



MultiMedia Information Retrieval: stato dell'arte e prospettive di applicazione

Roberto Raieli*, Perla Innocenti**

* Università degli Studi "Roma Tre", Roma, raieli@uniroma3.it

** Politecnico di Milano, Milano, perla.innocenti@polimi.it

Parole chiave: Multimedia information retrieval - Content-based information retrieval - Visual retrieval - Video retrieval - Audio retrieval - Digital library - ICT - Sistemi informativi - Intelligenza artificiale - Interazione uomo-computer - Image processing - Computer vision - Audio processing

Abstract: Il volume promosso da AIDA riguarda lo stato dell'arte del multimedia information retrieval (MMIR): un innovativo sistema di information retrieval content-based per il trattamento dell'informazione digitale multimediale, applicabile nelle basi di dati più avanzate grazie agli strumenti offerti dalle tecnologie odierne nei processi informativi e comunicativi. Nel volume vengono delineati il panorama evolutivo delle nuove problematiche del multimedia information retrieval e le aree di ricerca più significative, presentando metodi ed esperienze internazionali in diversi campi di applicazione. Scopo della pubblicazione è quello di contribuire alla diffusione in Italia dell'applicazione del sistema, in vista del suo impiego da parte di enti di ricerca, istituzioni e organizzazioni nel settore della documentazione, e di offrire un ulteriore approfondimento a quanti si occupano di sistemi informativi, image e audio processing, computer vision, intelligenza artificiale e interazione uomo-computer.

MultiMedia Information Retrieval: state of the art and application perspectives. The article presents a volume on multimedia information retrieval (MMIR), promoted by the Italian Association of Advanced Documentation (AIDA). The volume outlines the state of the art of MMIR, an innovative system of content-based information retrieval to process digital information in multimedia databases. It offers a panorama of multimedia information retrieval issues and challenging research areas and a collections of chapters by international specialists in the computer science and information retrieval community. The aim of the book is to contribute to dissemination of MMIR in Italy, envisioning its potential applications by public and private organizations in a variety of fields within the documentation area. For technically inclined readers dealing with information systems, artificial intelligence, image and audio processing, computer vision, human-computer interaction and high performance computing, the book will provide a deeping lecture on familiar topics.

L'esigenza di creare un volume che faccia il punto sullo stato dell'arte del *multimedia information retrieval* nasce dalla constatazione della sempre maggiore inevitabilità di un processo di rivoluzione dei sistemi di trattamento dell'informazione digitale, necessitata dalla già avvenuta rivoluzione dei documenti digitali e dei processi informativi e comunicativi.

Si è oramai reso necessario, infatti, tracciare un panorama evolutivo organico della problematica, nonché presentare e discutere in modo complessivo la teoria di base di ognuna delle possibili realizzazioni ed applicazioni, data la rarità nel contesto internazionale di trattazioni complete e organiche dell'argomento e l'assoluta mancanza in Italia di una pubblicazione di tale genere. Questa occasione, dunque, ci auguriamo possa essere utile in particolare al lancio e allo sviluppo di un dibattito tutto italiano sul *multimedia information retrieval* e i suoi sottoinsiemi specifici.

Fino a non molto tempo fa l'*information retrieval* si è prospettato come un dominio privilegiato dei bibliotecari, dei documentalisti e degli specialisti dell'informazione, che avevano adottato un'interpretazione *user-centered* focalizzando l'attenzione sui modi testuali con cui gli utenti interpretano

e utilizzano l'informazione, in opposizione alle modalità automatiche di strutturazione, immagazzinamento e recupero dei dati. Ma l'IR si è notevolmente evoluto negli ultimi anni, parallelamente alle sfide poste dall'espansione del Web e dallo sviluppo delle tecnologie dell'informazione e della comunicazione [1]. Il delinearsi di nuove modalità di comunicazione e condivisione delle informazioni ha dato luogo a quella che è stata definita la "convergenza al digitale" [2].

