la rilevanza del documento recuperato.

Il cosiddetto «Web semantico» (*Semantic Web*) sembra offrire un qualche rimedio: dati strutturati e significativi che mettano i processi automatizzati in grado di abbassare il richiamo, aumentando la pertinenza e la rilevanza delle risorse informative recuperate. Il ricorso ai metadati – nome di conio relativamente recente ma concetto ben conosciuto nell'ambito delle scienze bibliografiche – e ai linguaggi d'indicizzazione controllati rappresentano forse il metodo più praticato per conferire quella struttura e quel significato atti a migliorare il richiamo e la precisione delle risposte ottenute.

In ogni caso, l'uso di questa metodologia impone l'assunzione di rigide convenzioni non solo per la semantica, ma anche per la sintassi e la struttura. Un metadato è un attributo o, anche, una proprietà, un aspetto chiamato a descrivere la cosiddetta risorsa elettronica; il linguaggio controllato dovrebbe poi garantire l'uso di una terminologia condivisa da parte di chi descrive (indicizzatore) e di chi cerca e consulta (utente) grazie a liste, soggetti e tesauri per l'appunto comuni, sia a livello di post- che di preordinazione – per quanto recenti indagini quantitative e qualitative abbiano avanzato alcuni dubbi in proposito.

Da ultime ma non per importanza le questioni dell'interoperabilità semantica (*semantic interoperability*) e dell'usabilità (*usability*): la prima si pone non appena fuori dell'ambito bibliografico normato dai vari standard ISBD o per l'impiego non ancora soddisfacente di *de facto* standard come Dublin Core Metadata Initiative; la seconda riguarda invece l'approccio a e il rapporto con l'utente: in buona sostanza, interfacce d'interrogazione e visualizzazione dei set di risultati.

La fotografia mediante l'immagine

Gli archivi consultabili sono due: quello reale presso l'Istituto; l'altro, ideale, sotto forma d'immagini digitali con descrizioni memorizzate in una base di dati XML. Il modello descrittivo è costituito da un insieme di metadati relativi per così dire al «contenente» (oggetto, autore, editore, materia, tecnica, ecc.) e da un set attinente al contenuto (soggetto), articolato secondo l'impostazione divenuta ormai classica che consiste in una ripresa o, meglio, in un adattamento della teoria dei «livelli di significato» formulata da Erwin Panofsky: *ofness* (oggetto = quanto oggettivamente si vede) e *aboutness* (soggetto = interpretazione di quanto si vede). A questo proposito sono stati previsti due livelli descrittivi:

- 1) soggetto-descrizione per la descrizione oggettiva
- 2) soggetto-identificazione per l'interpretazione soggettiva.

L'esame di qualche esempio aiuta a chiarire la differenza fra i due livelli. Dal punto di vista dell'*ofness*, nella fig. 1



fig. 1

è raffigurato un interno con colonna e pavimento a mosaico, dove sono un uomo con vesti liturgiche che legge un libro, mentre alla sua sinistra un ragazzo con vesti liturgiche tiene in mano una candela accesa; davanti a loro un uomo regge in braccio un neonato adagiato su un cuscino con accanto due donne, un uomo in abito civile e uno in divisa militare. Dal punto di vista dell'*aboutness*, si tratta di un battesimo tenuto ad Aquileia nel 1915.



fig. 2

Nella fig. 2 osserviamo soldati a cavallo che sfilano lungo una strada urbana con soldati schierati ai lati; dietro vi sono edifici semidistrutti (*ofness*). Dal punto di vista dell'*aboutness*, si tratta di un momento della visita compiuta da Carlo d'Asburgo a Gorizia nel 1917.

