

WINTER SCHOOL  
INTELLIGENZA ARTIFICIALE NEI BENI CULTURALI

WINTER SCHOOL  
INTELLIGENZA ARTIFICIALE NEI BENI CULTURALI  
*A cura di S. Bandini, L. Bordoni, G. Mantegari*

2007 ENEA  
Ente per le Nuove tecnologie  
l'Energia e l'Ambiente

Lungotevere Thaon di Revel, 76  
00196 Roma

ISBN 88-8286-187-2



WINTER SCHOOL  
INTELLIGENZA ARTIFICIALE  
NEI BENI CULTURALI

A CURA DI

S. BANDINI, L. BORDONI, G. MANTEGARI

## *Ringraziamenti*

La realizzazione della Winter School è stata possibile grazie al supporto di:

- CSAI – Complex Systems and Artificial Intelligence Research Center – Università di Milano – Bicocca
- Laboratorio Nomadis – Università di Milano – Bicocca
- AI\*IA – Associazione Italiana per l'Intelligenza Artificiale

Hanno offerto il proprio patrocinio:

- DISCo – Dipartimento di Informatica, Sistemistica e Comunicazione – Università di Milano – Bicocca
- ENEA – Ente per le Nuove tecnologie, l'Energia e l'Ambiente
- Provincia di Milano – Progetto Monza e Brianza
- Università degli Studi di Bologna – Facoltà di Conservazione dei Beni Culturali

Desideriamo inoltre ringraziare chi, a vario titolo, ha contribuito al successo della Winter School.

In particolare:

Andrea Bonomi  
Bernardo Rondelli  
Sebastian Stride  
Giuseppe Vizzari

## Indice

Prefazione	7
Cultural Resources Management: prospettive e contributi dell'Intelligenza Artificiale <i>Stefania Bandini, Bernardo Rondelli, Giuseppe Vizzari</i>	11
Un framework per l'accesso intelligente all'informazione <i>Marco de Gemmis, Giovanni Semeraro</i>	25
Analisi multimediale e indicizzazione semantica: l'esperienza di PrestoSpace <i>Roberto Basili</i>	45
Visual Information Retrieval. Indicizzazione e reperimento di dati pittorici <i>Gianluigi Ciocca, Raimondo Schettini</i>	65
OMERO. Un sistema per permettere di "toccare" la realtà virtuale <i>Fabio De Felice, Floriana Renna, Giovanni Attolico, Arcangelo Distante</i>	83
IA e musei: l'allestimento automatico <i>Francesco Amigoni, Viola Schiaffonati</i>	103
αρωγή'. Un laboratorio distribuito su base geografica per la ricomposizione virtuale assistita di frammenti <i>Nicola Mosca, Giovanna Carlomagno, Floriana Renna, Giovanni Attolico, Arcangelo Distante</i>	119
L'Intelligenza Artificiale in Archeologia <i>Juan Antonio Barceló</i>	137



## PREFAZIONE

La Winter School “Intelligenza Artificiale nei Beni Culturali”, svoltasi presso l’Università degli Studi di Milano - Bicocca dal 5 al 9 febbraio 2007, ha costituito un’occasione preziosa e generosa d’incontro per tutti i partecipanti.

L’evento, organizzato dal Centro di Ricerca CSAI (Complex Systems & Artificial Intelligence Research Center - Università di Milano - Bicocca; <http://www.csai.disco.unimib.it/>) e ospitato dal Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo), è stato coordinato da *Stefania Bandini* (CSAI), *Luciana Bordonì* (ENEA-FIM e coordinatore del Gruppo di lavoro “Intelligenza Artificiale nei Beni Culturali” dell’AI\*IA) e *Glauco Mantegari* (CSAI e programma di dottorato e ricerca avanzata QUA\_SI - Laboratorio Nomadis).

Gli Atti della Winter School intendono rappresentare una testimonianza sullo stato dei lavori e sui risultati finora conseguiti in quel percorso di ricerca che muove dalla necessità di indagare nel variegato e complesso ambito dei beni storico-artistici, archeologici, librari e della catalogazione, con l’obiettivo di mostrare come alcune tra le più significative metodologie e tecniche informatiche dell’Intelligenza Artificiale rappresentino strumenti ormai insostituibili per l’analisi, la conservazione, la fruizione, il restauro e la tutela dei Beni Culturali.

Conformemente all’articolazione della Scuola, gli Atti sono organizzati in tre principali sezioni tematiche a forte coerenza interna e organiche fra loro. La prima sezione riguarda il supporto alla documentazione ed al *retrieval* “intelligente” di documenti in forma digitale; la seconda, l’accessibilità alle risorse culturali; la terza, il supporto alle attività di ricerca di esperti del settore dei Beni Culturali.

Il contributo di *Stefania Bandini*, *Bernardo Rondelli* e di *Giuseppe Vizzari* sulle prospettive e sui contributi dell’Intelligenza Artificiale nel contesto del Cultural Resource Management introduce alla prima sezione che si apre con il contributo di *Giovanni Semeraro* e di *Marco de Gemmis* (Università di Bari).

Gli autori propongono un’ampia esposizione e discussione dei concetti fondamentali, dei modelli formali e delle tecniche alla base dei sistemi per il reperimento automatico di documenti in forma digitale. In particolare viene presa in esame l’interazione tra le tecniche di base ed i metodi di apprendimento automatico in grado di introdurre intelligenza nell’accesso all’informazione. Vengono inoltre presentati sistemi che permettono di operare il filtraggio personalizzato dell’informazione, la costruzione automatica di profili utente, la ricerca semantica, la disambiguazione del linguaggio naturale e la classificazione automatica di documenti. A tale proposito viene presentato un applicativo realizzato per la *3rd International Semantic Web Conference*, che prevede l’implementazione di un *Conference Participant Advisor*, in grado, per esempio, di creare un programma personalizzato sulla base del profilo dell’utente.

La sezione comprende inoltre il contributo di *Roberto Basili* (Università di Roma - Tor Vergata) che affronta il tema della semantica dei testi e del *machine learning*, con particolare riferimento ai Beni Culturali.

Viene presentata, attraverso l'esperienza del progetto europeo "PrestoSpace", l'applicazione di tecnologie del linguaggio nell'estrazione automatica di metadati semantici da archivi multimediali. Tale progetto è dedicato alla digitalizzazione, preservazione e indicizzazione semantica dei grandi archivi audiovisivi europei. Tecnologie quali il riconoscimento automatico del parlato, l'analisi delle immagini e l'estrazione automatica di informazioni convergono in "PrestoSpace" all'automazione del processo di documentazione dei dati contenuti negli archivi multimediali.

*Gianluigi Ciocca e Raimondo Schettini* (Università di Milano - Bicocca) presentano un'introduzione, di ampia prospettiva, sui *Visual Information Systems*, affrontando, in particolare, il problema della ricerca di immagini per contenuto all'interno di un database di immagini. In questo ambito vengono discussi anche gli aspetti concernenti i meccanismi per il raffinamento delle interrogazioni pittoriche, la classificazione semantica e l'annotazione automatica delle immagini. Viene inoltre presentato "QuickLook", un sistema per il *retrieval* di contenuti multimediali attraverso l'impiego, per esempio, di immagini campione e di schizzi tracciabili dall'utente, con la possibilità di combinare, con queste modalità, ricerche su descrizioni testuali.

Nella seconda sezione degli Atti *Fabio De Felice, Floriana Renna, Giovanni Attolico, Arcangelo Distante* (ISSIA CNR, Bari) prendono in esame le possibilità offerte dalle tecnologie legate alla realtà virtuale e all'interazione evoluta uomo-macchina nello specifico ambito delle applicazioni per i Beni Culturali. Il contributo propone l'illustrazione e la discussione del progetto "Omero", che indaga le possibilità di fruizione di informazioni da parte di persone con disabilità visive, siano esse ipovedenti o non vedenti. Tale fruizione è resa possibile dall'impiego di metodologie e di tecnologie di interazione di tipo aptico/acustico, in grado di fornire agli utenti innovative possibilità di esplorazione di ambienti e oggetti virtuali.

*Francesco Amigoni e Viola Schiaffonati* (Politecnico di Milano) delineano alcune possibili relazioni fra Intelligenza Artificiale e musei analizzando, in particolare, l'ambito dell'allestimento automatico di musei virtuali. Vengono presentati i recenti sviluppi del sistema "Minerva", un progetto dedicato all'allestimento automatico di musei virtuali attraverso l'impiego di agenti intelligenti; viene proposta inoltre, una discussione sul concetto stesso di museo virtuale, con particolare attenzione alle possibilità da questo offerte nell'estensione delle attuali prospettive museali.

La terza sezione ospita il contributo di *Nicola Mosca, Giovanna Carlomagno, Floriana Renna, Giovanni Attolico e Arcangelo Distante* dedicato alla tematica del restauro virtuale, inteso come strumento di supporto agli operatori coinvolti in operazioni di restauro reale. In particolare si presenta il caso della ricomposizione di frammenti sulla vela del San Matteo di Cimabue nella basilica superiore di S. Francesco in Assisi, che ha permesso di attuare una significativa sperimentazione del sistema realizzato su un caso applicativo particolarmente rilevante.

Il contributo finale di *Juan A. Barceló*, prestigioso docente presso la *Universitat Autònoma de Barcelona* (Spagna), fornisce il punto di vista e le esperienze dell'autore su metodi e strumenti di Intelligenza Artificiale per il supporto alle attività di ricerca archeologica, con particolare riguardo all'impiego di Reti Neurali Artificiali.

Barceló articola il suo contributo in due sezioni principali. La prima riguarda la presentazione dei fondamenti epistemologici alla base dello studio archeologico, contestualizzando, in questo ambito, i concetti di problema, di *problem solving* e di *machine learning*, e introducendo la distinzione tra classificazione e *clustering*. La seconda propone un'introduzione alle Reti Neurali Artificiali, sottolineando quali possano essere i vantaggi del loro impiego nella ricerca archeologica.

Concludendo, possiamo affermare che la raccolta di tali contributi conferma il raggiungimento degli obiettivi che la Winter School si era proposta, riuscendo anche a definire uno spazio di approfondimento interdisciplinare piuttosto significativo all'interno del panorama nazionale. Sebbene sia evidente l'interesse sempre più spiccato nei confronti delle tematiche legate al rapporto tra nuove tecnologie e Beni Culturali, la ricerca fatica ancora a definirne con precisione i termini e le caratteristiche, così come a coinvolgere i soggetti che di queste ricerche dovrebbero essere i protagonisti attivi, sia nell'ambito umanistico sia in quello informatico. Quanto realizzato nell'ambito della Scuola costituisce pertanto un'esperienza da valorizzare ulteriormente, cercando di marcare con un segno di continuità il lavoro intrapreso nell'ottica di una reale prospettiva interdisciplinare.

*Stefania Bandini  
Luciana Bordoni  
Glauco Mantegari*



# CULTURAL RESOURCES MANAGEMENT: PROSPETTIVE E CONTRIBUTI DELL'INTELLIGENZA ARTIFICIALE

*Stefania Bandini, Bernardo Rondelli, Giuseppe Vizzari*

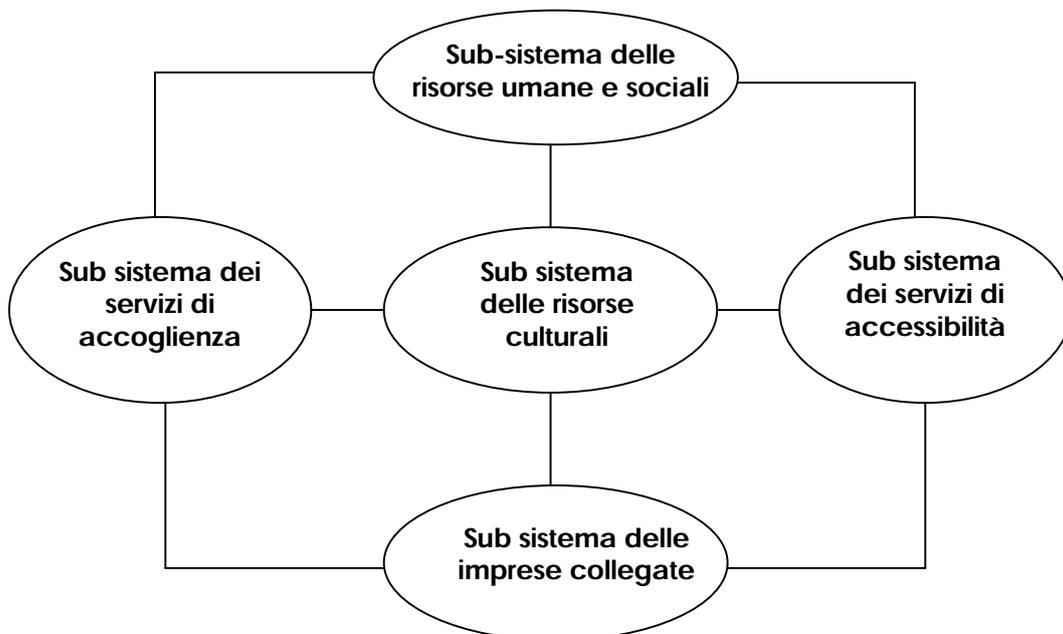
Complex Systems and Artificial Intelligence (CSAI) research center  
Dipartimento di Informatica Sistemistica e Comunicazione (DISCo)  
Università degli studi di Milano-Bicocca  
Viale Sarca, 336 - 20126 Milano, Italy

bandini@disco.unimib.it, rondelli@disco.unimib.it,  
vizzari@disco.unimib.it  
<http://www.csai.disco.unimib.it/>

## 1 Introduzione

La tradizione degli studi sui beni culturali (BBCC) è storicamente legata all'ambito materiale, abbracciando le discipline che intorno ad essa gravitano, facendo del "bene" l'oggetto stesso della loro indagine, quale espressione di contesti storici, sociali e culturali. L'importante riflessione, nata principalmente in seno all'archeologia e all'antropologia negli ultimi 20 anni, ha portato a riconsiderare l'oggetto materiale e ad allargare la riflessione all'ambito immateriale (aspetti simbolici, rituali ecc.) e all'ambiente-paesaggio.

Parallelamente si è avviata una transizione dal concetto di bene culturale (BC), quale insieme di "frammenti del passato" da preservare, al concetto di "risorsa" culturale [1], assumendo rilievo per il ruolo che essa svolge o può svolgere nel processo di sviluppo territoriale, ossia integrandosi e supportando la valorizzazione del sistema delle risorse locali in cui si colloca.



**Figura 1 - Integrazione in un sistema territoriale: lo schema è tratto da Valentino, P.A., *Le trame del territorio. Politiche di sviluppo dei sistemi territoriali e distretti culturali*, Milano, 2003**

Questa proposta ha aperto la strada ad un confronto più ampio nell'ambito della stessa ricerca applicata ai BBCC, e, più in generale, nel contesto del *Cultural Resource Management* (CRM)<sup>1</sup>.

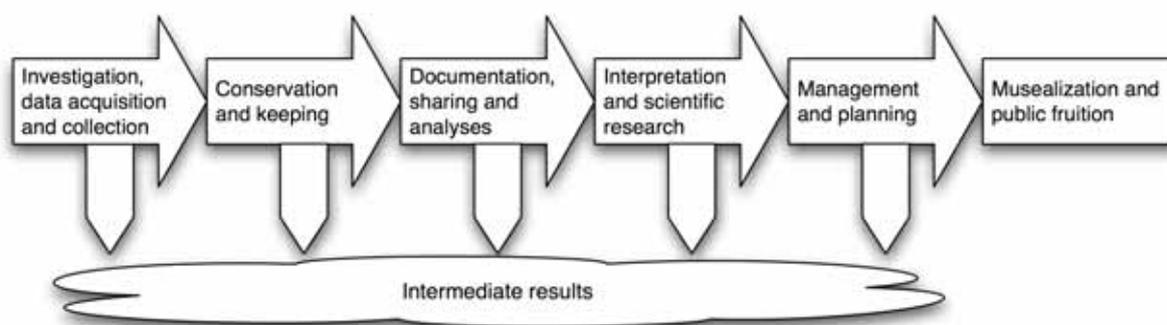
Il contesto del CRM ha assunto, sempre più, caratteristiche e connotati multidisciplinari, catturando l'attenzione e la riflessione scientifica di diversi contesti disciplinari su metodologie di indagine, valutazione, analisi e conservazione del contesto culturale in oggetto. Da sistema di controllo preventivo, costituito di organismi di tutela e codifica legislativa, il CRM si è trasformato in un proficuo ambiente di ricerca, in cui diversi studiosi mirano ad occuparsi del problema della gestione dei BBCC, come analisi, valorizzazione e fruizione della "conoscenza passata" in sistemi territoriali integrati<sup>2</sup>.

## 2 Applicazioni Informatiche nel contesto dei Beni Culturali

In questo contesto, l'impatto delle nuove tecnologie proprie della *Computer Science* (CS) hanno determinato e continuano a rappresentare un contributo fondamentale.

L'ampiezza dello spettro delle applicazioni informatiche nel contesto dei BBCC rende impraticabile la realizzazione di una panoramica esaustiva. L'obiettivo di questa sezione è fornire uno schema di riferimento nel quale collocare questi lavori e indicare alcuni riferimenti significativi per supportare il lettore desideroso di approfondire questi temi.