La centralità dei documenti multimediali nell'odierna società dell'informazione e gli strumenti offerti dalle tecnologie digitali hanno favorito la creazione di basi di dati multimediali, di ben altro livello di complessità rispetto alle basi di dati tradizionali, sviluppando pienamente le premesse del concetto di *digital library* [3]. Le *multimedia digital library* sono caratterizzate da una gestione e un accesso integrati per documenti tipologicamente eterogenei, attraverso l'impiego di specifici sistemi per l'indicizzazione, la ricerca e l'estrazione automatica di dati rappresentativi del complesso contenuto dei documenti multimediali (immagini, video, audio), che si vanno ad affiancare ai tradizionali sistemi manuali di analisi e indicizzazione terminologica dei documenti testuali o audiovisivi.

I moderni sistemi hanno perciò raccolto una serie di nuove sfide nell'accesso all'informazione, dato l'alto livello di interattività offerto dalle tecnologie di rete e a fronte del sempre maggiore incremento della domanda di un accesso efficace e veloce. Poiché la qualità del recupero delle informazioni è largamente influenzata dall'interazione dell'utente con il sistema, anche nei confronti dell'utente molto deve cambiare, e il sistema di approccio ai database multimediali deve essere riformulato sulla base delle più complesse esigenze di definire la *query* con dati *visivi* e *sonori* e non soltanto con dati *terminologici*. Alle tradizionali interfacce dei database testuali, che consentono la ricerca in un indice composto esclusivamente di termini estratti dai documenti o inseriti in metadati testuali, si devono quindi sostituire interfacce che permettano di formulare le *query* in diverse dimensioni, non solo tramite i termini ma anche attraverso le immagini e i suoni: consentendo così la ricerca in indici composti da testi estratti dalle didascalie o dal parlato, da immagini chiave di una sequenza, da semplici figure, da melodie, da forme, colori e suoni, senza con ciò escludere l'importanza che continuano a mantenere i dati testuali, descrittivi o classificatori, relativi ad aspetti non specificamente audiovisivi del documento.

Il crescente utilizzo dell'IR in ambito commerciale e scientifico ha stimolato un preciso interesse da parte della *computer science*, che a differenza della biblioteconomia e della documentazione ha affrontato tali tematiche con una prospettiva focalizzata sugli algoritmi e sulle tecniche di elaborazione dei dati. Da un'ottica *computer-centered* il problema consiste nella costruzione di indici efficienti, nell'elaborazione di *query* con un alto livello di *performance* e nello sviluppo di algoritmi di *ranking* che migliorino la qualità della risposta. Le ricerche si sono perciò estese alla modellizzazione dei dati, alla categorizzazione dei documenti, alle architetture di sistema, alle interfacce interattive, ai processi di *filtering* e alle *multimedial query* [4].

Seguendo questa tendenza, ai tradizionali sistemi di indicizzazione e ricerca *term-based*, comunque basati sull'uso di termini descrittivi, devono sostituirsi progrediti sistemi di archiviazione e recupero *content-based*, dove i "descrittori" saranno dei veri *metadati*, caso per caso della stessa natura dei dati cui si riferiscono, e di cui possono consentire nel modo più funzionale l'analisi e la ricerca.

È dunque limitativo continuare ad impostare la ricerca informativa nei termini di un generico *information retrieval*, riportando ogni tipo di ricerca documentale alle condizioni di una ricerca tramite linguaggio testuale. È invece necessario considerare un più ampio criterio di *multimedia information retrieval* (MMIR), dove ogni genere di documento digitale venga trattato, immagazzinato e richiamato tramite gli elementi di linguaggio, o di *metalinguaggio*, propri dello specifico oggetto digitale.