Un'altra differenza fra i due livelli consiste nella diversa convenzione adottata nella strutturazione dei relativi dati: a «soggetto-descrizione» è associata un'indicizzazione di tipo libero, nel senso che la descrizione prevede l'utilizzo di parole-chiave non controllate strutturate in frasi, mentre in «soggetto-identificazione» viene impiegata un'indicizzazione controllata mediante descrittori presi da una lista terminologica controllata redatta durante la fase di catalogazione e strutturati in stringhe. Nei due casi presi in esame, le rispettive schede presentano le stringhe seguenti:

- 1) Aquileia – Battesimo – 1915
- 2) Gorizia – Carlo d'Asburgo – Visita – 1917

Utenti ed esigenze informative

Gli interessi e le curiosità dell'utenza di un qualsiasi sistema informativo possono riguardare gli argomenti più vari, e le richieste essere formulate con minore o maggiore precisione; chi progetta e implementa il sistema può esprimere al riguardo congetture e previsioni praticamente infinite.

Nel caso in questione, un utente potrebbe essere ad esempio interessato alle immagini di una certa località; un altro alle immagini di quella località in un determinato anno; un altro ancora a tutte le manifestazioni di carattere sociale o religioso svoltesi in quella località in un certo anno!

L'obiettivo di rispondere con la precisione migliore tanto alle richieste generiche quanto alle più specifiche implica l'assunzione di scelte adeguate a livello di classificazione dell'informazione prima e poi di consultazione della stessa. Di fronte per così dire alle devianze della «classificazione enumerativa» e della Classificazione Decimale Dewey in particolare, che offrono un «singolo atomo d'informazione, codificato da un numero», S.R. Ranganathan sostenne che non si doveva fare così e si poteva invece «classificare ogni aspetto dell'oggetto». Prendiamo in considerazione queste stringhe relative all'*aboutness* di altrettante immagini:

- 1) Cormons – Chiesa – Altare (Facciata, Navata, Presbiterio)
- 2) Cormons – Villa – Facciata (Parco, Salone d'onore, Scalone d'onore)
- 3) Gorizia – Seminario – Cappella (Dormitorio, Palestra, Refettorio)
- 4) Gorizia – Villa – Facciata (Parco, Salone d'onore, Scalone d'onore)

Classificando in modo classico-enumerativo, avremmo

- Cormons
 - Chiese
 - Altari
 - Facciate
 - Navate
 - Presbiteri
 - Ville
 - Facciate
 - Parchi
 - Saloni d'onore
 - Scaloni d'onore
- Gorizia
 - Seminari
 - Cappelle
 - Dormitori
 - Palestre

- Refettori
- Ville
 - Facciate
 - Parchi
 - Saloni d'onore
 - Scaloni d'onore

Se un utente è interessato a immagini di tipologie edilizie e loro parti per città, scorrendo un elenco classificato di questo tipo probabilmente sarà soddisfatto, perché la gerarchia è ordinata dapprima per città, poi per edificio e, infine, per parti di questi: proprio esattamente quanto serve all'utente. Ma se l'utente avesse un interesse diverso? Se fosse per esempio interessato direttamente agli edifici o alle parti di questi? Gli servirebbe una classificazione che partisse non più dal punto di vista del luogo, ma iniziasse a categorizzare da uno degli altri due punti di vista. Tuttavia, essendo la gerarchia articolata diversamente, dovrà comunque passare da ogni città per reperire i dati d'interesse.

L'esempio appena discusso illustra bene il fatto che fondere insieme in una sola gerarchia differenti tipi di dato (città, edifici, parti di edifici) non è l'impostazione adatta a soddisfare le più diverse esigenze informative.

Proviamo ora a considerare separatamente i vari tipi di dati:

- 1) Aquileia, Cormons e Gorizia sono nomi di città e, come altri nomi di luogo nel sistema, appartengono alla categoria «luoghi».
- 2) Battesimo è una cerimonia ed appartiene a questa categoria.
- 3) Cappelle, facciate, parchi, ecc. sono parti di o adiacenze di edifici e possono essere ricondotti a questa categoria.
- 4) Carlo d'Asburgo è una persona e in questa categoria può essere incluso.
- 5) 1915 e 1917 sono date e come tali possono essere categorizzate.