Le applicazioni informatiche, infatti, abbracciano tutto il "flusso" della gestione dei BBCC, dal supporto all'acquisizione, documentazione e archiviazione dei dati, all'aspetto analitico e interpretativo, alla pianificazione e fruizione.



**Figura 2 - Flusso delle attività di gestione dei BBCC, con evidenziati i risultati intermedi e la fase finale di musealizzazione e/o fruizione da parte del vasto pubblico**

<sup>1</sup> Alla fine degli anni '70 è stato coniato il termine di risorsa culturale, come espressione contrapposta al concetto di "risorse naturali", con l'intento iniziale di regolare la tutela e valorizzazione delle espressioni materiali della "cultura" passata, poi esteso anche al patrimonio immateriale, fino alla ratifica della *Convenzione per la salvaguardia del patrimonio culturale immateriale* (Parigi, 2003 - <http://www.unesco.org/culture/ich/index.php>).

<sup>2</sup> Attraverso una strategia che integri i processi di valorizzazione delle risorse e del territorio, privilegiando la realizzazione di programmi allargati di azione a discapito di singoli interventi puntuali, irrobustendo le interdipendenze tra attività di ricerca, industria culturale e gli altri settori produttivi presenti sul territorio.

Solo per citare alcuni strumenti informatici che hanno profondamente influito sulle pratiche degli operatori in contesto gestione di BBCC possiamo indicare:

- i sistemi di gestione di basi di dati e le tecnologie dei sistemi informativi hanno supportato la realizzazione di grandi *repository* di informazioni anche eterogenee, per l'archiviazione, memorizzazione ed accesso organizzato ad informazioni inerenti i BBCC;
- il mondo del web e le tecnologie Internet correlate possono supportare la realizzazione di un vasto spettro di modalità di accesso personalizzato all'informazione per differenti tipi di utenti, per finalità che vanno dall'interpretazione e ricerca scientifica fino alla fruizione da parte del vasto pubblico;
- la natura intrinsecamente spaziale delle risorse culturali non può essere ignorata e i sistemi informativi geografici (GIS) rappresentano un supporto adeguato alla rappresentazione, memorizzazione e gestione (incluse differenti modalità di analisi) di dati georeferenziati;
- le tecnologie della mobilità delineano nuove possibilità tanto per gli operatori che lavorano sul campo, nelle fasi di acquisizione dati, conservazione e documentazione, quanto per innovative modalità di fruizione.

Due contesti in particolare che sembrano oggi catturare l'attenzione di molti ricercatori nell'ambito della CS nel contesto dei BC sono in particolare<sup>3</sup>:

- le Digital Library
- i Musei Virtuali e i Musei 'Aumentati'.

Oggi, infatti, gli archivi, le biblioteche e i musei di tutta Europa affrontano sfide simili, nel cercare di trarre vantaggio dall'enorme potenziale che l'uso delle tecnologie dell'informazione e della comunicazione promette alle istituzioni della memoria. Queste sfide non sono di mera natura tecnologica, ma coinvolgono il cuore stesso delle istituzioni per la conservazione dei beni culturali.

Essere digitali, per molti archivi, biblioteche e musei europei non è più una possibilità, ma una necessità. Tali istituzioni si sono trasformate in "ibridi" (nel caso dei musei si può parlare di realtà "aumentate", ovvero ambienti fisici dotati di estensioni virtuali finalizzate a migliorare l'esperienza del visitatore) che si prendono cura tanto delle risorse culturali analogiche quanto di quelle digitali. La conversione di tutte le tipologie di contenuti culturali in *bits and bytes* apre una dimensione completamente nuova per chi voglia raggiungere tanto il pubblico tradizionale quanto un pubblico nuovo, offrendo l'accesso a risorse culturali secondo modalità inimmaginabili un decennio fa.

In tal direzione, oltre alle specifiche realizzazioni applicative e all'utilizzo di modelli e tecnologie, si deve registrare, anche nell'ambito dei BBCC, la crescente diffusione di progetti di ricerca finalizzati alla realizzazione di portali, network di ricerca e strumenti per il lavoro collaborativo e la condivisione di dati e informazioni.

---

<sup>3</sup> In relazione a questi temi, il volume include alcuni contributi significativi, come si vedrà in seguito.

Tra questi si possono segnalare alcuni tra i riferimenti più accreditati in ambito internazionale:

1. EPOCH (<http://www.epoch-net.org/>): European Research Network on Excellence in Processing Open Cultural Heritage
2. ADS (<http://ads.ahds.ac.uk/>): Archaeological Data Service
3. Alexandria Archive (<http://www.alexandriaarchive.org/>)
4. ECAI (<http://ecai.org/>): Electronic Cultural Atlas Initiative
5. IDP (<http://idp.bl.uk/>): International Dunhuang Project.

Il proliferare di attività di ricerca nell'ambito dell'informatica nel contesto dei BBCC ha permesso anche la diffusione di diverse riviste specializzate, conferenze e workshop in contesto nazionale e internazionale.

Per una disamina abbastanza esaustiva cfr. <http://www.digicult.info/pages/resources.php>; <http://caa.leidenuniv.nl/index.htm>; <http://www.u.arizona.edu/~mlittler/>.

### **3 Applicazioni dell'Intelligenza Artificiale nel contesto del CRM**

I metodi, le tecnologie e le esperienze applicative della IA consentono oggi di affrontare la gestione del patrimonio culturale con approcci innovativi e con maggiore consapevolezza delle esigenze degli operatori di settore. Le possibilità offerte dai paradigmi computazionali per la rappresentazione e il trattamento della conoscenza permettono infatti di volgere l'attenzione non solo agli aspetti tecnologici per la fruizione, ma anche per il disegno di nuove metodiche per la formazione, la ricerca e la progettazione di servizi rivolti al pubblico. Inoltre, la valorizzazione delle risorse culturali di un territorio si esplica nella individuazione di corretti indirizzi e metodi per l'acquisizione e gestione della loro conoscenza, così da orientare la pianificazione verso lo sviluppo di potenzialità compatibili con il territorio stesso. In questa prospettiva, gli strumenti concettuali e computazionali dell'Intelligenza Artificiale impattano direttamente sul vasto settore del CRM, in termini di creazione di sistemi per la condivisione di informazioni e la loro rappresentazione per favorire una riflessione profonda e attuale sui concetti di cultura, identità e dinamica storica.

È infatti importante concepire il patrimonio culturale nella sua accezione materiale attraverso il reperimento e la gestione di un bene (storico, archeologico, paesaggistico, quale espressione tangibile di specifiche azioni e scelte umane) per poi curarne la valorizzazione in termini di interpretazione e rappresentazione, così da permettere una diffusione del suo significato come espressione di relazioni, come identità storica, sociale, antropologica, o nel suo più ampio valore di memoria collettiva.

La vastità dei temi, delle aree di intervento, delle finalità delle applicazioni dell'Intelligenza Artificiale nel solo settore dei BBCC rende impraticabile la realizzazione di una descrizione esaustiva dello stato dell'arte. Preferiamo in questa sede proporre uno schema che permetta di classificare i diversi tipi di contributo presenti in letteratura, ed in particolare ci intendiamo soffermare sui diversi tipi di 'utenti' di queste soluzioni.

Un rilevante numero di esperienze di applicazione di tecniche derivate dall'Intelligenza Artificiale è stato volto a migliorare l'esperienza di *fruizione* pubblica del bene culturale, con particolare riferimento alla realizzazione di *repository* dotati di funzionalità avanzate di accesso, ricerca e navigazione di informazioni e documenti.

L'iniziativa del cosiddetto *Semantic Web* vede nel contesto dei BBCC un'area di applicazione naturale, come anche testimoniato dalla organizzazione di eventi specifici<sup>4</sup>; va inoltre citato, sempre in quest'area, lo sforzo di realizzazione di ontologie di riferimento per la descrizione di beni culturali che ha portato alla definizione del CIDOC Conceptual Reference Model (CRM)<sup>5</sup>. In particolare uno degli obiettivi esplicitamente formulati dell'iniziativa è anche la realizzazione di un modello utile a mediare ed integrare differenti sorgenti di informazioni inerenti i BBCC, quali quelle pubblicate da musei, biblioteche ed archivi.

Il vasto pubblico non è però l'unico tipo di utente delle applicazioni di Intelligenza Artificiale in questo settore: la complessità di alcune delle attività svolte dagli operatori, dagli studiosi e dai ricercatori è tale infatti da presentare numerosi scenari nei quali andare a sperimentare tecniche volte appunto alla risoluzione (o al supporto alla risoluzione) di problemi che generalmente sono delegati ad esperti umani dotati di conoscenza e competenza di valore. L'esperienza della Winter School dell'anno 2007 ha cercato di selezionare dei rappresentanti significativi di questo genere di esperienze, cercando di promuovere un confronto tra le nuove tecnologie, a disposizione e a supporto dell'esperto e, parallelamente, proporre una riflessione sul concetto stesso di "gestione" del BC, in relazione alle considerazioni più profonde in seno alla stessa disciplina dell'IA. In particolare si è cercato di indirizzare la riflessione non tanto sull'aspetto di controllo e tutela legislativa del bene culturale, quanto sulla proposta di strumenti concettuali e analitici utili all'organizzazione e fruizione della conoscenza che intorno al bene culturale orbita.

#### **4 Prospettive e Linee di sviluppo**

Negli ultimi anni, il settore dei BBCC ha riscosso un'attenzione sempre maggiore anche presso il mondo politico [3], in virtù del suo potenziale economico e della sua importanza per lo sviluppo del mercato nella Società dell'Informazione. Anche all'interno dei governi nazionali e delle amministrazioni regionali, grandi sono le aspettative che le istituzioni per la conservazione dei beni culturali divengano soggetti attivi nella nascente economia dell'informazione<sup>6</sup>. Tuttavia, misurare i beni culturali in termini meramente economici non ne coglierebbe il reale valore<sup>7</sup>. Il reale valore che le istituzioni per la conservazione dei beni culturali mettono a disposizione della società è spesso indiretto e non economicamente misurabile, dal momento che esse cercano di offrire piacere intellettuale e di innalzare la consapevolezza dell'importanza della cultura e del sapere storico.

Le entrate aggiuntive o la capacità di creare entrate spesso si manifestano indirettamente, per altri settori economici, come per esempio lo sviluppo regionale, il turismo, l'editoria e i media.

---

<sup>4</sup> Si veda ad esempio il First International Workshop on Cultural Heritage on the Semantic Web, che si terrà in concomitanza dell'International Semantic Web Conference 2007 (<http://www.cs.vu.nl/~laroyo/CH-SW.html>).

<sup>5</sup> <http://cidoc.ics.forth.gr/>

<sup>6</sup> In accordo con le politiche regionali europee, che indicano tra le proprie linee guida *l'identità regionale* e le *nuove tecnologie dell'informazione* quali strumenti utili allo sviluppo sostenibile regionale (Regional Poloct-Infolegio).

<sup>7</sup> Come rileva Jim McGuigan [2], l'idea che un prodotto culturale sia valutabile secondo il suo prezzo di mercato, determinato dalle scelte del consumatore sovrano e dalla legge della domanda e dell'offerta, è ancora prevalente, sebbene profondamente incrinata. Il suo difetto fondamentale è la riduzione di ogni valore, la cui natura è invece manifestamente varia e opinabile, ad una logica economica e unidimensionale, la logica del 'libero mercato'.

In quanto principali soggetti erogatori di finanziamenti i governi nazionali e le amministrazioni regionali dovrebbero essere consapevoli del fatto che ciò che stanno finanziando va molto al di là del valore economico, e anzi è, per l'intera società, un caposaldo della creazione dell'identità culturale.

Sebbene le istituzioni per la conservazione dei beni culturali affrontino alti rischi in relazione alle incertezze sui rapidi mutamenti in campo tecnologico, assumere un atteggiamento attendista sarebbe da parte loro una strategia erranea. Le attuali metodologie di conservazione a lungo termine, come la conservazione tecnologica, la transmigrazione e l'emulazione, sono considerate insufficienti per conservare gli oggetti digitali a lungo termine: soluzioni a breve termine per un problema di lungo periodo, in particolare in merito a problemi relativi a supporti fisici e standard di formato.

Le istituzioni stesse, invece, dovrebbero sviluppare capitali e interventi solidi per la creazione e l'acquisizione di materiali digitali. In aggiunta, è necessario che le autorità a livello nazionale e regionale intraprendano azioni immediate e formulino strategie per la conservazione digitale intesa come parte integrante di una politica nazionale dell'informazione. È inoltre necessaria un'immediata politica di interventi in merito al sempre crescente volume dei materiali "nati digitali", cioè quelle risorse che sono state create con l'aiuto della tecnologia dell'informazione e che richiedono *software* e *hardware* particolari per poter essere consultate. L'esplosione del numero di materiali pubblicati elettronicamente esercita attualmente un'enorme pressione sulle istituzioni per la conservazione dei beni culturali, dal momento che esse mancano del quadro normativo che le autorizza a raccogliere, immagazzinare, rendere accessibili e conservare correttamente le risorse pubblicate in rete. Poiché molte risorse web svaniscono in un breve periodo di tempo, in assenza di una legislazione di questo tipo e di altri meccanismi che consentano alle istituzioni di raccogliere questi dati, una grande quantità del nostro patrimonio culturale del futuro andrà inevitabilmente perduta.

Alla luce di queste brevi e certamente non esaurienti considerazioni, si possono ora indicare alcune proposte, suggerite in particolar modo da una recente indagine di DigiCULT (2006) quale supporto alle azioni della Commissione Europea nel settore dei BBCC<sup>8</sup>:

#### Area 1: Navigazione e guida utente intelligenti

- Sviluppo di strumenti intelligenti e narrativi di contestualizzazione per i dati relativi ai beni culturali;
- Strumenti di analisi automatizzata dell'immagine per dati storico-pittorici.

#### Area 2: Automazione della digitalizzazione e archiviazione di massa;

- Casistica delle buone pratiche in materia di archiviazione di massa;
- Digitalizzazione automatizzata di massa;
- Acquisizione di metadati all'atto della digitalizzazione e integrazione con la gestione della collezione.

---

<sup>8</sup> Cfr. <http://www.digicult.info/pages/resources.php>.

Area 3: Conservazione a lungo termine di complesse risorse digitali e ricerca nel campo degli oggetti digitali dinamici:

- Strategie di conservazione a lungo termine per risorse digitali culturali complesse;
- Casistica sulle buone pratiche in materia di emulazione e di strategie di conservazione a lungo termine;
- Soluzioni per la conservazione di oggetti digitali dinamici;
- Nuovi approcci alla denominazione nel web e ulteriori sviluppi di sistemi permanenti di identificazione;
- Innalzamento della consapevolezza sulle questioni riguardanti la conservazione a lungo termine al di fuori della comunità dei beni culturali.

Area 4: Nuovi strumenti:

- Strumenti ad alta produttività per utenti non tecnici (per esempio, un sistema autore su *knowledge-base*);
- Interattività, tramite un'ampia gamma di interfacce uomo-macchina;
- Strumenti collaborativi capaci di supportare vari modelli;
- Sistemi intelligenti capaci di supportare gli utenti a vari livelli;
- Ricerca nel campo dell'uso di tecnologie avanzate per applicazioni sui beni culturali.

I beni culturali, pertanto, rappresentano un utile banco di prova per le attuali e future tecnologie di rappresentazione e di gestione della conoscenza.

In conclusione una sfida necessaria sarà quella di permettere la transizione da sistemi informativi e repository di dati e informazioni verso la condivisione di conoscenze utili a supportare:

- l'analisi e l'interpretazione dei contesti da parte degli esperti di settore;
- la pianificazione e la gestione del patrimonio culturale per trasformarlo realmente in sistemi integrati di valorizzazione delle risorse;
- la fruizione e la formazione del grande pubblico.

## **5 CSAI e CRM**

In conclusione del capitolo presentiamo brevemente due iniziative multidisciplinari, collocabili nell'area 4 del sopraccitato schema DigiCULT, che vedono la collaborazione del centro di ricerca CSAI e differenti realtà nazionali ed internazionali attive nel settore della gestione dei BBCC.

### 5.1 Progetto Po-Basyn (*Po Valley and Bronze Age Settlement dYNamics*)

Il progetto<sup>9</sup> è di natura fortemente multidisciplinare e vuole rappresentare una piattaforma di sperimentazione per soluzioni innovative che possano sfruttare nuovi strumenti concettuali, modelli, tecniche e tecnologie derivati dall'approccio scientifico a supporto della gestione delle risorse culturali, operando su due linee principali:

- studio delle dinamiche evolutive delle società antiche in chiave regionale;
- realizzazione di strumenti per la condivisione, la valorizzazione e la fruizione di informazioni e conoscenze sul patrimonio culturale.

La scelta del contesto in cui sperimentare questo approccio è lo studio del popolamento dell'età del Bronzo (2300 ca. – 900 ca. a.C.) nella Pianura Padana. Il contesto oggetto di studio offre un corpus di dati tali da permettere di sperimentare differenti modalità di analisi delle informazioni e di gestione della conoscenza; inoltre, ci poniamo nella prospettiva di contribuire ad un miglioramento delle ricerche stesse, cercando di proporre nuovi metodi per la lettura dei fenomeni del popolamento protostorico nella Pianura Padana.