Il distacco principale tra i sistemi di archiviazione e recupero di documenti testuali e quelli di documenti multimediali si focalizza nel sistema di analisi ed estrazione degli elementi indicatori del contenuto del documento e dei descrittori specifici delle sue caratteristiche. Se i sistemi di indicizzazione e ricerca tradizionali sono basati sui termini, sulla logica delle *parole chiave* - che ci si propone naturalmente, data la nostra abitudine ad esprimere e rappresentare verbalmente e testualmente qualunque oggetto di conoscenza e informazione - i sistemi di recupero più innovativi - che riflettono l'evoluzione di una

cultura sempre più visuale, nonché multimediale - richiedono invece una riflessione più avanzata sulle caratteristiche intrinseche dei file digitali, sulla semantica e sulla strutturazione degli elementi che compongono un documento, sulle modalità di interazione e restituzione dei risultati delle *query* all'utente.

In generale si possono quindi attualmente distinguere:

- sistemi di *information retrieval term-based*, basati su informazioni testuali (termini estratti dal linguaggio naturale, schemi di classificazione e soggettazione, thesauri), per la ricerca di documenti testuali (e finora spesso anche di documenti audiovisivi);
- sistemi di *information retrieval content-based* o *multimedia information retrieval*:
 1. sistemi di *visual retrieval*, in cui i file d'immagine 2D e 3D sono cercati e recuperati tramite dati visivi interni al file, quali ad esempio colore, *texture* e *pattern*, forma, orientamento e distribuzione spaziale, ecc.;
 2. sistemi di *video retrieval*, dove per il recupero di documenti audiovisivi si utilizza il linguaggio audiovisivo, cioè elementi di ricerca ricavati dalle immagini del filmato, dal movimento degli oggetti nelle inquadrature, dall'analisi degli stacchi di montaggio o della traccia sonora;
 3. sistemi di *audio retrieval*, nei quali l'informazione sonora è ricercata in misure di suoni, ricavando quindi i dati di *query* dall'analisi dei volumi, delle sonorità, dei ritmi o delle melodie.

Il senso e il carattere radicalmente rivoluzionario dei sistemi di MMIR consistono dunque nel fatto che, se negli archivi dove il contenuto dei documenti è di tipo testuale appare appropriato che le chiavi che ne consentono l'accesso siano descrittori testuali estratti *dall'interno* del documento, negli archivi multimediali si rivela invece semplificato e impreciso attribuire *dall'esterno* una descrizione testuale a contenuti che si fondano su un diverso regime di senso. Tuttavia, per raggiungere un buon livello di precisione nel recupero dei documenti da una base di dati multimediale, sembra comunque auspicabile la compresenza dei sistemi di *information retrieval term-based* e *content-based*: l'interrogazione *term-based* può costituire un ottimo metodo preliminare per selezionare una parte della grande quantità di documenti di un archivio, e per centrare la ricerca in base a dati quali gli ambiti d'appartenenza, le tipologie, le classi, i titoli, gli autori; successivamente può essere un sistema di ripulitura finale dall'inevitabile rumore specifico di un'interrogazione *content-based*. In tutto ciò i due procedimenti possono operare sempre in armonia e in interazione costante, in un'unica interfaccia utente.

Negli ultimi anni il *multimedia information retrieval* è stato oggetto di un numero crescente di ricerche e applicazioni a livello internazionale, attestando l'interesse di varie comunità d'uso che si occupano della gestione di documenti multimediali per le nuove possibilità di valorizzazione di archivi e attività professionali: dalle applicazioni biochimiche (ad esempio per l'indicizzazione delle molecole) all'architettura e all'archeologia (per la ricerca di costruzioni, siti e piante in base alle similarità nei particolari), dagli archivi di film e di video al giornalismo (dove vengono notevolmente potenziate le ricerche di personaggi e situazioni dispersi in migliaia di minuti di girato), dai beni culturali al design, dalla didattica all'*e-commerce* e al settore dello spettacolo, dai GIS ai sistemi di *remote sensing*, dal settore medico (giacché le diagnosi vengono spesso effettuate servendosi di ricognizioni audiovisuali di condizioni anormali) ai sistemi di identificazione e di sorveglianza.