Si avrà dunque questa nuova formulazione delle gerarchie:

- Cerimonie
 - Battesimi
- Date
 - 1915
 - 1917
- Luoghi
 - Aquileia
 - Cormons
 - Gorizia
- Parti di edifici
 - Cappelle
 - Facciate
 - Parchi
- Persone
 - Carlo d'Asburgo

Da qui è possibile ricorrere a gerarchie più semplici e sintetiche, combinando le quali un utente potrà consultare per luoghi e edifici, per edifici e luoghi, per cerimonie e date, per persone e date e così via. Non solo: si raggiunge anche l'obiettivo di non atomizzare l'informazione e di evidenziare ogni aspetto relativo ad un «oggetto» – come auspicato da Ranganathan inventore del concetto di faccetta (*facet*).

Nonostante i continui miglioramenti, la ricerca online o, meglio, le modalità secondo le quali viene proposta continuano a porre questioni di usabilità con significative implicazioni nella progettazione dei sistemi informativi: uno dei problemi maggiori è senza dubbio rappresentato proprio dall'architettura informativa delle interfacce di consultazione.

L'interfaccia di consultazione

Offre tre modalità di consultazione/interrogazione: la prima (fig. 3) riprende l'impostazione appena descritta offrendo per ora otto categorie (faccette) relative all'*aboutness* e, dunque, al livello «soggetto-identificazione» – una volta testato il prototipo verranno implementate anche altre categorie relative al «contenente» e sarà sviluppata la gerarchia interna dove necessario: per ora le categorie sono a sviluppo monodimensionale.

INTERFACCIA DI CONSULTAZIONE

RICERCA LIBERA, LISTE TERMINOLOGICHE E CATEGORIE

I visitatori collegati sono in questo momento 2

[Note sul sistema](#) [Note sulla catalogazione](#)

[Suggerimenti, osservazioni e commenti](#)

CATEGORIE DI RICERCA

[Associazioni](#) [Cerimonie](#) [Date](#) [Edifici](#) [Luoghi](#) [Nomi](#) [Parti di edifici](#) [Persone](#)

fig. 3

Cliccando su ogni categoria, subito sotto appare l'elenco degli argomenti relativi a questa (fig. 4): si tratta delle voci di soggetto della faccetta «Luoghi».



fig. 4

Cliccando poi su un argomento vengono visualizzate le immagini relative. A questo riguardo possono verificarsi due situazioni: set di risultati fino a nove immagini, set di oltre dieci immagini – l'ordinamento è sempre cronologico. Nel primo caso (fig. 5), l'utente esaminerà direttamente le immagini recuperate.

Risultati della ricerca

Luoghi > Capriva del Friuli (Centa)

Le immagini sono 9 in ordine cronologico

Numero dell'immagine Valore dell'ingrandimento



fig. 5

Nel secondo caso, l'utente avrà una risposta «paginata» ovvero di dieci immagini alla volta e, cosa più importante, apparirà sul lato destro un nuovo elenco di categorie con relative voci di soggetto: l'argomento già scelto potrà essere combinato o, meglio, raffinato con ogni voce di soggetto evidenziata (fig. 6).

Luoghi > Cormons

Le immagini sono 67 in ordine cronologico

Pagina 1 di 7

Immagini 1 - 10 di 67

Numero dell'immagine Valore dell'ingrandimento

Raffinamenti per Associazioni, Cerimonie, Date, Edifici, Nomi, Parti di edifici e Persone

[Benedizioni](#) [Commemorazioni](#)
[Congressi eucaristici](#) [Incoronazione della Madonna](#) [Processioni](#)

1898 1899 1907 1910 1913
1915 1917 1920-1925 1920-1930
1922 1924 1930-1942 1931
1932 1933 1935 1936 1937
1940-1942 1942 1945-1950 1950-1960
1960 1955-1958 1960-1970 1961-1971

[Case](#) [Chiese](#) [Conventi](#) [Santuari mariani](#) [Ville](#)

[Corpus Domini](#) [Cumano-Perusini \(villa\)](#) [Madonna del Rosario](#) [Rosa Mistica \(chiesa\)](#) [Rosa Mistica \(santuario mariano\)](#) [San Giovanni Battista](#) [San Leopoldo](#) [San Quirino](#) [Santa Maria Sant'Adalberto](#) [Sant'Antonio da Padova](#) [Tomadoni \(casa\)](#)

[Altari](#) [Facciata](#) [Navate](#) [Presbiteri](#)

fig. 6

Sarà possibile restringere il set di 67 immagini relative a Cormons ad esempio con la voce «Chiese» della categoria «Edifici»: anche in questo caso l'insieme d'immagini recuperato potrà essere ulteriormente raffinato perché i documenti recuperati sono più di nove (fig. 7).