In particolare, oggetto di indagine è la regione come un *sistema aperto*, cioè in relazione con l'esterno, per osservare gli effetti che il grado di apertura della regione sistemica produce. È interessante sottolineare come questo tipo di approccio consenta di lavorare su un piano multiscale: nel momento in cui consideriamo un sistema territoriale non lo consideriamo, infatti, 'isolato' dalla realtà ad esso circostante, bensì come parte integrante di un sistema più grande, sottoposto a processi che partono da attori interni ad esso o da attori ad esso esterni.

Tale approccio rappresenta una sfida quanto mai attuale per la rivalutazione dell'identità culturale in chiave regionale. L'identità culturale, infatti, non coincide semplicemente con la provenienza geografica, perché non può essere considerata come un'entità fissa e immutabile, ma va interpretata come effetto e risultante di processi storici ed è quindi, per sua natura, plurale e composita. Le culture hanno sempre un carattere relazionale in quanto si costituiscono, agiscono, "vivono" solo in relazione alla storia e al confronto con sistemi culturali differenti. Per questo motivo, il progetto si propone di studiare le dinamiche del popolamento come momento di intersezione tra diverse culture e di indagare quali effetti si siano originati da tali relazioni.

Il Progetto nasce e si sviluppa con l'intento di perseguire dei seguenti obiettivi:

#### 1. *Azioni di Networking*

- a. Realizzazione di un Network multidisciplinare di ricerca specificatamente orientato al raggiungimento degli obiettivi del progetto lungo le sue molteplici dimensioni.

---

<sup>9</sup> Il Progetto è il risultato di una collaborazione tra il Centro CSAI e il Centro CuDAM dell'Università di Milano Bicocca, il Dipartimento di Scienze dell'Antichità dell'Università Statale di Milano, il Dipartimento di Archeologia dell'Università di Bologna.

## 2. Condivisione, valorizzazione e fruizione della conoscenza

- a. Creazione di una base di conoscenza il più esaustiva possibile sulle testimonianze archeologiche dell'età del Bronzo attraverso la realizzazione di un progetto di *carta archeologica* basata su tecnologie per la gestione e la diffusione di informazioni geografica in Rete (*webGIS*)<sup>10</sup>. Le informazioni scientifiche disponibili per ciascun sito sono ad oggi estremamente disperse tra i vari gruppi di ricerca e di tutela e la loro qualità è piuttosto difforme. In questo contesto, uno degli obiettivi primari che ci si propone è quello di ottenere un sufficiente livello di uniformità tra le differenti esperienze, attraverso la definizione di approcci e metodologie condivise che possano trovare nel progetto il proprio punto ideale di aggregazione. Tale base di conoscenza potrà essere utilizzata a diversi livelli, dallo studio specialistico alla fruizione culturale in senso più allargato.

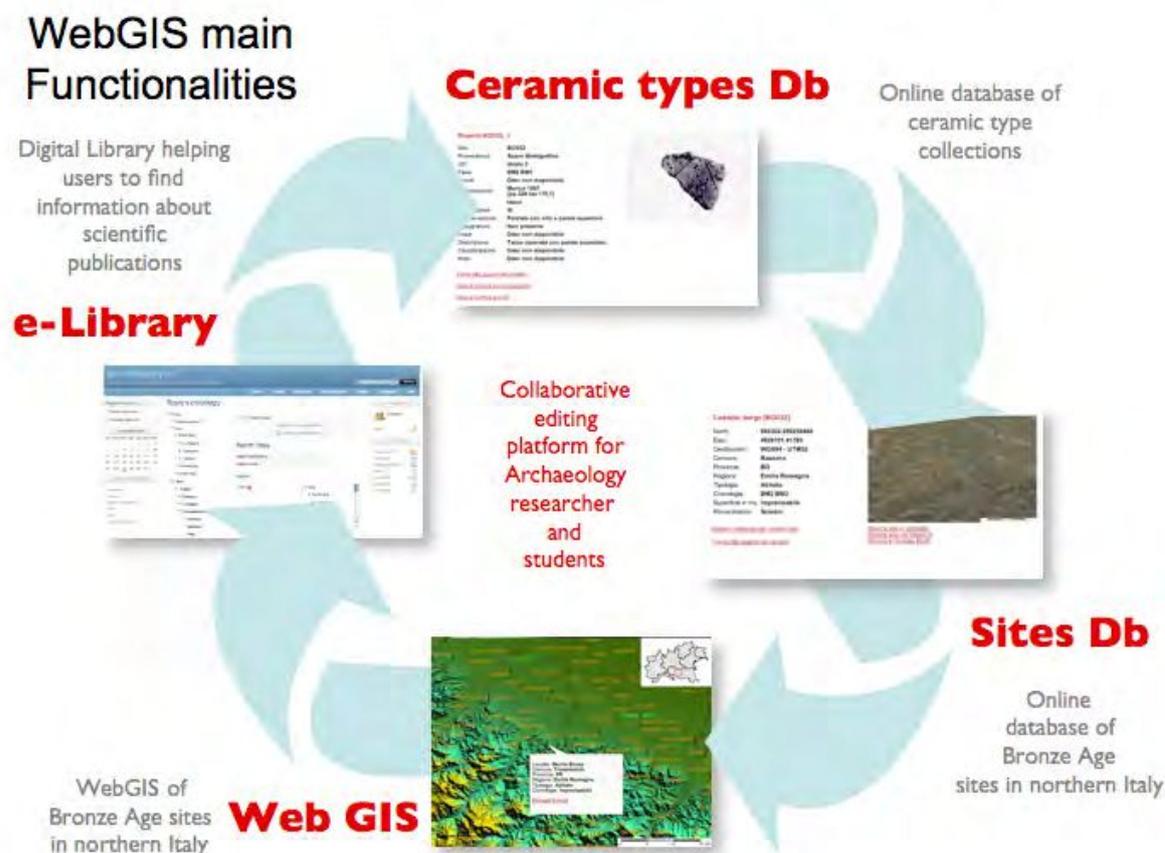


Figura 3 - Funzionalità previste per il sistema di organizzazione, gestione e condivisione di informazione e documenti riguardanti il patrimonio culturale oggetto di indagine

<sup>10</sup> Un'esperienza pilota nell'ambito delle applicazioni di tecnologie *web-based* per la condivisione di informazioni scientifiche è già da qualche tempo in atto nell'ambito del progetto del portale "ArcheoServer" ([www.archeoserver.it](http://www.archeoserver.it)); il portale, nonostante la sua recente attivazione, sta già dimostrando le proprie potenzialità che sono testimoniate dal costante incremento del numero di visite giornaliere e dal suo proficuo impiego da parte dei Dipartimenti di Archeologia delle Università coinvolte nel progetto [4][5].

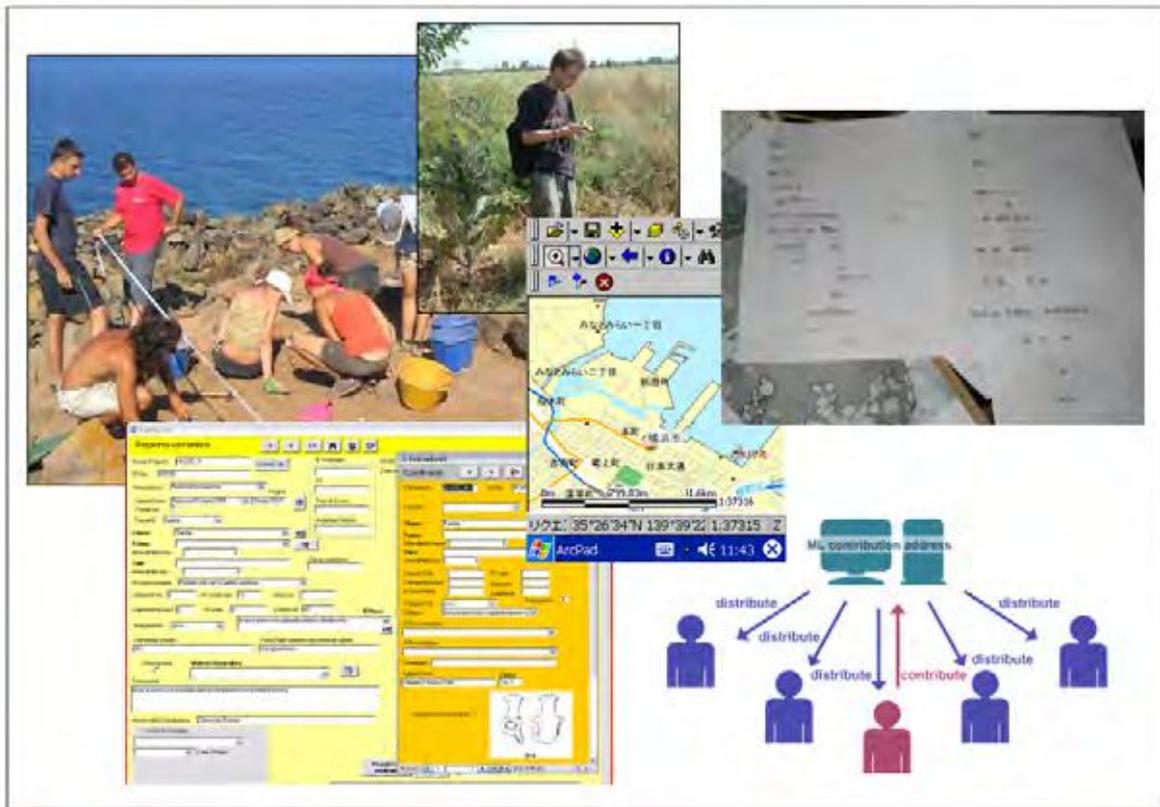
- b. Sviluppo di nuove piattaforme tecnologiche per la *condivisione della conoscenza* (“esperta”, oltre che “manualistica”) tra studiosi, ricercatori e studenti impegnati nell’investigazione di problematiche inerenti il progetto, che supportino funzionalità avanzate di visualizzazione, ricerca, selezione e aggiornamento dei dati in un ambiente aperto e collaborativo.
- c. Progettazione e implementazione di uno strumento “web-based” per la *gestione del Patrimonio Archeologico* che funga da supporto alla valorizzazione, alla pianificazione territoriale e alla fruizione di quest’ultimo da parte di un pubblico esperto e non.

### 3. *Rappresentazione, analisi e interpretazione*

- a. Definizione di nuovi metodi, modelli, strumenti di analisi e di rappresentazione formale dell’informazione e della conoscenza, volti a supportare la ricerca di una maggior comprensione delle problematiche inerenti le dinamiche evolutive del popolamento, con particolare riferimento al contesto specifico dell’età del Bronzo nella Pianura Padana.
- b. Applicazione di tecnologie di datazione assoluta per la valutazione di sequenze cronologiche, per le verifiche e per il miglioramento della comprensione del record archeologico.
- c. Studio del paesaggio per una analisi in chiave evolutiva delle dinamiche del popolamento, attraverso l’interpretazione della cultura materiale e della sua distribuzione.
- d. Approfondimento del concetto di “regione” nell’età del Bronzo. La nozione di regione sarà analizzata adottando una prospettiva sistemica finalizzata ad un accrescimento delle conoscenze circa l’identificazione degli attributi, delle estensioni, delle tendenze e delle traiettorie evolutive di quest’ultima.
- e. Approntamento di un laboratorio virtuale per la corroborazione e la falsificazione delle teorie e delle ipotesi formulate nel contesto multidisciplinare, mediante la definizione di opportuni modelli computazionali volti alla progettazione ed implementazione di sistemi di simulazione di dinamiche evolutive delle relazioni popolazione-risorse-territorio.

### 4. *Lavoro cooperativo, Logistica e Coordinamento*

Rilettura delle strategie di ricerca sul campo (scavo e analisi territoriale) e definizione di nuove modalità di indagine attraverso la valutazione progressiva dei risultati acquisiti nella attività di organizzazione sistematica delle informazioni scientifiche.



**Figura 4 - Situazioni, pratiche, documenti, schemi relativi alle attività svolte sul campo da un archeologo**

Il contributo di questo progetto di ricerca si colloca nell'ambito della gestione dei BBCC nella loro accezione più ampia (come sopra descritto), promuovendo lo studio di una realtà archeologica, quale espressione materiale di un contesto culturale, osservandone la sua natura composita e dinamica, al fine di proporre, da una parte, una maggior comprensione scientifica di tali fenomeni e, dall'altra, permetterne la valorizzazione e fruizione di tali risorse come espressioni culturali, in un sistema integrato, in relazione al territorio in cui si collocano.

### **5.2 Progetto SilkRoDE (*Silk Roads in the Digital Era*)**

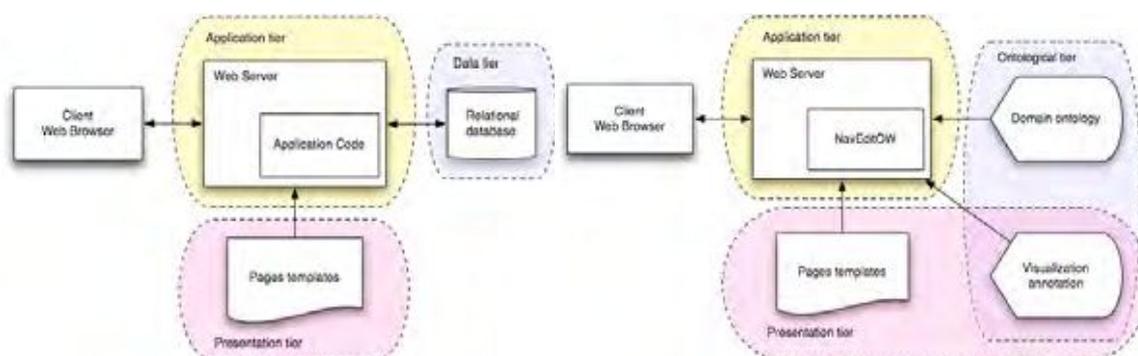
SilkRoDE nasce come progetto finalizzato alla raccolta, organizzazione, gestione e diffusione della conoscenza inerente i BBCC dell'Asia Centrale, coinvolgendo discipline quali l'archeologia, la storia, la geografia, la sociologia, la linguistica e l'etnografia. Il progetto, nato dalla collaborazione di oltre 30 istituzioni che in diverso modo operano sul contesto dell'Asia Centrale, intende realizzare una rete di competenze e conoscenze indirizzate su tre diversi tipi di utenti finali [5]:

- esperti: supporto all'analisi, interpretazione e pianificazione
- studiosi e studenti: formazione di competenze
- pubblico generico: superamento delle barriere culturali preconcette

Il Progetto si sviluppa con l'intento di perseguire i seguenti obiettivi:

- recupero e salvaguardia di tutte le importanti indagini e ricerche condotte nell'area durante l'Unione Sovietica e oggi fortemente a rischio di essere dimenticate e perdute;
- proposta di tecniche analitiche proprie dell'IA per lo studio dei fenomeni e processi storici e dell'evoluzione socio-culturale;
- creazione di un *Knowledge Management System*, grazie alla collaborazione di esperti di dominio, ed esperti di tecniche e modelli propri della CS e IA, operando nell'ambito:
  - della realizzazione di sistemi esperti di supporto all'acquisizione dei dati (*field activities*)
  - realizzazione di una *digital library* e di un WebGIS che permetta di raccogliere, diffondere e interrogare documenti d'archivio, pubblicazioni, cartografie ecc. (*document repository and management*)
  - esplicitare le competenze e le conoscenze implicite, proprie di chi opera nel settore (*knowledge acquisition and engineering*)
  - proporre strumenti di simulazione come laboratori sperimentali per lo studio dei fenomeni socio-economici, investigando il comportamento di sistemi, esplorando le diverse alternative possibili e mettendo alla prova le proprie alternative di scelta ed intuizioni (*data modeling and simulation*)
  - organizzazione e gestione delle relazioni tra studiosi, teams e istituzioni che operano nel contesto (*networking and cooperation*).

Una delle prime attività concrete del progetto ha portato alla realizzazione di un portale dedicato alla descrizione del progetto stesso, alla raccolta di informazioni e documenti sulle persone coinvolte, sulle aree di interesse, sulle ricerche in corso. Il prototipo di tale portale<sup>11</sup> è attualmente online, e permette un'incrementale definizione ed inserimento di tali dati, informazioni e documenti (attualmente in corso). L'approccio adottato è però differente da quello che tipicamente caratterizza un portale web dinamico (schematizzato nella Figura 5 a sinistra): in questo genere di schema le informazioni che vengono utilizzate per popolare le pagine del portale vengono organizzate e memorizzate in modo persistente in una base dati (tipicamente di tipo relazionale).