A fronte di questo interesse e delle potenzialità dell'utilizzo di tali sistemi in ambiti documentali ad elevato contenuto specialistico, AIDA ha promosso un volume dal titolo *MultiMedia Information Retrieval. Metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione* [5], curato da chi scrive e finalizzato alla diffusione ed alla promozione di tali metodologie e delle loro applicazioni.

Il volume è diretto innanzitutto a quanti si occupano di documentazione e in generale delle *library & information sciences*, ponendosi come un'introduzione *per exempla* al MMIR attraverso metodi, ricerche ed esperienze sviluppati in ambito internazionale. Ma poiché il MMIR si colloca al crocevia di molteplici ambiti che includono i sistemi informativi, l'intelligenza artificiale, l'*image* e l'*audio processing*, la

statistica, la *computer vision*, l'*high performance computing*, l'interazione uomo-computer, il libro può rappresentare un'occasione di ulteriori approfondimenti da parte di tutti quanti si occupano di tali settori.

Il volume è articolato in un capitolo introduttivo alle tematiche del *multimedia information retrieval*, scritto dai curatori, e due sezioni complementari sullo stato dell'arte dei metodi e delle esperienze più significativi in questo ambito, con contributi di ricercatori e studiosi che approfondiscono alcune delle principali tematiche teoriche e applicative.

La **prima sezione** del volume inquadra le iniziative condotte negli ultimi anni in Italia e in Europa. È presente un saggio fondamentale di Peter Enser, della University of Brighton, fondatore del gruppo VIR [Visual Information Retrieval Research] [6], tra i primi e più attenti teorici europei delle possibili applicazioni del *visual retrieval* alla documentazione. Molto importante il contributo italiano di Alessandro Micarelli [7], Marco Anelli ed Enver Sanginetto, dell'Università degli Studi "Roma Tre", che approfondiscono la tecnica pratica del *visual retrieval* ed illustrano il metodo specifico da loro messo a punto. Nell'ambito del *video retrieval*, Pasquale Savino, Giuseppe Amato e Claudio Gennaro, dell'ISTI-CNR [Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - Consiglio Nazionale delle Ricerche] di Pisa, trattano del sistema ECHO [European CHronicles On-line] [8], da loro creato per la dettagliata analisi, archiviazione e ricerca di video e filmati di ogni complessità. Tra i maggiori studiosi italiani del *visual* e del *video retrieval*, Raimondo Schettini [9], dell'Università degli Studi di Milano-Bicocca e dell'ITIM-CNR [Istituto per le Tecnologie Informatiche Multimediali - Consiglio Nazionale delle Ricerche] di Milano, esamina le teorie e le pratiche più avanzate del sistema. Un panorama delle prospettive di *audio retrieval* unitamente alla discussione delle proprie sperimentazioni e applicazioni viene presentato da Andreas Ribbrock e Frank Kurth, dell'Università di Bonn. Infine, Jürgen Herre e Markus Cremer, del Fraunhofer IIS [Institut für Integrierte Schaltungen] tedesco, discutono le concrete applicazioni di *audio retrieval* del sistema AudioID [10], da loro messo a punto.

La **seconda sezione** è dedicata al ricco e variegato panorama extraeuropeo. Eric Paquet, del National Research Council canadese, membro del gruppo Visual Information Technology [11], illustra nel suo contributo un nuovo metodo per la descrizione di oggetti multidimensionali applicabile nell'ambito del *visual retrieval*. Per il settore del *video retrieval* un saggio ricco di riflessioni ed esperienze è quello di John R. Smith [12], dell'IBM T. J. Watson Research Center, creatore di sistemi come VisualSEEk e WebSEEk, che tratta del *multi-source video clustering* tramite modelli semantici. Di grande rilevanza nello stesso settore il contributo di Nevenka Dimitrova [13], Lalitha Agnihotri e Mauro Barbieri della Philips Research, relativamente alle tecniche di *video summarization*. Ancora nell'ambito del *video retrieval* è il saggio di Susan Tennant e Jeffrey Huang, della Indiana University-Purdue University Indianapolis (IUPUI), in merito alle esperienze del progetto CLIOH [Cultural Digital Library Indexing Our Heritage] [14]. E un ulteriore approfondimento del settore si offrirà nel saggio sulle applicazioni del formato Annodex di Silvia Pfeiffer e Claudia Schremmer [15] della Common Scientific & Industrial Research Organization (CSIRO). Il recupero delle informazioni sonore viene affrontato in ambito musicale da George Tzanetakis, della University of Victoria [16], creatore di innovativi sistemi di *audio retrieval* e interessanti interfacce grafiche. Conclude la sezione extraeuropea una panoramica delle ricerche e le applicazioni di *audio retrieval*, presentata da Michael Casey [17], della City University di Londra, già ricercatore presso i Mitsubishi Electric Research Laboratories (MERL).