Luoghi > Cormons raffinato con Chiese
 Le immagini sono 13 in ordine cronologico

Numero dell'immagine Valore dell'ingrandimento

1907	1910	1913	1940-1942
1942	1945-1950	1950-1960	1955-1958
1960-1970			

San Giovanni Battista [San Leopoldo](#) [San Quirino](#) [Santa Maria](#) [Sant'Adalberto](#) [Sant'Antonio da Padova](#)
 Facciate [Navate](#) [Presbiteri](#)

fig. 7

La possibilità di restringere o raffinare il set d'immagini recuperato verrà proposta ogni volta che questo superi il limite prefissato.

La pagina sulla quale avvengono le successive visualizzazioni continua a proporre nella parte alta l'elenco di categorie: in ogni momento, la ricerca può dunque ripartire dalla stessa categoria – cliccando nuovamente su di essa – ma da voce diversa oppure da un'altra categoria con relativo elenco di voci.

Il limite numerico che determina l'assenza o la presenza dei raffinamenti, vale a dire la possibilità o meno di combinare fra loro voci di soggetto provenienti da categorie diverse è del tutto provvisorio: l'hanno segnalato test condotti su analoghe esperienze online; il *feedback* e riscontri d'altra natura chiariranno se è adeguato o andrà rivisto.

L'utenza dell'approccio a faccette

L'uso della classificazione a faccette è assai diffuso sul Web: molti Web designer sono giunti a questa soluzione per via indipendente, del tutto ignari delle teorie di Ranganathan e del successivo lavoro del Classification Research Group. Forse perché – come è stato notato – si tratta di un metodo naturale di organizzare le «cose» o, meglio, d'un approccio dettato dal buon senso.

Al di là di ogni considerazione generale o tecnica, l'immediatezza, l'intuitività e la libertà offerte dall'approccio risultano evidenti dagli esempi appena presentati e discussi. A ben vedere, l'utente che «naviga» in questo modo attraverso la raccolta ad ogni passaggio successivo costruisce implicitamente una richiesta (*query*) basata su AND altrettanto implicito: anche questo modo soft, amichevole di ricorrere al primo gradino della logica associativa possono spiegarne la duttilità.

Si tratta di un approccio adatto a tutti, ma in particolare all'utenza non-esperta ovvero non-professionale, dove per non-esperto o non-professionale non significa digiuno d'Internet e di motori di ricerca, ma non tecnico, non studente o studioso o, più semplicemente, non in possesso di interessi specifici verso la documentazione e la relativa informazione.

Non-esperta o non-professionale è pertanto la ricerca che muove da interessi generali e da richieste generiche, che ben poco frutto e magra soddisfazione porterebbe sul piano di una *query* formulata mediante parole-chiave.

La ricerca per parole-chiave

L'utente esperto o professionale è dunque mosso per varie ragioni da interessi maggiormente specifici e centrati; e solitamente pone richieste precise al sistema. Ad esso è rivolta la terza modalità di ricerca, chiamata «Ricerca libera» sull'interfaccia di consultazione – modulo che presenta un primo abbozzo di ricerca avanzata.

Anche qui la ricerca avviene per ora sui soli attributi di contenuto (*ofness* e *aboutness*) e mediante un confronto parziale *case-insensitive* (insensibile a maiuscole e minuscole); è possibile visualizzare direttamente le schede oppure una sequenza di miniatura dell'immagine con relativa frase di *ofness* o stringa di *aboutness* linkate alle rispettive schede. Ambedue gli elenchi sono debitamente paginati e con tutte le informazioni del caso.

Liste terminologiche

In mezzo l'ultima modalità di ricerca: le liste terminologiche. Si tratta di due elenchi in ordine alfabetico dei termini liberi di *ofness* e dei descrittori controllati di *aboutness*. Si tratta del corpus terminologico preliminare sulla base del quale è stata avviata la stesura di un tesoro integrato di termini generici e descrittori specifici opportunamente relazionati e categorizzati.