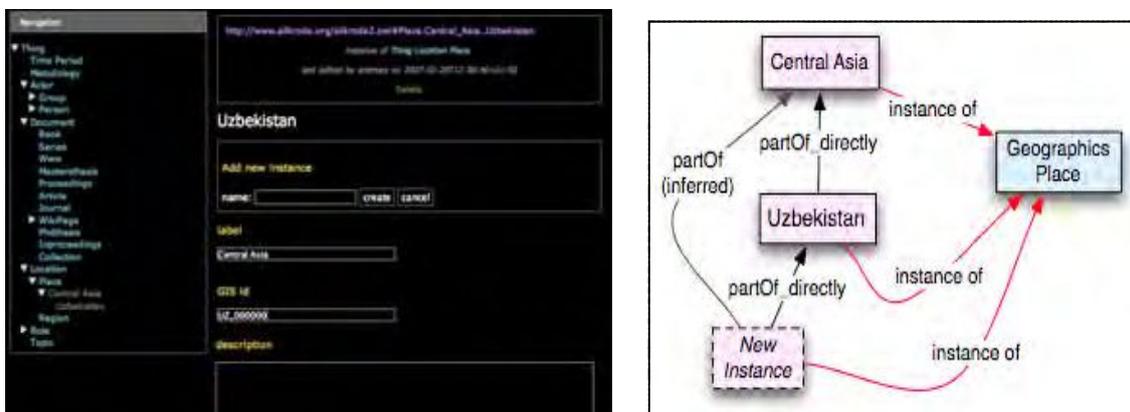


**Figura 5 - Una tradizionale architettura di un sistema web (a sinistra) confrontata all'architettura adottata per il portale del progetto SilkRoDE (a destra)**

<sup>11</sup> [www.silkrode.org](http://www.silkrode.org)

La necessità di avere una definizione ed inserimento incrementale di schemi, informazioni e conoscenze, ha spinto all'esplorazione di un approccio che adotti una vera e propria ontologia di dominio come strumento per organizzare, memorizzare e gestire le informazioni che vanno a comporre le pagine del portale (in uno schema illustrato nella Figura 5 a destra).

Questo approccio, descritto più approfonditamente in [6], supporta le tradizionali modalità di definizione e fruizione di normali portali web, consentendo al tempo stesso la definizione di modalità innovative di navigazione basata su percorsi semantici (si veda, ad esempio, la schematizzazione di relazioni di tipo spaziale in Figura 6), così come di ricerca di informazioni all'interno del portale, apertura all'interoperabilità e all'accesso dei dati ivi contenuti da parte di applicativi e sistemi di terze parti capaci di interpretare il contenuto dell'ontologia (espresso in un linguaggio standard quale OWL<sup>12</sup>).



**Figura 6 - Una schermata del portale all'atto di inserire e descrivere un elemento dell'ontologia di dominio ed uno schema concettuale di alcune delle relazioni spaziali che possono essere sfruttate per definire e realizzare percorsi di navigazione semantici**

<sup>12</sup> <http://www.w3.org/TR/owl-features/>

## 6 Bibliografia

- [1] J. MCGUIGAN, *Culture and the Public Sphere*. 1996
- [2] T. KING, *Thinking about Cultural Resource Management: Essays from the Edge*, Walnut Creek, California, 2002, Altamira Press.
- [3] M. CATTANI, G. MANTEGARI, G. VIZZARI, Tecnologie Open source e servizi web per la per la condivisione di conoscenze sulla preistoria e la protostoria in Italia, in *Atti del Workshop "Open Source, Free Software e Open Formats nei processi di ricerca archeologici"*, Grosseto, 8 aprile 2006.
- [4] M. CATTANI, R.C. DE MARINIS, G. MANTEGARI, G. VIZZARI, Towards a Web-based Environment for Italian Prehistory and Protohistory, in *Atti del Convegno "Computer Applications and Quantitative Methods in Archaeology"*, Fargo (ND) - USA, 18-21 aprile 2006, in corso di stampa.
- [5] B. RONDELLI, S. STRIDE, G. VIZZARI, ed., Silk Roads in the Digital Era Project, in *Atti del First International Workshop on Silk Roads in the Digital Era*, University of Barcelona on the 15<sup>th</sup> and 16<sup>th</sup> of December 2006, in corso di stampa.
- [6] ANDREA BONOMI, GIUSEPPE VIZZARI, BERNARDO RONDELLI, SEBASTIAN STRIDE. An Ontology Driven Web Site and its Application in the Archaeological Context, 2nd International Workshop on Ontology, Conceptualization and Epistemology for Software and System Engineering (ONTOSE 2007).

# UN FRAMEWORK PER L'ACCESSO INTELLIGENTE ALL'INFORMAZIONE

*Marco de Gemmis, Giovanni Semeraro*

Dipartimento di Informatica

Università degli Studi di Bari

Via E. Orabona, 4 – 70126 Bari, Italy

degemmis@di.uniba.it, semeraro@di.uniba.it

Nell'era di Internet, grandi quantità di documenti sono disponibili a tutti, in ogni luogo ed in ogni momento e questo comporta però un sovraccarico informativo. Tale problema è anche conosciuto con il termine inglese di *information overload*. Anche utilizzando un motore di ricerca, spesso molti dei risultati restituiti non appagano i reali bisogni informativi dell'utente. Un possibile modo per superare questo problema è rappresentato dallo sviluppo di sistemi intelligenti che forniscano servizi personalizzati per l'accesso all'informazione. Una sfida nella costruzione di questi sistemi è costituita dal catturare informazioni che descrivano l'utente per memorizzarle in un profilo utilizzato nella fase di selezione dei contenuti da suggerire agli utenti.

Questo lavoro descrive un *framework* per l'accesso intelligente a contenuti testuali. Il processo di selezione dei contenuti è *intelligente* perché *personalizzato*, ossia in grado di adattarsi all'utente in quanto impiega profili utente acquisiti attraverso tecniche di apprendimento automatico, e *semantico*, ossia basato su tecniche avanzate di elaborazione del linguaggio naturale che consentono di estrarre dai documenti i concetti rilevanti.

Il lavoro presenta infine un servizio per l'accesso ad articoli scientifici realizzato sfruttando l'architettura del framework proposto.

## 1 Introduzione

Nell'era della moderna società dell'Informazione, la ricerca online dovrà essere sempre più vicina all'utente e ai suoi bisogni informativi, sempre più facile ed efficace. I motori di ricerca di nuova generazione conosceranno i nostri interessi e daranno risposte personalizzate. Gli ostacoli posti alla diffusione delle conoscenze umane saranno sormontati con l'aiuto di sistemi *intelligenti*.

Secondo quanto dichiarato da Sergey Brin, cofondatore con Larry Page del motore di ricerca Google, in una recente intervista pubblicata su *L'espresso* (21 giugno 2007), la ricerca online è in piena evoluzione verso strategie intelligenti e personalizzate. Ad esempio, una query deve essere processata diversamente a seconda del dispositivo attraverso il quale è posta: se l'utente digita il titolo di un film sul cellulare, probabilmente è interessato a conoscere i cinema dove è proiettato, piuttosto che leggerne una lunga recensione. Ancora, una ricerca sul cuore fatta da un giornalista ha probabilmente un obiettivo diverso da quella fatta da un medico. Queste osservazioni portano a concludere che saranno richiesti nuovi algoritmi in grado di *interpretare* le richieste da un punto di vista logico e grammaticale, e di contestualizzarle.

In sintesi, il motore di ricerca di successo sarà quello che riuscirà ad adattarsi meglio e con maggiore rapidità alle richieste formulate da ogni singolo utente.

In quest'ottica risultano di particolare interesse sistemi che riescono a filtrare i documenti, mostrando all'utente solo quelli che hanno contenuti corrispondenti ai propri interessi (memorizzati nel suo profilo).

Un fattore determinante per l'*interpretazione* dei documenti è la *semantica*. Risulta evidente che strategie basate ancora su *string matching* sono superate.

L'approccio analizzato in questo lavoro tiene invece conto del significato delle parole (producendo profili semantici ed indicizzando i documenti semanticamente) ed utilizza algoritmi di disambiguazione che risolvono i classici problemi di sinonimia (due differenti parole con lo stesso significato) e polisemia (due parole morfologicamente identiche, ma con significati differenti), presenti invece negli algoritmi di string-matching. I profili vengono appresi automaticamente dal sistema, sulla base delle interazioni dell'utente.

Il lavoro presenterà dunque una proposta di framework per l'accesso intelligente a repository di documenti testuali. Elementi cardine della proposta sono: una componente per l'*interpretazione* dei testi che si avvale dell'ontologia linguistica WordNet [8] per applicare una procedura di disambiguazione del linguaggio naturale (*word sense disambiguation*, WSD), ed una componente di profilazione utente in grado di apprendere automaticamente un profilo degli interessi dell'utente per utilizzarlo nella fase di raccomandazione dei documenti.

## 2 Analisi di lavori correlati

La ricerca è stata ispirata principalmente dai seguenti lavori. "Syskill & Weibert" [10] apprende profili utente come classificatori bayesiani capaci di raccomandare pagine web, ma rappresenta i documenti tramite l'utilizzo di parole chiave (keyword). LIBRA [9] adotta un classificatore bayesiano per produrre raccomandazioni basate sul contenuto dei libri utilizzando le descrizioni dei prodotti fornite dalle pagine web di Amazon. I documenti sono rappresentati utilizzando parole chiave e sono suddivisi in "slot", ognuno corrispondente ad una specifica sezione del documento. Come in "Syskill & Weibert", la limitazione principale di questo lavoro è che la rappresentazione dei documenti è basata su parole chiave.

"SiteIF" [5] sfrutta una rappresentazione basata su sensi per costruire un profilo utente sotto forma di rete semantica, i cui nodi rappresentano i significati, i sensi, delle parole presenti nei documenti richiesti dall'utente. Nella fase di modellazione, "SiteIF" considera i synset (significati in WordNet) nel documento esplorato durante una sessione di navigazione dell'utente. I synset sono riconosciuti da una procedura di *Word Domain Disambiguation* (WDD), che è una variante di WSD nella quale, per ogni nome nel testo, viene scelta un'etichetta di dominio (Letteratura, Religione...) invece di una etichetta di significato. In un secondo passo, i synset associati ai domini sono aggiunti alla rappresentazione del documento. Il sistema costruisce la rete semantica inserendo nei nodi i synset presenti nei documenti visitati dall'utente, assegnando ad ogni nodo un peso che è inversamente proporzionale alla frequenza del synset nel corpus. Gli archi tra i nodi rappresentano le co-occorrenze dei synset all'interno dello stesso documento.

Il nostro approccio è differente sia nel processo di disambiguazione che in quello di costruzione del profilo dell'utente. Nel framework proposto, il processo di disambiguazione non è limitato solo ai nomi e non assegna domini, ma prova ad assegnare il synset più appropriato ad ogni parola in un documento, utilizzando una strategia diversa per ogni categoria lessicale. Per quanto concerne il modello dell'utente, apprendiamo una distribuzione di probabilità dei significati presenti nel corpus dei documenti sui quali l'utente ha espresso un giudizio di rilevanza rispetto ai propri interessi.

“OntoSeek” [3] è un sistema progettato per realizzare una modalità di ricerca basata sul contenuto operante su pagine gialle ed i prodotti dei cataloghi online. Il sistema ha evidenziato il ruolo delle ontologie linguistiche nei sistemi di knowledge-retrieval. L’approccio ha mostrato che la rappresentazione strutturata del contenuto associata alle ontologie linguistiche può migliorare sia il richiamo che la precisione dei sistemi di ricerca basati sul contenuto.

Prendendo in esame quanto appreso dai lavori citati in precedenza, abbiamo introdotto nel nostro framework:

- una componente che realizza una rappresentazione dei documenti basata sul significato delle parole, ottenuta sfruttando l’ontologia linguistica WordNet;
- una componente di *profilazione utente e raccomandazione di contenuti* che apprende profili utente sotto forma di classificatori bayesiani di testi. Il processo di apprendimento è effettuato su documenti suddivisi in slot (come LIBRA) e rappresentati mediante i concetti di WordNet (come SiteIF).

La strategia che utilizziamo per passare da una rappresentazione del documento basata su parole chiave ad una basata sul significato è quella di integrare la conoscenza lessicale nel processo di indicizzazione dei documenti.

Molti metodi sono stati proposti per assolvere a questo compito. Scott e Matwin [12] hanno proposto di includere l’informazione di Wordnet direttamente negli attributi, espandendo ogni parola nel training set con tutti i suoi sinonimi in WordNet, allo scopo di evitare un processo di WSD. Questo approccio ha mostrato un peggioramento dell’efficacia nel classificatore ottenuto, dovuto fondamentalmente all’ambiguità delle parole, suggerendo l’opportunità di un processo di disambiguazione. Lavori più recenti come [7] e [2] hanno fornito evidenza sperimentale dell’utilità di integrare il processo di WSD nel processo di classificazione, specialmente in presenza di un numero limitato di esempi etichettati (già classificati), situazione che caratterizza il processo di costruzione del profilo utente. In [7], WordNet è usato come un thesaurus gerarchico sia per il processo di WSD che per quelli di classificazione, mentre il nostro approccio si avvale dell’organizzazione gerarchica unicamente per il calcolo della similarità semantica tra i synset.

In [2] gli autori sperimentano vari modi per collegare le parole con i loro significati, riscontrando risultati positivi sui diversi corpora: Reuters 25178, OSHUMED e FAODOC. I risultati migliorati possono essere imputati ad una identificazione dei concetti espressi da più parole (“Intelligenza Artificiale”), al riconoscimento dei sinonimi, e all’utilizzo delle strutture dell’ontologia per la generalizzazione. Nel nostro approccio di knowledge-based WSD, la generalizzazione è usata solo per determinare il concetto-padre più specifico tra più concetti al fine di calcolare la similarità semantica.

### 3 Definizione del framework

L'architettura concettuale del framework proposto è descritta in Figura 1. Il modulo *Content Analyzer* consente l'introduzione di un livello semantico nel processo di raccomandazione dei documenti, attraverso l'analisi dei documenti stessi con procedure avanzate di elaborazione del linguaggio naturale che ne identificano i concetti più rappresentativi.

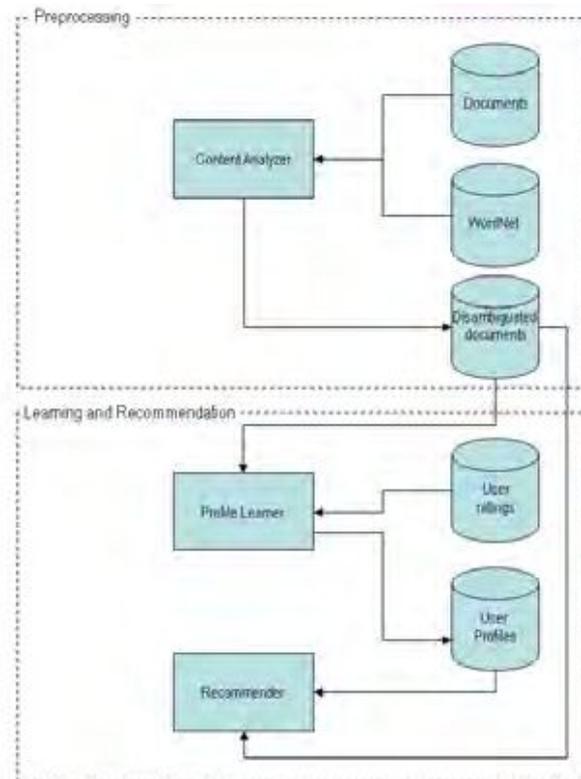


Figura 1 - Architettura del framework

L'idea alla base dell'identificazione dei concetti è quella di adottare una procedura di disambiguazione che seleziona, tra tutti i possibili significati (o sensi) di una parola polisemica, quello più indicato in base al contesto in cui la parola compare. In questo modo, ogni documento può essere rappresentato ed indicizzato usando concetti piuttosto che *keyword*, realizzando così una soluzione che tenta di superare il problema dell'imprecisione delle raccomandazioni dovuta all'ambiguità del linguaggio naturale.

Il risultato finale del processo realizzato dal modulo *Content Analyzer* di analisi è un repository di documenti disambiguati. Questo processo di indicizzazione semantica è fortemente basato su tecniche di WSD che sfruttano che conoscenza linguistica ingegnerizzata nell'ontologia lessicale WordNet.

Il modulo *Profile Learner* implementa tecniche di apprendimento supervisionato con l'obiettivo di inferire un modello probabilistico degli interessi dell'utente, considerando come esempi di addestramento i documenti precedentemente considerati dall'utente come rilevanti (o non rilevanti) rispetto ai propri interessi.

Tale modello è il *profilo semantico* dell'utente, poiché appreso da documenti indicizzati semanticamente, che contiene i concetti maggiormente rappresentativi dei suoi interessi.

I profili semantici sono utilizzati dal modulo *Recommender* per suggerire documenti rilevanti, effettuando un *matching* probabilistico tra i concetti presenti nel profilo utente e quelli presenti nei documenti tra i quali selezionare quelli da raccomandare.

Nell'articolo sarà presentato un servizio di raccomandazione progettato per partecipanti a conferenze scientifiche. In particolare, il servizio è stato realizzato per i documenti dell'edizione 2004 della "International Semantic Web Conference (ISWC)", con l'obiettivo di costruire un programma personalizzato per ogni partecipante, in base agli interessi di ricerca inseriti nel suo profilo inferito dagli articoli maggiormente rilevanti delle edizioni 2002 e 2003 della stessa conferenza. Il profilo è confrontato con tutti gli articoli presentati a ISWC 2004 al fine di identificare quelli maggiormente rilevanti, che saranno evidenziati nel programma personalizzato.

Nel seguito si descriveranno in dettaglio le tecniche utilizzate dai moduli dell'architettura.

#### **4 Content Analyzer: indicizzazione semantica dei documenti**

I documenti testuali non possono essere direttamente interpretati dagli algoritmi di *machine learning*. Deve essere applicata una procedura di indicizzazione che trasforma un documento in una rappresentazione compatta del suo contenuto. Una scelta tipica per l'indicizzazione dei documenti è il classico approccio bag-of-words (BOW), in cui ogni documento è rappresentato come un vettore di attributi. Ogni coordinata rappresenta una parola del vocabolario (l'insieme di tutte le parole distinte nella collezione da indicizzare) ed è avvalorata con uno score numerico che ne indica l'importanza all'interno del documento [13].