Se, parafrasando Michael Lew, «digital audiovisual information is here to stay» [18], poiché essa permea ormai ogni aspetto della nostra vita, i sistemi di *multimedia information retrieval* rappresentano certamente una delle aree di ricerca e sviluppo più promettenti del futuro prossimo venturo, che troveranno efficace applicazione in molteplici settori e attività. La strada è dunque aperta anche per l'Italia...

Note

1 - Cfr. Ricardo Baeza-Yates - Berthier Ribero-Neto, *Modern information retrieval*. New York : Addison-Wesley Longman, 2000 e David A. Grossman - Ophir Frieder, *Information retrieval. Algorithms and heuristics*. Boston : Kluwer Academic Publishers, 1998.

- 2 - Fabio Ciotti - Gino Roncaglia, *Il mondo digitale. Introduzione ai nuovi media*. Roma; Bari : Laterza, 2000, p. 348.
- 3 - Cfr. la famosa definizione utilizzata nel *Workshop on distributed knowledge environment* di Santa Fé del 1997, secondo cui una biblioteca digitale «is not merely equivalent to a digitized collection with information management tools. It is rather an environment to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge» www.si.umich.edu/SantaFe.
- 4 - Cfr. William I. Grosky, *Managing multimedia information in database systems*. "Communications of the ACM", 40 (1997), n. 12, p. 73-80.
- 5 - *MultiMedia Information Retrieval. Metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione*, a cura di Roberto Raieli e Perla Innocenti. Roma : AIDA, 2004.
- 6 - Homepage del gruppo di *Visual Information Retrieval Research* at the University of Brighton: www.cmis.brighton.ac.uk/Research/vir/VIR1.HTM.
- 7 - Homepage di Micarelli a: www.dia.uniroma3.it/~ia2/materialeDidattico.html#materialeDidattico.
- 8 - Sito Web del sistema ECHO a: pc-erato2.iei.pi.cnr.it/echo.
- 9 - Homepage di Schettini a: www.itim.mi.cnr.it/Staff/Schettini.
- 10 - Sito Web di AudioID a: www.emt.iis.fhg.de/produkte/audioid.
- 11 - Sito Web del gruppo di *Visual Information Technology* di Paquet: www.iit-iti.nrc-cnrc.gc.ca; www.cleopatra.nrc.ca/Nefertiti/EricPaquet/index.html.
- 12 - Homepage di Smith a: www.ctr.columbia.edu/~jrsmith.
- 13 - Homepage di Dimitrova a: www.ctr.columbia.edu/~nevenka.
- 14 - Sito Web del progetto CLIOH: www.ctr.columbia.edu/~nevenka; homepage di Huang a: www.cs.iupui.edu/~huang.
- 15 - Homepage di Pfeiffer a: www.silvia-pfeiffer.de/index.htm.
- 16 - Homepage di Tzanetakis a: www.cs.uvic.ca/~gtzan.
- 17 - Homepage di Casey a: www.merl.com/people/casey.
- 18 - Michael S. Lew (ed.), *Principles of visual information retrieval*. London : Springer-Verlag, 2001.

Il volume *MultiMedia Information Retrieval. Metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione* è disponibile per prenotazioni a condizioni vantaggiose sul sito AIDA www.aidaweb.it/cgibuonodordine.html.