La finalità delle liste è molteplice: 1) rappresentano uno strumento di consultazione e di ricerca diretta, con una visualizzazione dei risultati che è quella proposta dalla modalità a faccette senza tuttavia i raffinamenti; 2) dovrebbero comunicare qualcosa sulle differenze fra indicizzazione libera e controllata; e 3) più in generale dovrebbero costituire una sorta di cerniera fra la ricerca a faccette e quella libera, e viceversa.

Forse non tutto è stato detto, ma solo di funzioni o di utilità che si presenteranno da sé. Ancora una volta è bene ribadire che, essendo a livello di prototipo, il sistema presenta incoerenze e oscillazioni che saranno superate dopo un opportuno «rodaggio» e un adeguato *feedback*. Quanto agli inevitabili errori, se non perdonati vengano almeno compresi, considerato il lavoro di quasi un anno portato avanti non da un team ma da due sole persone.*

Andrea Cuna
Univ. di Udine
andrecuna2002@libero.it

In attesa di riprendere con ben altro metro gli argomenti accennati in questa presentazione, mi limito a qualche indicazione bibliografica di puro orientamento. Rassegne sull'indicizzazione delle immagini: A.A. Goodrum, *Image Information Retrieval: An Overview of Current Research*, <http://www.informingscience.org/Articles/Vol3/v3n2p63-66.pdf>; M. De Marsicoi, L. Cinque e S. Levalde, *Indexing Pictorial Documents by their Content: A Survey of Current Techniques*, «Image & Vision Computing», 15, 1997, 119-141; K. Vézina, *Survol du monde de l'indexation des images*, <http://www.ebsi.umontreal.ca/cursus/vol4no1/vezina.htm>. – CDD, metadati, vocabolari controllati: L. Mai Chan, *Exploiting LCSH, LCC and DDC to Retrieve Networked Resources: Issues and Challenges*, http://www.loc.gov/catdir/bibcontrol/chan_paper.html; L. Dempsey e R. Heery, *A Review of Metadata: a Survey of Current Resource Description Formats*, <http://www.ukoln.ac.uk/metadata/desire/overview> e T. Koch, *Controlled vocabularies, Thesauri and Classification Systems*, <http://www.lub.lu.se/metadata/subject-help>. – Utenti: C. Jörgensen, *Image Indexing: An Analysis of Selected Classification Systems in Relation to Image Attributes Named by Naïve Users*, http://www.oclc.org/oclc/research/publications/review99/Jorgensen_e_T.L._Wells-Angerer,_A_Study_of_Retrieval_Success_with_Original_Works_of_Art_Comparing_the_Subject_Index_Terms_Provided_by_Experts_in_Art_Museums_with_those_Provided_by_Novice_and_Intermediate_Indexers, <http://etd.ils.unc.edu/dspace/bitstream/1901/135/1/tammywellsangerer.pdf>. – *Ofness* e *aboutness*: S. Shatford Layne, *Analysing the Subject of a Picture: A Theoretical Approach*, «Cataloging & Classification Quarterly», 6, 1986, 39-62 e anche E. Svenonius, *Access to Nonbook Materials: The Limits of Subject Indexing for Visual and Aural Languages*, «Journal of the American Society for Information Science», 45, 1994, 600-606. – Interoperabilità: O. Signore, *Tecnologie XML: un elemento essenziale per l'interoperabilità*, <http://www.w3c.it/papers/webdav2003-interoperability.pdf>. – Teoria della *facet analysis* (Ranganathan e Classification Research Group – CRG): L. Spiteri, *A Simplified Model for Facet Analysis*, «Canadian Journal of Information and Library Science», 23, 1998, 1-30 e http://iainstitute.org/pg/a_simplified_model_for_facet_analysis.php. Restano

fondamentali per tutto il contesto qui affrontato I due contributi di W. Denton, *Putting Facets on the Web: An Annotated Bibliography*, <http://www.miskatonic.org/library/facet-biblio.html> e *How to Make a Faceted Classification and Put It On the Web*, <http://www.miskatonic.org/library/facet-web-howto.html>.