Il modello BOW è esteso per ottenerne uno in cui ogni documento è rappresentato dal senso identificato dalle parole in esso contenute, insieme alle loro rispettive occorrenze. In questo caso, "senso" è utilizzato come un sinonimo di "significato". Questo modello di indicizzazione semantica è sfruttato dall'algoritmo di *machine learning* per costruire profili utente semantici. Ogni implementazione di indicizzazione di un documento basata su significato (sense-based) deve affrontare il problema che, in contrapposizione all'indicizzazione basata sulle sole parole, è difficile individuare l'eventuale ripetersi dei concetti all'interno di un documento a causa della complessità di interpretazione del contesto.

Come conseguenza, si rende necessario prevedere una procedura per assegnare i significati alle parole. Questo compito, conosciuto con il nome di *Word Sense Disambiguation (WSD)*, consiste nel determinare quale dei significati di una parola ambigua è pertinente in un particolare uso di quella parola [6]. Inoltre, lo scopo di un algoritmo di WSD è di associare ogni parola  $w_i$  che occorre in un documento con il suo significato più appropriato (o senso)  $s$ , sfruttando il contesto  $C$  nel quale  $w_i$  si trova, definito comunemente come un insieme di parole che precedono e seguono  $w_i$ . Il senso  $s$  è selezionato da un insieme predefinito di possibilità, comunemente noto come *sense inventory*. Nell'algoritmo proposto, WordNet è usato come repository di significati.

WordNet è un lexicon semantico per la lingua Inglese. Raggruppa le parole in insiemi di sinonimi chiamati *synset*, fornisce brevi descrizioni generali, e registra le varie relazioni semantiche tra i vari insiemi di sinonimi.

WordNet ha un duplice scopo: produrre una combinazione di dizionario e thesaurus che sia intuitivamente usabile, e supportare l'analisi automatica del testo da parte delle applicazioni che ne fanno utilizzo. WordNet distingue nomi, verbi, aggettivi ed avverbi poiché essi seguono differenti regole grammaticali. Ogni synset corrisponde dunque ad un concetto ed è costituito dai sinonimi che possono essere utilizzati per esprimerlo. Il significato dei synset è inoltre chiarito da brevi descrizioni testuali. Un tipico esempio di synset con una descrizione è riportato di seguito:

*{01611196} good, right, ripe -- (most suitable or right for a particular purpose; "a good time to plant tomatoes"; "the right time to act"; "the time is ripe for great sociological changes")*

Si possono distinguere, nell'ordine: il codice identificativo del synset, i sinonimi che ne fanno parte, e la relativa descrizione, chiamata anche *glossa*. Molti synset sono connessi ad altri tramite un certo numero di relazioni semantiche. Nei nostri algoritmi, usiamo le seguenti relazioni semantiche per i nomi ed i verbi:

- iperonimia: Y è un iperonimo di X se ogni X è un (tipo di) Y;
- iponimia: Y è un iponimo di x se ogni Y è un (tipo di) X.

WordNet fornisce inoltre il conteggio della polisemia di una parola come numero di synset che contengono quella parola. Se una parola compare in più synset (cioè ha più significati), allora di solito alcuni significati sono più comuni di altri. WordNet quantifica questa situazione tramite un punteggio di frequenza. In vari corpora campione, che la comunità scientifica considera di riferimento, le parole sono etichettate "semanticamente" con il synset corrispondente, consentendo così una stima frequentista di quanto spesso la parola appare con un significato specifico.

#### **4.1 JIGSAW: un algoritmo per la disambiguazione del linguaggio naturale**

Considerando che le prestazioni degli algoritmi di WSD variano in base al part-of-speech (POS) della parola che deve essere disambiguata, l'algoritmo JIGSAW proposto si basa sull'idea di combinare tre strategie differenti per disambiguare sostantivi, verbi, aggettivi e avverbi.

Un adattamento dell'algoritmo di Lesk per WSD basata su dizionari è stato usato per disambiguare aggettivi ed avverbi [1]. Per la disambiguazione dei sostantivi, si è adottata una versione modificata dell'algoritmo proposto da Resnik [11]. L'algoritmo che abbiamo sviluppato per disambiguare i verbi sfrutta i sostantivi nel contesto del verbo ed i sostantivi presenti sia nelle definizioni che nelle frasi che WordNet utilizza per descrivere l'uso del verbo.

L'algoritmo disambigua solo le parole che hanno almeno un synset. JIGSAW prende come input un documento  $d = (w_1, w_2, \dots, w_h)$  codificato come una lista di parole così come appaiono nel testo originale e fornisce in output una lista di synset di WordNet  $X = (s_1, s_2, \dots, s_k)$ , ( $k \leq h$ ), nella quale ogni elemento  $s_j$  è ottenuto disambiguando la parola obiettivo  $w_i$  basata sull'informazione ottenuta da WordNet. L'algoritmo è composto da tre procedure: JIGSAW<sub>nouns</sub> per la disambiguazione dei sostantivi, JIGSAW<sub>verbs</sub> per la disambiguazione dei verbi e JIGSAW<sub>others</sub> per la disambiguazione degli aggettivi e degli avverbi.

*JIGSAW<sub>nouns</sub>*. La procedura è ottenuta apportando alcune modifiche all’algoritmo progettato da Resnik [11] per la disambiguazione di un insieme di sostantivi. Dato un insieme di sostantivi  $W = \{w_1, w_2, \dots, w_n\}$ , ottenuto dal documento  $d$ , dove ad ogni  $w_i$  è associato un insieme di possibili sensi  $S_i = \{s_{i1}, s_{i2}, \dots, s_{ik}\}$ , lo scopo è quello di associare a ciascun  $w_i$  il senso più appropriato  $s_{ih} \in S_i$ , a seconda della similarità di  $w_i$  con le altre parole in  $W$  (il contesto di  $w_i$ ). L’idea è di definire una funzione  $\varphi(w_i, s_{ij})$ ,  $w_i \in W$ ,  $s_{ij} \in S_i$ , che calcoli un valore in  $[0,1]$ , che rappresenti la confidenza con cui il senso  $s_{ij}$  può essere associato a  $w_i$ .

L’intuizione dietro questo algoritmo è essenzialmente la stessa intuizione sviluppata da Lesk [4]: l’assegnazione più plausibile di sensi a co-occorrenze multiple di parole è quella che massimizza la correlazione (*relatedness*) del significato tra i concetti scelti. *JIGSAW<sub>nouns</sub>* differisce dall’algoritmo originale di Resnik nella misura di similarità usata per calcolare la relazione tra due sensi. Abbiamo adottato la misura Leacock-Chodorow, che è basata sulla lunghezza del cammino tra i concetti nella gerarchia *IS-A*. L’idea alla base di questa misura è che la similarità tra due synset,  $s_1$  e  $s_2$ , è inversamente proporzionale alla loro distanza nella gerarchia *IS-A* di WordNet. La distanza è calcolata:

- cercando il sussuntore più specifico (Most Specific Subsumer - MSS) tra  $s_1$  e  $s_2$  (ogni antenato sia di  $s_1$  che di  $s_2$  nella gerarchia di WordNet è un sussuntore, MSS è quello al livello più basso);
- contando il numero di nodi nel cammino tra  $s_1$  e  $s_2$  che attraversa il loro MSS.

Ad esempio, la Figura 2 mostra che la lunghezza del percorso tra *cat* (mammifero felino) e *mouse* (roditore) attraversando il concetto *mammiferi placentari* è 5. La similarità tra i due synset è calcolata dalla funzione descritta dall’Algoritmo 1. Nell’esempio,  $\text{SYNSIM}(\textit{feline mammal}, \textit{rodent}) = 0.806$ . Abbiamo esteso questa misura introducendo il parametro  $k$  che restringe la ricerca per MSS a  $k$  antenati (cioè si sale nella gerarchia *IS-A* di WordNet finché viene trovato MSS oppure sono stati esplorati  $k + 1$  antenati sia di  $s_1$  che di  $s_2$ ). Questo garantisce che MSS “più astratti” (cioè meno informativi) vengano ignorati.

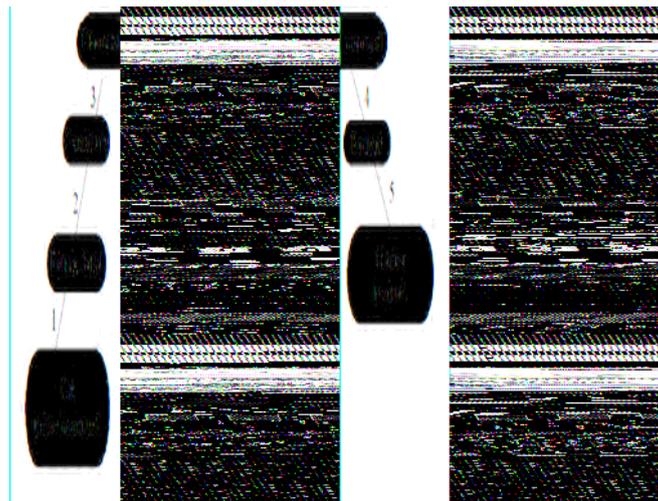


Figura 2 - Un frammento della gerarchia di WordNet

---

**Algorithm 1** The Leacock-Chodorow similarity function between synsets

---

```
1: function SYNSIM( $s_1, s_2$ )           ▷ The similarity for the synsets  $s_1$  and  $s_2$ 
2:    $MSS \leftarrow MSS(s_1, s_2)$        ▷  $MSS(s_1, s_2)$  returns the most specific subsumer
   between  $s_1$  and  $s_2$ 
3:    $N_p \leftarrow$  the number of nodes in the shortest path  $p$  from  $s_1$  to  $s_2$  by traversing
   MSS
4:    $D \leftarrow$  maximum depth of the taxonomy
5:    $r \leftarrow -\log(N_p/2D)$ 
6:   return  $r$ 
7: end function
```

---

---

**Algorithm 2** The modified similarity function between synsets

---

```
function SYNSIM( $s_1, s_2, k$ ) ▷ The similarity for the synsets  $s_1$  and  $s_2$ , search for
MSS restricted to  $k$  ancestors
 $MSS \leftarrow MSS(s_1, s_2, k)$  ▷  $MSS(s_1, s_2, k)$  returns the most specific subsumer
between  $s_1$  and  $s_2$ ,  $k$  ancestors explored, at the most
 $r \leftarrow 0$ 
if MSS not null then
   $N_p \leftarrow$  the number of nodes in the shortest path  $p$  from  $a$  to  $b$  by traversing
  MSS
   $D \leftarrow$  maximum depth of the taxonomy
   $r \leftarrow -\log(N_p/2D)$ 
end if
return  $r$ 
end function
```

---

Nell'esempio,  $\text{SYNSIM}(\textit{feline mammal}, \textit{rodent}, 4) = 0.806$ ,  $\text{SYNSIM}(\textit{feline mammal}, \textit{rodent}, 2) = 0$ . Prima di descrivere l'intera procedura  $JIGSAW_{\textit{nouns}}$ , abbiamo bisogno di definire la similarità semantica tra due parole  $w_1$  e  $w_2$ , come descritta nell'Algoritmo 3. Essa è calcolata come similarità massima tra tutte le possibili coppie di concetti corrispondenti alle due parole. In sostanza, è il valore di SYNSIM tra i due concetti più simili che le due parole possono esprimere.

In aggiunta alla funzione di similarità semantica, la  $JIGSAW_{\textit{nouns}}$  differisce dall'algoritmo di Resnik nell'uso di:

- un fattore Gaussiano che tiene in considerazione la distanza nel testo tra le parole che devono essere disambiguate;
- un fattore che dà maggiore importanza ai synset che sono più comuni degli altri, in accordo con il punteggio di frequenza;
- una ricerca parametrizzata per il MSS tra due concetti (la ricerca è limitata ad un dato numero di antenati).

---

**Algorithm 3** The similarity function between words

---

```
function SIM( $w_1, w_2, k$ )    ▷ The similarity for the words  $w_1$  and  $w_2$ , search for
MSS restricted to  $k$  ancestors
 $S_1 \leftarrow \{s_{11}, s_{12}, \dots, s_{1n}\}$     ▷ sense inventory for  $w_1$ 
 $S_2 \leftarrow \{s_{21}, s_{22}, \dots, s_{2m}\}$     ▷ sense inventory for  $w_2$ 
for all  $s_{1i} \in S_1$  and  $s_{2j} \in S_2$  do
     $MSS_{ij} \leftarrow MSS(s_{1i}, s_{2j}, k)$     ▷ returns the most specific subsumer between
     $s_{1i}$  and  $s_{2j}$ ,  $k$  ancestors explored, at the most
     $N_{ij} \leftarrow$  the number of nodes in the shortest path from  $s_{1i}$  to  $s_{2j}$  by traversing
     $MSS_{ij}$ 
end for
 $MSS_{xy} \leftarrow$  the MSS that minimizes  $N_{ij}$  ▷  $MSS_{xy}$  is the synset that subsumes
both  $w_1$  and  $w_2$ , in any sense of either word
 $r \leftarrow SynSim(s_{1x}, s_{2y}, k)$     ▷ the maximum value of similarity for all possible
pairs of senses associated with  $w_1$  and  $w_2$ 
return  $r$ 
end function
```

---

L'Algoritmo 4 descrive la procedura completa per la disambiguazione dei sostantivi. Questo algoritmo considera le parole in  $W$  a coppia. Per ogni coppia  $(w_i, w_j)$ , si identifica il sussuntore più specifico  $MSS_{ij}$ , riducendo la ricerca al più a *depth1* antenati. Quindi la similarità  $sim(w_i, w_j, depth2)$  tra due parole è calcolata riducendo la ricerca per MSS al più a *depth2* antenati.  $MSS_{ij}$  è considerato come *supporto di evidenza* per i synsets  $s_{ik}$  in  $S_i$  e  $s_{jh}$  in  $S_j$  che sono discendenti di  $MSS_{ij}$ . L'ammontare del supporto che ha contribuito alla comparazione della coppia è il valore di similarità calcolato secondo la funzione descritta nell'Algoritmo 3, pesata attraverso un fattore gaussiano che prende in considerazione la posizione di  $w_i$  e  $w_j$  in  $W$  (più breve è la distanza tra le parole, maggiore è il peso). Il valore  $\varphi(i, k)$  assegnato ad ogni synset candidato  $s_{ik}$  per la parola  $w_i$  è la somma dei due elementi. Il primo è la proporzione del supporto ricevuto, sul supporto possibile, calcolato come:  $support_{ik}/normalization_i$  nello pseudocodice.

---

**Algorithm 4** The procedure for disambiguating nouns derived from the algorithm by Resnik

---

```

1: procedure JIGSAWnouns(W, depth1, depth2)                                ▷ finds the proper
   synset for each polysemous noun in the set  $W = \{w_1, w_2, \dots, w_n\}$ , depth1 and
   depth2 are used in the computation of MSS
2:   for all  $w_i, w_j \in W$  do
3:      $sim \leftarrow sim(w_i, w_j, depth1) * Gauss(position(w_i), position(w_j))$           ▷
       Gauss(x,y) is a Gaussian function which takes into account the difference
       between the positions of  $w_i$  and  $w_j$ 
4:      $MSS_{ij} \leftarrow MSS(w_i, w_j, depth2)$     ▷  $MSS_{ij}$  is the most specific subsumer
       between  $w_i$  and  $w_j$ , search for MSS restricted to depth2 ancestors
5:     for all  $s_{ik} \in S_i$  do
6:       if is-ancestor( $MSS_{ij}, s_{ik}$ ) then          ▷ if  $MSS_{ij}$  is an ancestor of  $s_{ik}$ 
7:          $support_{ik} \leftarrow support_{ik} + sim$ 
8:       end if
9:     end for
10:    for all  $s_{jh} \in S_j$  do
11:      if is-ancestor( $MSS_{ij}, s_{jh}$ ) then
12:         $support_{jh} \leftarrow support_{jh} + sim$ 
13:      end if
14:    end for
15:     $normalization_i \leftarrow normalization_i + sim$ 
16:     $normalization_j \leftarrow normalization_j + sim$ 
17:  end for
18:  for all  $w_i \in W$  do
19:    for all  $s_{ik} \in S_i$  do
20:      if  $normalization_i > 0$  then
21:         $\varphi(i, k) \leftarrow \alpha * support_{ik} / normalization_i + \beta * R(k)$ 
22:      else
23:         $\varphi(i, k) \leftarrow \alpha / |S_i| + \beta * R(k)$ 
24:      end if
25:    end for
26:  end for
27: end procedure

```

---

L'altro elemento che contribuisce a  $\varphi(i, k)$  è il fattore  $R(k)$  che tiene conto del punteggio di  $s_{ik}$  in WordNet, ovvero quanto comune è il senso  $s_{ik}$  per la parola  $w_i$ .  $R(k)$  è calcolato come:

$$R(K) = 1 - 0.8 * \frac{k}{n-1}$$

**Equazione 1**

dove  $n$  è la cardinalità dell'insieme dei sensi  $S_i$  per  $w_i$  e  $k$  è il punteggio di  $s_{ik}$  in  $S_i$ , partendo da 0. Infine, entrambi gli elementi sono pesati attraverso due parametri:  $\alpha$ , che controlla il contributo dato a  $\varphi(i, k)$  attraverso il supporto normalizzato, e  $\beta$ , che controlla il contributo dato dal punteggio di  $s_{ik}$ . Abbiamo scelto i valori  $\alpha = 0.7$  e  $\beta = 0.3$ .

Il synset assegnato ad ogni parola in  $W$  è quello con il valore più alto di  $\varphi$ . Notiamo che usiamo due parametri differenti,  $depth1$  e  $depth2$  per assegnare la massima profondità per la ricerca di MSS:  $depth1$  limita la ricerca per MSS calcolato nella funzione di similarità, mentre  $depth2$  limita il calcolo di MSS usato per assegnare supporto ai synset candidati. Ad esempio, impostando  $depth1 = 6$  e  $depth2 = 3$ , permettiamo di salire nella gerarchia di WordNet per ricercare MSS fino ad un alto livello di astrazione, ma imponiamo un vincolo più forte per il calcolo di MSS usato per assegnare il supporto. Questo significa che soltanto i synset che sono discendenti del più “specifico” MSS riceveranno supporto.

$JIGSAW_{verbs}$ . Prima di descrivere la procedura  $JIGSAW_{verbs}$ , è necessario definire la *descrizione* di un synset. Essa è la stringa ottenuta concatenando la glossa e le frasi che WordNet usa per spiegare l’uso di una parola.

Ad esempio la glossa per il synset corrispondente al senso n.2 del verbo look ( $\{look, appear, seem\}$ ) è “*give a certain impression or have a certain outward aspect*”, mentre alcuni esempi dell’uso del verbo sono: “*She seems to be sleeping*”; “*This appears to be a very difficult problem*”. La descrizione del synset è “*give a certain impression or have a certain outward aspect She seems to be sleeping This appears to be a very difficult problem*”. In primo luogo  $JIGSAW_{verbs}$  include nel contesto  $C$  per il verbo target  $w_i$ , tutti i nomi nella finestra delle  $2 \cdot n$  parole che circondano  $w_i$ . Per ciascun synset candidato  $s_{ik}$  di  $w_i$ , l’algoritmo calcola  $nouns(i,k)$ , che è l’insieme dei nomi nella descrizione per  $s_{ik}$ . Nell’esempio citato,  $nouns(look,2)=\{impression, aspect, problem\}$ . Quindi, per ogni  $w_j$  in  $C$  e per ogni synset  $s_{ik}$ , viene calcolato il seguente valore:

$$\max_{jk} = \max_{w_j \in nouns(i,k)} \{SIM(w_j, w_i, depth)\}$$

**Equazione 2**

dove  $SIM(w_j, w_i, depth)$  è definito come nell’Algoritmo 3. In altre parole,  $\max_{jk}$  è il più alto valore di similarità per  $w_j$ , rispetto ai nomi collegati al  $k$ -mo senso per  $w_i$ . Infine, il punteggio complessivo di similarità tra  $s_{ik}$  e l’intero contesto  $C$  è calcolato come:

$$\varphi(i,k) = R(k) \cdot \frac{\sum_{w_j \in C} Gauss(position(w_i), position(w_j)) \cdot \max_{jk}}{\sum_h Gauss(position(w_i), position(w_k))}$$

**Equazione 3**

dove  $R(k)$  è definita come nell’Equazione 1 e  $Gauss(position(w_i), position(w_j))$  è lo stesso fattore Gaussiano utilizzato in  $JIGSAW_{nouns}$  che attribuisce un peso più alto alle parole più vicine alla parola target. Il synset assegnato a  $w_i$  è quello con il più alto valore di  $\varphi$ .

$JIGSAW_{others}$ . La procedura è basata sull’algoritmo WSD proposto in [1]. L’idea è di confrontare le glosse di ogni senso candidato per la parola target con le glosse di tutte le parole nel suo contesto. Indichiamo con  $W_i$  l’insieme di sensi per la parola target  $w_i$ . Per ogni  $s_{ik}$  in  $W_i$ ,  $JIGSAW_{others}$  computa due stringhe:  $targetGloss_{ik}$ , che contiene le parole nella glossa di  $s_{ik}$ , e  $contextGloss_i$ , che contiene le parole nelle glosse di tutti i synset corrispondenti ad ogni parola nel contesto per  $w_i$ .

Alla fine, la procedura computa la sovrapposizione (overlap) fra  $contextGloss_i$  e  $targetGloss_{ik}$ , assegnando a  $w_i$  il synset con il più alto punteggio di overlap. Il punteggio è calcolato conteggiando le parole che occorrono sia in  $contextGloss_i$  che in  $targetGloss_{ik}$ .

## 4.2 Dall'indicizzazione per parole chiave a quella per concetti

La procedura WSD è fondamentale per ottenere una rappresentazione *vector space* [13] basata su *synset* che abbiamo chiamato Bag-Of-Synsets (BOS). In questo modello, ogni documento è codificato come un vettore di *synset*, piuttosto che come un vettore di parole, come nel classico modello Bag-Of-Words (BOW). Un'altra caratteristica chiave dell'approccio è che ogni documento è rappresentato da un insieme di *slot*, dove ogni slot è un campo testuale corrispondente ad una specifica caratteristica del documento. In questo modo, si può tentare di sfruttare nel meccanismo di indicizzazione anche una eventuale struttura dei documenti. Ad esempio, ipotizzando di indicizzare articoli scientifici, si possono considerare almeno tre slot:

- *title*, il titolo dell'articolo;
- *authors*, la lista dei nomi degli autori;
- *abstract*, il breve testo che presenta i punti principali dell'articolo.

Il testo in ogni slot è rappresentato in accordo al modello BOS, conteggiando separatamente le occorrenze di un *synset* negli slot in cui appare. Più formalmente, assumiamo di avere una collezione di  $N$  documenti rappresentati usando  $M$  slot. Indichiamo con  $m$  l'indice dello slot, per  $n=1,2,\dots,N$ , l' $n$ -esimo documento è ridotto a  $M$  bag of *synset*, uno per ogni slot:

$$d_n^m = \langle t_{n1}^m, t_{n2}^m, \dots, t_{nDnm}^m \rangle \quad m = 1, \dots, M$$

dove  $t_{nk}^m$  è il  $k$ -esimo *synset* nello slot  $s_m$  del documento  $d_n$  e  $D_{nm}$  è il numero totale di *synset* che appaiono nel  $m$ -esimo slot del documento  $d_n$ . Per ogni  $n, k, m$ ,  $t_{nk}^m$  appartiene a  $V_m$ , che è il vocabolario per lo slot  $s_m$  (l'insieme di tutti i *synset* distinti trovati nello slot  $s_m$ ). Il documento  $d_n$  è dunque rappresentato nel *vector space* da  $M$  vettori *synset-frequenza*:

$$f_n^m = \langle w_{n1}^m, w_{n2}^m, \dots, w_{nDnm}^m \rangle$$

dove  $w_{nk}^m$  è il peso del *synset*  $t_k$  nello slot  $s_m$  del documento  $d_m$ . Tale peso può essere calcolato in modi diversi: può semplicemente essere il numero di volte che il *synset*  $t_k$  appare nello slot  $s_m$  o uno score TF-IDF più complesso. La nostra ipotesi è che la rappresentazione del documento proposta aiuti ad ottenere profili capaci di raccomandare documenti semanticamente più vicini agli interessi dell'utente. La differenza rispetto ai profili keyword-based è che sono utilizzati gli identificatori univoci dei *synset* al posto delle parole. La prossima sezione descrive gli algoritmi di apprendimento adottati per costruire profili utente semantici, utilizzando la rappresentazione BOS dei documenti. Ovviamente, nel caso in cui l'indicizzazione sia effettuata utilizzando il modello BOW, dalla fase di apprendimento si otterranno profili basati su parole chiave, mentre un'indicizzazione semantica produrrà come effetto profili semantici basati su *synset* di WordNet. Si utilizzerà nel seguito il termine "token" per indicare una parola o un *synset*.

## 5 Apprendimento di profili utente semantici per la raccomandazione di documenti

Il modulo *Profile Learner* del framework utilizza un algoritmo Naïve Bayes per la costruzione induttiva di profili utente sotto forma di classificatori di testi. L'algoritmo bayesiano rappresenta un approccio probabilistico all'apprendimento induttivo. Il modello probabilistico appreso stima la probabilità *a posteriori*,  $P(c_j | d_i)$ , che il documento  $d_i$  appartenga alla classe  $c_j$ . Per classificare un documento  $d_i$ , è selezionata la classe con la più

alta probabilità. Come modello per il classificatore Naïve Bayes, utilizziamo quello ad eventi multinomiale:

$$P(c_j | d_i) = \frac{P(c_j)}{P(d_i)} \prod_{w \in V_{d_i}} P(t_k | c_j)^{N(d_i, t_k)}$$

**Equazione 4**

dove  $N(d_i, t_k)$  è definito come il numero di volte in cui il token (parola o synset)  $t_k$  appare nel documento  $d_i$ . Bisogna notare che, piuttosto di fare il prodotto di tutte i token distinti nel corpus,  $V$ , utilizziamo soltanto il subset del vocabolario,  $V_{d_i}$ , che contiene i token che appaiono nel documento  $d_i$ .

Poiché ogni istanza può essere codificata come un vettore di BOS (o di BOW), uno per ogni slot, l'Equazione 4 diventa:

$$P(c_j | d_i) = \frac{P(c_j)}{P(d_i)} \prod_{m=1}^{|S|} \prod_{k=1}^{|b_{im}|} P(t_k | c_j, s_m)^{n_{kim}}$$

**Equazione 5**

dove  $S = \{s_1, s_2, \dots, s_{|S|}\}$  è l'insieme di slot,  $b_{im}$  è la BOS (BOW) nello slot  $s_m$  del documento  $d_i$ ,  $n_{kim}$  è il numero di occorrenze del token  $t_k$  in  $b_{im}$ . Il sistema di profilazione IItem Recommender (ITR), che include le componenti *Profile Learner* e *Recommender*, implementa questo approccio per classificare documenti come interessanti o non interessanti per un particolare utente. Poiché il termine  $P(d_i)$  è costante rispetto alla classe, lo si può trascurare nel calcolo della formula 5, in cui è necessario di conseguenza stimare soltanto  $P(c_j)$  e  $P(t_k | c_j, s_m)$  nella fase di training del sistema. I documenti utilizzati per addestrare il sistema devono essere etichettati dagli utenti con *rating* che ne esprimono il grado di interesse. I rating sono espressi attraverso una scala discreta da MIN a MAX, solitamente comprendente un numero dispari di valori (Es.: 1-5). Un documento etichettato con un rating  $r$ , con  $r \leq (MIN + MAX / 2)$ , appartiene alla classe  $c_-$  (user-dislikes), altrimenti si considera nella classe  $c_+$  (user-likes). Ogni rating è stato normalizzato per ottenere un range di valori fra 0 ed 1:

$$w_+^i = \frac{r-1}{MAX-1}; \quad w_-^i = 1 - w_+^i$$

**Equazione 6**

I valori  $w_+^i, w_-^i$  in Equazione 6 sono utilizzati per pesare le occorrenze di un token in un documento e per stimare le probabilità dal training set TR. Le probabilità a priori delle classi sono così computate:

$$\hat{P}(c_j) = \frac{\sum_{i=1}^{|TR|} w_j^i + 1}{|TR| + 2}$$

**Equazione 7**

Per stimare  $P(t_k | c_j, s_m)$ , è stata adottata la stima di Witten-Bell [15] tenendo conto del fatto che i documenti sono strutturati in slot e che le occorrenze delle parole sono pesate utilizzando i pesi dell'Equazione 6:

$$\hat{P}(t_k | c_j, s_m) = \begin{cases} \frac{N(t_k, c_j, s_m)}{|V_{c_j}| + \sum_i N(t_i, c_j, s_m)} & \text{se } N(t_k, c_j, s_m) \neq 0 \\ \frac{|V|}{|V_{c_j}| + \sum_i N(t_i, c_j, s_m)} \frac{1}{|V| - |V_{c_j}|} & \text{se } N(t_k, c_j, s_m) = 0 \end{cases}$$

**Equazione 8**

dove  $N(t_k, c_j, s_m)$  è il numero delle occorrenze pesate del token  $t_k$  nei dati di training per la classe  $c_j$  nello slot  $s_m$ ,  $V_{c_j}$  è vocabolario della classe  $c_j$  e  $V$  è il vocabolario di tutte le classi.  $N(t_k, c_j, s_m)$  è calcolato come:

$$N(t_k, c_j, s_m) = \sum_{i=1}^{|TR|} w_j^i n_{kim}$$

**Equazione 9**

Nella formula 9,  $n_{kim}$  è il numero di occorrenze del token  $t_k$  nello slot  $s_m$  della  $i$ -esima istanza. La somma di tutti i  $N(t_k, c_j, s_m)$  nel denominatore dell'Equazione 8 denota la lunghezza totale pesata dello slot  $s_m$  nella classe  $c_j$ . In altre parole,  $\hat{P}(t_k | c_j, s_m)$  è stimato come il rapporto fra le occorrenze pesate del token  $t_k$  nello slot  $s_m$  della classe  $c_j$  e la lunghezza totale pesata dello slot. Il risultato finale del processo di apprendimento è un modello probabilistico utilizzato per classificare un nuovo documento nella classe  $c_+$  o  $c_-$ . Il modello può essere utilizzato per costruire un profilo personale che include i token che risultano maggiormente indicativi delle preferenze dell'utente, in accordo con il valore delle probabilità condizionate nella formula 8. Esperimenti dettagliati che confrontano l'accuratezza dei profili synset-based con quella dei profili keyword-based sono riportati in [14].

## 6 Il servizio Conference Participant Advisor

Il servizio "Conference Participant Advisor" è basato su ITR ed è stato progettato per fornire un supporto per ricercatori che partecipano a conferenze di tipo scientifico, in cui spesso il vasto programma non consente di seguire tutti gli articoli presentati. Diventa dunque fondamentale selezionare gli articoli più rilevanti.

Il profilo semantico del partecipante registrato al servizio è utilizzato per costruire un programma personalizzato della conferenza. Il prototipo del servizio è stato realizzato utilizzando i dati dell'edizione 2004 della "International Semantic Web Conference". È stata creata una copia (non ufficiale) del sito nella quale il servizio è stato integrato a scopo dimostrativo (Figura 3).



**Figura 3 - Home page della versione dimostrativa del servizio**

Il partecipante si può iscrivere al servizio fornendo un indirizzo di posta elettronica.

Successivamente al passo di registrazione, l'utente può navigare il repository di articoli delle edizioni 2002 e 2003, fornendo un *rating* di rilevanza rispetto ai propri interessi al fine di raccogliere l'insieme di documenti per l'addestramento del sistema di profilazione.

Il servizio integra anche un motore di ricerca, anch'esso semantico, che consente all'utente di ritrovare articoli sui quali esprimere un giudizio di rilevanza, ottenendo così esempi di addestramento appropriati per la componente di profilazione.

Anche il motore di ricerca si basa sul modello BOS in modo da implementare un meccanismo di ricerca semantica che prescinde dalle parole chiave usate dall'utente per descrivere la query.

Supponiamo che un utente componga la query "categorization"; il motore di ricerca la analizza e seleziona da WordNet il *sense inventory* di ogni keyword (in questo caso la query si compone di una sola parola chiave). Tra tutti i sensi nel *sense inventory*, l'utente seleziona quello più appropriato.

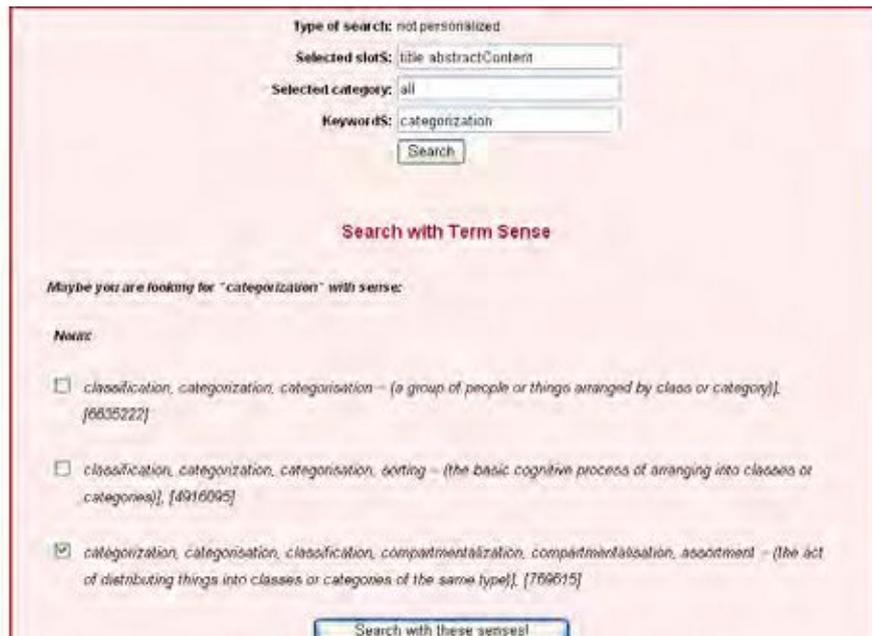


Figura 4 - Scelta del senso per "categorization".

Nello scenario proposto, supponiamo che l'utente sia interessato ad articoli relativi al tema "text categorization", ossia il processo di assegnazione di una categoria (topic), selezionata da un insieme finito di scelte, a documenti appartenenti ad un corpus.

Dunque, il senso più appropriato da associare a "categorization" in questo contesto è il terzo nel sense inventory (Figura 4). Ad ogni articolo ritrovato l'utente può associare un rating di rilevanza, su scala discreta, come mostrato in Figura 5.

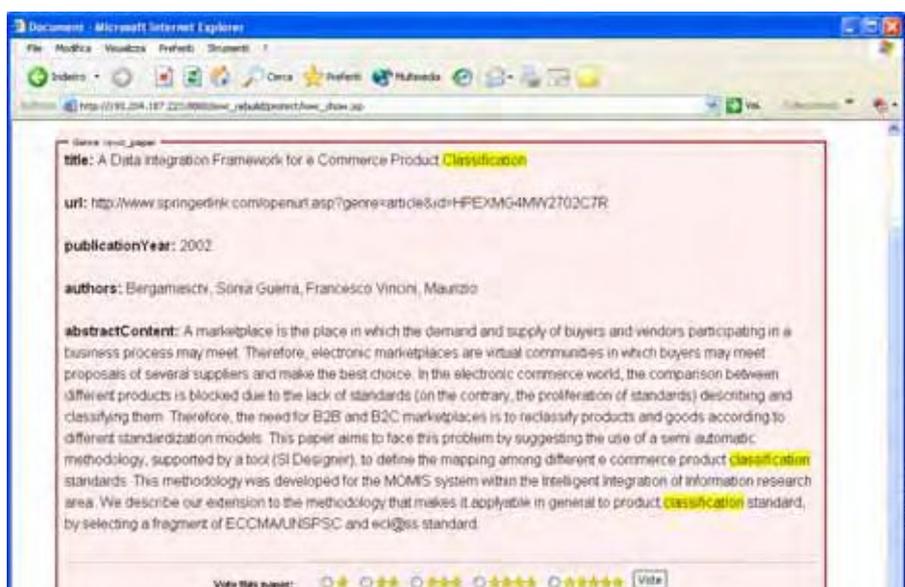


Figura 5 - Uno dei documenti nel result set della query "categorization"

Si può notare come la parola che ha determinato il ritrovamento (“classification”, evidenziata dal motore di ricerca), è diversa da quella usata dall’utente per formulare la query.

Il ritrovamento è avvenuto grazie al fatto che le due parole sono sinonime e pertanto figurano nello stesso synset, consentendo così al sistema di effettuare un *matching* semantico o per concetti. Il servizio di ritrovamento risulta pertanto efficace poiché consente un’accurata selezione degli esempi di addestramento: il documento ritrovato nell’esempio precedente non sarebbe stato ritrovato con una ricerca tradizionale per parole chiave.

Una volta fornito un numero adeguato di esempi di addestramento, il sistema apprende il profilo semantico dell’utente (attualmente la soglia minima per ottenere un profilo accurato è 20) come descritto nella sezione precedente. Nel profilo, sono memorizzati i synset maggiormente rappresentativi degli interessi del partecipante.

Gli articoli nel programma finale di ISWC 2004 sono infine classificati secondo il profilo appreso per ogni utente registrato, al fine individuare gli articoli più interessanti per ognuno di essi, evidenziati nella versione personalizzata del programma inviata individualmente per posta elettronica. Un esempio di programma personalizzato è riportato in Figura 6.

TECHNICAL PROGRAM		
Tuesday 9th		
Opening Ceremony	09:00-09:30	[Room 1]
Invited Speaker: Edward Feigenbaum	09:30-10:30 [chair: F. van Harmelen]	[Room 1]
The Semantic Web Story – It's already 2004. Where are we?		
Coffee Break	10:30-11:00	[Room 4]
Session 1: Integration and Interoperability	11:00-12:30 [chair: M.C. Rousset]	[Room 1]
An Extensible Directory Enabling Efficient Semantic Web Service Integration	Ion Constantinescu, Walter Binder, Bor Fábregas	
Opening Up Maggie via Semantic Services	Martin Dabov, Enrica Motta, John Domingue	
Working with Multiple Ontologies on the Semantic Web	Bernardo Cuenca Grau, Bijan Parsia, Evren Sirin	
Session 2: Searching and Querying	11:00-12:30 [chair: R. Studer]	[Room 2]
A Comparison of RDF Query Languages	Peter Haase, Jean Broekstra, Andreas Eberhart, Raphael Volz	
Generating On the Fly Queries for the Semantic Web: The ICS-FORTH Graphical RQL Interface (GRQL)	Nikos Athanasi, Vassilia Christopidou, Dimitra Kotzinos	
Information Retrieval Support for Ontology Construction and Use	Willem van Hage, Maarten de Rijke, Maarten Marx	
Lunch	12:30-14:00	
Session 3: SW Services	14:00-15:30 [chair: K. Sycara]	[Room 1]
A Conceptual Architecture for Semantic Web Services	Chris Preist	
Applying KaaS Services to Ensure Policy Compliance for Semantic Web Services	Andrzej Uszok, Jeff M. Bradshaw, Renia Jeffers, Austin Tate, Jeff Pflaum	

Figura 6 - Esempio di programma personalizzato per ISWC 2004

## 7 Conclusioni e sviluppi futuri

L'obiettivo del lavoro è stato quello di presentare un framework che realizzi funzionalità di raccomandazione di documenti attraverso la costruzione di profili utente *semantici* inferiti con tecniche di apprendimento Bayesiane. I profili sono definiti *semantici* poiché ottenuti da documenti indicizzati con i concetti riconosciuti analizzando i termini presenti nei documenti stessi con una procedura di WSD che sfrutta la conoscenza lessicale di WordNet.

La rappresentazione dei documenti attraverso i concetti (synset) di WordNet ha prodotto un modello di indicizzazione più accurato di quello basato su parole-chiave. Tale modello può essere sfruttato con successo da algoritmi di apprendimento che inferiscono profili utente a partire da documenti considerati rappresentativi degli interessi degli utenti stessi.

È stato mostrato come i profili semantici producano suggerimenti accurati, attraverso la progettazione di un servizio di raccomandazione di articoli scientifici basato sul framework proposto. Il servizio supporta il partecipante ad una conferenza nella selezione degli articoli più rilevanti.

In futuro, si prevede di estendere il modulo Content Analyzer con nuove procedure che sfruttino ontologie di dominio, oltre che WordNet, e di introdurre componenti per il riconoscimento di entità, al fine di produrre un modello di indicizzazione più accurato.

## 8 Bibliografia

- [1] BANERJEE S. AND T. PEDERSEN, An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet, in *CICLing '02: Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*, pages 136-145, London, UK, 2002. Springer-Verlag.
- [2] S. BLOEDHORN AND A. HOTH, Boosting for text classification with semantic features, in *Proceedings of 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Mining for and from the Semantic Web Workshop*, pp. 70-87, 2004.
- [3] N. GUARINO, C. MASOLO, AND G. VETERE, Content-based access to the web, *IEEE Intelligent Systems*, 14(3):70-80, 1999.
- [4] M. LESK, Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone, in *Proceedings of the 1986 SIGDOC Conference*, pp 20-29, 1986.
- [5] B. MAGNINI AND C. STRAPPARAVA, Improving user modelling with content-based techniques, in *Proceedings of the 8th International Conference User Modeling*, pp 74-83. Springer, 2001.
- [6] C. MANNING AND H. SCHÜTZE, *Foundations of Statistical Natural Language Processing*, chapter 7: Word Sense Disambiguation. The MIT Press, Cambridge, US, 1999.
- [7] D. MAVROEIDIS, G. TSATSARONIS, M. VAZIRGIANNIS, M. THEOBALD, AND G. WEIKUM, Word sense disambiguation for exploiting hierarchical thesauri in text classification, in *Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, pp 181-192. Springer, 2005.

- [8] MILLER, G. A., WordNet: a lexical database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39-41.
- [9] MOONEY R. J., ROY L., Content-Based Book Recommending Using Learning for Text Categorization, in *Proceedings of the 5th ACM Conference on Digital Libraries*, San Antonio, USA, pp. 195-204. 2000.
- [10] PAZZANI M., BILLISUS D., Learning and Revising User Profiles: The Identification of Interesting Web Sites, *Machine Learning* 27(3):313-331, 1997.
- [11] P. RESNIK, Disambiguating noun groupings with respect to WordNet senses, in *Proceedings of the Third Workshop on Very Large Corpora*, pp. 54-68. Association for Computational Linguistics, 1995.
- [12] S. SCOTT AND S. MATWIN, Text classification using wordnet hypernyms, in *COLING-ACL Workshop on usage of WordNet in NLP Systems*, pp. 45–51, 1998.
- [13] F. SEBASTIANI, *Machine learning in automated text categorization*. ACM Computing Surveys, 34(1), 2002.
- [14] SEMERARO G., DEGEMMIS M., LOPS P., AND BASILE P., Combining Learning and Word Sense Disambiguation for Intelligent User Profiling, *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence IJCAI-07*, 2856-2861, Morgan Kaufmann, San Francisco, California, 2007.
- [15] I. WITTEN AND T. BELL, The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression, *IEEE Transactions on Information Theory*, 37(4): 1085-1094, 1991.



## **ANALISI MULTIMEDIALE E INDICIZZAZIONE SEMANTICA: L'ESPERIENZA DI PRESTOSPACE**

*Roberto Basili*

Department of Computer Science, Systems and Production  
AI Research Group  
University of Roma, Tor Vergata  
Via del Politecnico 1, 00133 Roma, Italy  
basili@info.uniroma2.it

L'accesso alle risorse culturali è caratterizzato da alcuni aspetti le cui complessità peculiari non emergono in altri settori applicativi delle tecnologie intelligenti. La ricchezza ed eterogeneità delle singole opere in questo ambito costituiscono una sfida importante per la costruzione di risorse e di sistemi in grado di sfruttarne le caratteristiche semantiche e le finalità comunicative. Ne è esempio concreto l'insieme degli archivi radio-televisivi che grandi enti europei hanno la missione di creare e preservare nel tempo. Il progetto europeo PrestoSpace è dedicato allo sviluppo di tecnologie dedicate per il supporto informatico alla conservazione degli archivi multimediali televisivi, radiofonici e audiovisivi che fanno dell'Europa un fulcro culturale globale. All'interno del progetto lo studio e lo sviluppo di una piattaforma per l'accesso e la diffusione di metadati semantici (*Metadata Access and Delivery, MAD*) costituisce quindi un'esperienza importante per lo studio dell'applicabilità di tecnologie proprie dell'intelligenza artificiale ad un consistente insieme di beni culturali.

Il sistema *MAD* è dedicato alla generazione automatica, validazione e diffusione dei materiali d'archivio attraverso la creazione di metadati basata sulla sintesi dei contenuti semantici a partire dai dati digitali. Tecnologie quali il riconoscimento automatico del parlato, l'analisi delle immagini e l'estrazione automatica di informazioni convergono qui all'automazione del processo di documentazione dei dati degli archivi multimediali e trovano un'integrazione armonica nella piattaforma *MAD*. Strumenti intelligenti per il recupero delle informazioni sono quindi un effetto collaterale dei processi di analisi e aprono possibilità nuove all'automazione dei processi di accesso distribuito all'informazione multimediale. Questo lavoro, discutendo il quadro di riferimento tecnologico del progetto, fornisce una possibile lettura dei suoi risultati nell'ambito specifico dei beni culturali a cui questo volume è dedicato.

### **1 Accesso Intelligente all'informazione multimediale in PrestoSpace**

Nell'ambito dell'elaborazione di video ed immagini, viene spesso enfatizzato il cosiddetto "*semantic gap*" tra la semantica di alto livello necessaria all'indicizzazione di materiale audio-visivo e le proprietà di basso livello offerte dall'analisi automatica. Si presenta quindi la necessità di arricchire la semantica disponibile attraverso la fusione di quest'ultima con contenuti provenienti da altre forme espressive associate all'audio ed al video, come il testo ed il parlato. In tal senso gli attuali fornitori di servizi radio-televisivi hanno riscoperto che il valore dei loro archivi audiovisivi e degli approcci intesi al loro recupero può tradursi in un consistente risparmio nei processi di produzione dei loro programmi [1].

In tale contesto i metadati giocano un ruolo centrale. Infatti nello scenario degli archivi radio-televisivi caratterizzare informazioni e strumenti di ricerca ad alto livello è necessario per permettere agli utenti interessati di ritrovare efficacemente il materiale audio-visivo desiderato con significativi livelli di accuratezza [2], [3].

In tale dominio sono stati identificati quattro principali schemi di ricerca del materiale audiovisivo (AV):

- Ricerca di dati AV attraverso le informazioni. La ricerca in tal caso si basa sull'utilizzo dei vincoli imposti dai metadati sul materiale audio-visivo, utilizzati come specifiche selettive per l'identificazione del materiale richiesto. Questa modalità rappresenta il tipico metodo di utilizzo dell'informazione come "metadato".
- Ricerca di informazioni attraverso dati AV. L'accesso alle informazioni degli archivi in tal caso si basa sul materiale audio-visivo utilizzato come vettore dell'informazione a cui gli utenti sono interessati.
- Ricerca di informazioni attraverso informazioni. Qui l'informazione richiesta viene acquisita attraverso l'uso di altre informazioni che agiscono da "metadati" rispetto all'informazione target.
- Ricerca di dati AV attraverso altri dati AV. Il materiale audio-visuale è identificato e recuperato definendo e sfruttando il concetto di similarità all'interno del materiale stesso, cioè trascurando il significato ed i contenuti espressi.

I partner del gruppo di lavoro MAD (*Metadata Acquisition and Delivery*) nel progetto PrestoSpace<sup>1</sup> hanno svolto un'approfondita analisi dei temi e metodi sopra elencati, arrivando alla conclusione che l'informazione richiesta ai fini dell'utilizzo tipico di un archivio audiovisivo può appartenere alle seguenti quattro classi:

- *Informazioni di identificazione*, come titoli, titoli di coda o di testa, informazioni riguardanti la pubblicazione di un programma.
- *Informazioni sulle componenti editoriali di un programma*, quali le informazioni sulle sottovoci editoriali rilevanti (per esempio le notizie individuali in un telegiornale)
- *Informazioni legate al contenuto*, ad esempio le trascrizioni testuali del parlato, argomenti, descrizioni, caratteristiche descrittive di basso livello di materiale audio e video.
- *Informazioni di arricchimento*, cioè informazioni che provengono da fonti esterne, legate ai contenuti del programma in maniera generica o specifica.

### **1.1 Acquisizione e Disseminazione dei Metadati (MAD): una rassegna**

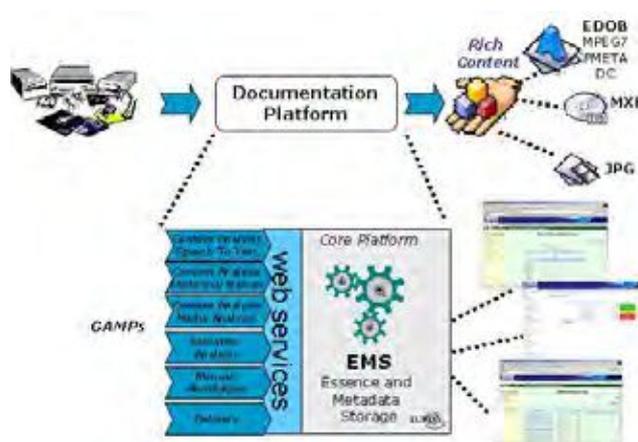
Al fine di soddisfare i requisiti sopra descritti, la Piattaforma MAD è stata dotata di un'architettura modulare e scalabile. Come mostrato in Figura 1, la *Piattaforma Documentale MAD* acquisisce file multimediali (audio e video) in ingresso e produce in uscita i key frames rilevanti, i movimenti camera ed i metadati semantici. Questo materiale opportunamente organizzato ed indicizzato viene pubblicato sul Web server che ospita la *Piattaforma di Pubblicazione MAD*.

---

<sup>1</sup> Prestospace è il progetto integrato FP6-507336 del Sesto Framework Europeo, per la innovazione tecnologica nell'accesso e la valorizzazione del patrimonio culturale europeo (<http://prestospace.org/project/index.en.html>).

La Piattaforma Documentale MAD è composta da un componente centrale, la *Core Platform*, e da un set di estensioni denominate *GAMP* (cioè “*Generic Activity MAD Processor*”). I servizi offerti sono i seguenti:

- il servizio di *Workflow Management*, per l’attivazione dei processi e per la gestione delle dipendenze tra i processori *GAMP*;
- il sistema di *Essence and Metadata Storage* (EMS), per l’archiviazione (delle sorgenti) del materiale audio-visivo e dei metadati associati;
- il servizio di *Concurrent Versioning*, per il tracciamento delle modifiche ai metadati svolte durante i processi gestiti dai *GAMP*; il tool che offre tale servizio è costruito su un motore standard CVS (*Concurrent Versioning System*);
- il servizio di *Delivery* per l’accesso ai metadati descritti semanticamente e a tutto il materiale ad essi connesso prodotto dalla Piattaforma Documentale.



**Figura 1 - MAD Platform: l’architettura**

Il tool EMS archivia il materiale su file system, e ne traccia la posizione attraverso un database relazionale. È possibile avere molte istanze dello stesso materiale, anche posizionato su diverse macchine, e accessibile tramite protocolli diversi (es. file, HTTP, FTP, SMB ...). I processori *GAMP* sono moduli software che estraggono metadati dal materiale digitalizzato. La *Core Platform* gestisce nel *workflow* le code per ogni *GAMP*; questi ultimi quindi attiveranno chiamate alla piattaforma per avere informazioni sulle attività da svolgere e per ricevere il materiale multimediale ed i relativi metadati prodotti fino al momento della richiesta.

L’attuale configurazione sperimentale della Piattaforma MAD utilizza tre diverse categorie di *GAMP*: Analisi del Contenuto, Analisi Semantica e Annotazione Manuale, ma grazie alla configurazione modulare dell’architettura MAD si può all’occorrenza prevedere l’estensione a nuovi *GAMP* di altro tipo.

I servizi offerti dalla Piattaforma sono resi disponibili attraverso le interfacce SOAP ai seguenti servizi Web: il *Workflow*, l’*EMS* e l’*Administration*. Attraverso questi servizi ogni *GAMP* può chiamare la *Core Platform* richiedendo un *job* e le relative risorse. Non appena completato il *job*, i dati prodotti vengono inviati e viene notificato il compimento del *job* al *Workflow Manager*. L’uso dei servizi Web permette di sviluppare i *GAMP* con linguaggi di programmazione diversi e di instanziarli su piattaforme e sistemi operativi diversi tra loro.

I vantaggi nell'utilizzo di una piattaforma di questo tipo sono i seguenti:

- *modularità*: i GAMP possono offrire funzionalità ed essere sviluppati secondo modalità completamente diverse, continuando quindi ad interagire con la *Core Platform*;
- *scalabilità*: aggiungere un nuovo GAMP implica, a livello architetturale, solo l'aggiunta di una nuova coda di processo al motore di *Workflow* della *Core Platform*;
- *indipendenza dalla piattaforma*: i GAMP possono essere sviluppati in ogni linguaggio di programmazione, a condizione che supportino i protocolli SOAP e Web service;
- *distribuzione a più livelli*: ogni GAMP può essere installato su sistemi fisici diversi, purché esista un *network link* alla *Core Platform*;
- *rilocabilità*: i componenti della Core Platform (cioè l'EMS, il motore di Workflow ed il DBMS) possono essere installati su server diversi, geograficamente distribuiti.

La complessità dei processi sopra descritti richiede la definizione di una architettura e di un *data model* adeguati. Questo *requirement* è stato soddisfatto attraverso la creazione di un formato documentale in XML, basato su due standard orientati su base nativa ai metadati audio-visivi, l'MPEG-7 [4] ed il P\_META [5].

## 2 Elaborazione dei Contenuti Multimediali

I metodi automatici di analisi dei contenuti AV vengono applicati in Prestospace per estrarre automaticamente metadati dal materiale multimediale e arricchire la descrizione del contenuto. I metadati estratti automaticamente sono usati per facilitare l'annotazione manuale durante la documentazione. La struttura del contenuto rilevata fornisce infatti un input alla analisi semantica ed alla indicizzazione degli oggetti AV. Una *survey* dei tool di stato dell'arte per l'analisi dei contenuti visuali, audio o congiunti è discussa in [7].

Un insieme di tool di analisi dei contenuti AV sono stati selezionati, progettati in un processo di documentazione ed integrati nell'infrastruttura MAD descritta sopra. Grazie alla sua architettura modulare ed estensibile, la piattaforma consente di distribuire, attraverso *client* dislocati geograficamente, i carichi di lavoro dei complessi processi coinvolti.

**Estrazione di proprietà visuali di basso livello.** L'estrazione di proprietà a basso livello descrive i fotogrammi chiave (*key frame*), o le scene, in base ai loro colori, *texture* e proprietà del movimento camera. Il tool estrae alcuni dei descrittori specificati nella parte visuale di MPEG7 ([4] parte 3), cioè il *ColorLayout*, il *ColorStructure*, il *DominantColor*, l'*EdgeHistogram* ed il *MotionActivity*. I descrittori forniscono una rappresentazione compatta ed efficiente del contenuto visuale di una scena e sono usati per determinare la similarità visiva tra scene.

**Riconoscimento dei cambi di scena.** Il tool di rilevamento dei cambi di scena segmenta un video nei suoi blocchi costitutivi principali, cioè le sue scene. È capace di rilevare sia dei tagli improvvisi sia delle transizioni graduali (ad esempio la dissolvenza o lo zoom). La segmentazione in scene è un requisito per altri algoritmi di analisi del contenuto video, per la strutturazione e per l'indicizzazione del contenuto, e fornisce anche un supporto alla navigazione in un video per il tool di documentazione manuale. L'approccio usato per il rilevamento dei cambi scena è un miglioramento del modello proposto in [9].

**Estrazione di *key frame* e *stripe image*.** Il riconoscitore dei *key frame* ne estrae un numero per ogni scena a seconda della quantità di cambiamento visuale. I *key frame* forniscono una rappresentazione sintetica delle scene e sono usate come ingresso per l'estrazione di proprietà di basso livello. Le *stripe image* invece sono rappresentazioni spazio-temporali del livello visuale, create dal contenuto di una colonna, fissa o in movimento, a partire dall'evolversi nel tempo dell'immagine. Forniscono un sostegno per una *survey* rapida del contenuto di una scena e per la navigazione, specialmente durante l'attività di documentazione manuale.

**Classificazione dei movimenti camera.** Tale processo analiticamente descrive 4 tipi di movimento rispetto al contenuto (*pan, tilt, zoom e roll*), una quantificazione approssimata dell'ammontare del movimento e della lunghezza dei segmenti in cui tali movimenti compaiono. L'algoritmo è basato sul tracciamento di proprietà visuali. L'approccio descritto ed i risultati ottenuti si trovano in [8]. Le informazioni riguardo i movimenti camere sono un criterio di ricerca molto importante nel riuso di materiale di archivio per nuove produzioni ed anche per l'acquisizione di informazioni di livello più alto.

**Strutturazione e segmentazione dell'audio.** Quest'analisi consiste nella classificazione dei segmenti audio in quattro categorie principali: silenzio, musica, parlato e rumore. Le informazioni prodotte sono principalmente considerate come supporto all'annotazione manuale.

**Segmentazione in parti editoriali.** Le parti editoriali sono considerate dagli archivisti televisivi moderni come le entità di base indispensabili per la documentazione di un programma d'archivio. Esse possono essere definite come le parti costituenti di un programma dal punto di vista editoriale, cioè quello dei creatori del programma, ad es. l'autore, il regista o il curatore. Nel caso dei telegiornali, esse coincidono con le notizie individuali. Sono state studiate molte tecniche per risolvere il difficile problema dell'identificazione di parti editoriali a partire dall'analisi delle proprietà AV di basso livello [7], sebbene nessuna di esse abbia per ora risolto il problema generale. In Prestospace, la segmentazione automatica è stata limitata alle notizie radiotelevisive, scegliendo un approccio multistrato che combina informazioni video ed audio finalizzato al rilevamento delle notizie all'interno delle edizioni di telegiornali generali.

**Riconoscimento di frammenti video di riferimento.** Ha il compito di riconoscere delle repliche di un clip di riferimento all'interno del flusso di un contenuto visivo. Questa semplice attività è molto utile se applicata a materiale di archivi televisivi ove i *jingle*, le barre di colore ed altri effetti sono usati come separatori visuali tra le componenti di un programma. L'informazione prodotta costituisce uno degli input per il ben più complesso *task* della segmentazione editoriale.

### **3 La elaborazione semantica dei dati multimediali**

La piattaforma MAD è finalizzata a sfruttare le tecnologie del linguaggio per l'*Information Extraction* dai dati audiovisivi resi disponibili da archivi su larga scala. La natura e la complessità della gestione, della ricerca e del riuso dei materiali di archivio richiedono funzionalità complesse di memorizzazione e recupero. Per queste attività si richiedono quindi:

- ❑ il riconoscimento e l'indicizzazione di *opportune concettualizzazioni* di interesse per l'archivio, come le entità classificate in persone, organizzazioni o luoghi;
- ❑ funzioni efficaci di *retrieval* che migliorino l'indicizzazione al semplice livello testuale e supportino il recupero dei dati in base ai concetti piuttosto che alle stringhe di caratteri;

- *interoperabilità* ai livelli di astrazione richiesti dal contenuto AV. Per esempio, i dati AV dovrebbero essere pubblicati, interrogati, scambiati in modo distribuito. Lo sviluppo di applicazioni Web per la pubblicazione dovrebbe rendere disponibili l'interrogazione, nonché l'istanziamento ed invocazione distribuite di servizi di tipo semantico. Le descrizioni dei dati semantici sono critiche in queste attività e si rendono quindi necessari modelli interoperabili (ontologie).

L'analisi semantica viene applicata in MAD per soddisfare i requisiti di astrazione complessi, dalle proprietà multimediali (per esempio, audio) sino alla costruzione di generalizzazioni e rappresentazioni di tipo ontologico.

Il contributo dell'Information Extraction nell'area del *Semantic Web* è la realizzazione del processo che va dai dati grezzi e testuali alle annotazioni ontologiche. Il punto di partenza dell'analisi semantica è quindi il *riconoscimento automatico del parlato* (ASR) dal contenuto dei dati audio. L'estrazione di testo dal contenuto parlato del materiale AV è un passo fondamentale per diversi task di documentazione, e inoltre fornisce un nucleo base di dati utili alla ricerca nella piattaforma di pubblicazione. Nel prototipo corrente della piattaforma di documentazione è stato applicato un motore di ASR, discusso in [6], in grado di estrarre testi da programmi in Inglese ed Italiano.

La ridondanza che oggetti AV garantiscono a livello dati deve essere esplorata per gestire la difficoltà complessiva del *retrieval* di informazioni soddisfacendo l'accuratezza richiesta. Tali problemi, dovuti alla natura rumorosa dei dati estratti (ad esempio gli errori nel motore di ASR, che produce informazione errata o fuorviante per il riconoscimento grammaticale), debbono poter essere opportunamente corretti. Qui l'obiettivo è estrarre un'informazione la più ricca possibile attraverso il processo complessivo di estrazione, e mettere quindi a disposizione delle componenti di ricerca di MAD un'ampia varietà di risorse. In questa prospettiva dovrebbero poter essere considerati anche insiemi di evidenze più ampie che non quelle rese disponibili dal mero dato AV sorgente. Le trascrizioni in ingresso possono essere elaborate ed arricchite da un insieme più ampio di informazioni di interesse, rappresentate come metadati semantici:

- le informazioni lessicali e terminologiche locali al dato AV in input (cioè pronunciate esplicitamente nel parlato);
- il riferimento (citazione) *a nomi di persone, luoghi o organizzazioni (Named Entities)* esplicito nel dato AV sorgente o derivabile dalle sorgenti esterne;
- la generazione automatica di *hyperlink* significativi che colleghino i dati AV e le sorgenti distribuite (per esempio, *link* dalle notizie del telegiornale al portale, o a pagine, di un giornale su Web). Tali sorgenti forniscono descrizioni testuali, consolidate e grammaticalmente corrette, di temi ed eventi collegati ai segmenti AV individuali;
- informazione ontologica, come classi/concetti (ad esempio, luoghi geografici o organizzazioni), gli individui (*John Coltrane* o *USA/United States*) ed infine le classi tematiche (*Istruzione, Economia, Sport* o *Politica Estera*).

L'estrazione, richiesta in MAD, di quest'ampia e ricca varietà di informazioni è l'obiettivo di GAMP specializzati, chiamati *Semantic Analysis GAMPs* (SA\_GAMP). I GAMP di questo tipo sono dipendenti dalla lingua, cosicché due diversi SA\_GAMP sono stati progettati rispettivamente per le sorgenti AV in Italiano ed in Inglese. Nel seguito si userà il processore SA\_GAMP per l'Italiano come esempio di riferimento durante la discussione, mentre i dettagli tecnici del SA\_GAMP per l'Inglese si trovano in [17].

### 3.1 Analisi Semantica in MAD

In MAD viene utilizzata una cascata di processi per l'arricchimento delle parti editoriali rappresentate in forma multimediale con metadati semantici. Tutti questi processi sono organizzati e sincronizzati dal *WorkFlow manager*. Questo modulo invoca i processori in accordo con le loro dipendenze così come mostrato in Figura 2. I diversi componenti coinvolti qui sono:

- un *Intaker* che gestisce il caricamento del programma e normalizza la sequenza dei segmenti forniti dalle trascrizioni in ASR coinvolte
- un *Classificatore Tematico*, che assegna una categoria specifica a ciascuno dei segmenti individuali in ingresso
- un *Parser grammaticale* ([10]), che riconosce le unità lessicali all'interno delle trascrizioni e fornisce un primo livello di informazione morfosintattica attraverso operazioni di *tagging* grammaticale (*POS tagging*)
- un *Classificatore di nomi propri* (*Named Entity, NE, recognizer*) che riconosce le citazioni a luoghi, persone o organizzazioni o ad altre entità interessanti (per esempio, le date)
- un *Analizzatore ontologico* che collega le citazioni a NE (cioè individui riconosciuti nel testo) alle entità (istanze) note nella ontologia di riferimento (vedi Sezione 3.2)
- un *Web aligner* che ricerca e seleziona le pagine Web che descrivono eventi correlati (o equivalenti) a quelli discussi nei segmenti AV in ingresso.

Il processo generale di analisi semantica coincide quindi con una cascata di diversi moduli di estrazione automatica. Il primo modulo (*Intaker*) colleziona e normalizza le notizie individuali del programma così come esse sono trascritte e segmentate dal GAMP di ASR. Poi viene invocato il *classificatore* (discusso in [18],[19]) che assegna una categoria tematica ad ogni notizia individuale (associandola ad un valore di confidenza), in base allo schema di classificazione prescelto.

In Prestospace, il classificatore implementa lo schema editoriale utilizzato dalla RAI e vengono riconosciute le classi giornalistiche principali (ad es. *Politica Estera*). In modo concorrente, le notizie possono essere analizzate grammaticalmente dal *parser* ([10]) per la generazione delle informazioni morfo-sintattiche che abilitano il *NE recognition* nel testo di persone, luoghi, organizzazioni o date.

Queste informazioni costituiscono un insieme di metadati molto significativi per il modulo di *Web alignment* che ricerca notizie correlate in sorgenti esterne. Una notizia AV infatti può trattare gli stessi eventi di una pagina Web e questa può essere facilmente accessibile tramite un processo di *spidering*. La similitudine tra le pagine Web e la notizia AV in esame dipende strettamente dagli individui, fatti luoghi riconosciuti nel dato AV così come dall'allineamento temporale tra le due. Le pagine scaricate sono per questo elaborate grammaticalmente ed indicizzate ad intervalli temporali regolari o *on-demand*. Una notizia AV in esame è così allineata ad una nel Web, selezionando quest'ultima tra le pagine dall'insieme dei candidati Web già recuperati. Il criterio di similitudine specifico dipende dalle NE rilevate, dalla distanza temporale e dall'analogia tematica tra le due notizie ([11]). Un allineamento accettato genera un nuovo (meta)dato della notizia AV che rappresenta un *hyperlink* diretto alla controparte Web (cioè la URL) della(e) notizia(e) esterna.

È da osservare che i contenuti di tale pagina “allineata” vengono anche sfruttati per estendere i metadati interni cioè quelli derivanti dalle sole informazioni dell’oggetto AV in esame: una pagina Web infatti contiene un testo completo e grammaticalmente valido. Allo scopo di migliorare l’accuratezza complessiva dell’analisi semantica, i metadati riguardati gli individui o i luoghi della notizia sono quindi fusi con quelli rilevati nella sorgente esterna: in questa fase viene quindi ridefinito (esteso o corretto) l’insieme dei metadati interni (potenzialmente incompleto a causa degli errori nel processo di ASR).

L’analizzatore ontologico infine viene applicato per allineare l’annotazione semantica di ogni notizia individuale ad una ontologia di riferimento: in Prestospace questo processo è gestito dalla piattaforma KIM, [13], discussa nella sezione successiva. Gli ulteriori dettagli riguardo il processo di analisi semantica, che la sintesi perseguita in questo lavoro non ci consente di discutere, sono presentati in [11]. Un lavoro che discute l’analizzatore semantico per la lingua inglese è [17].

### 3.2 Il ruolo dell’informazione ontologica

La componente ontologica in Prestospace è gestita dalla piattaforma KIM [14], che fornisce un servizio di *information extraction* in accordo con una ontologia ed un base di conoscenza estensiva. La piattaforma sostiene un processo di annotazione semantica come modello innovativo di arricchimento dei contenuti semantici [3], abilitando metodi complessi di recupero dell’informazione che estendono i modelli tradizionali di *Information Retrieval*. In questo modo KIM fornisce le applicazioni come l’*highlighting* delle occorrenze di concetti ontologici in un testo, l’indicizzazione, il retrieval la classificazione, la generazione di metadati di sintesi e la navigazione amichevole attraverso i testi (scarsamente strutturati) e basi di conoscenza di dominio.

L’approccio all’IE impiegato da KIM è motivato dal concetto che le entità a cui un contenuto D fa riferimento sono di significativa importanza per la rappresentazione del significato di D. Per chiarire il ruolo dei nomi propri nella semantica di un documento, si può osservare il semplice esempio di una frase come “*il primo presidente degli Stati Uniti*”: la comprensione del significato delle parole costituenti qui non è sufficiente a catturare correttamente tutti i significati dell’espressione.

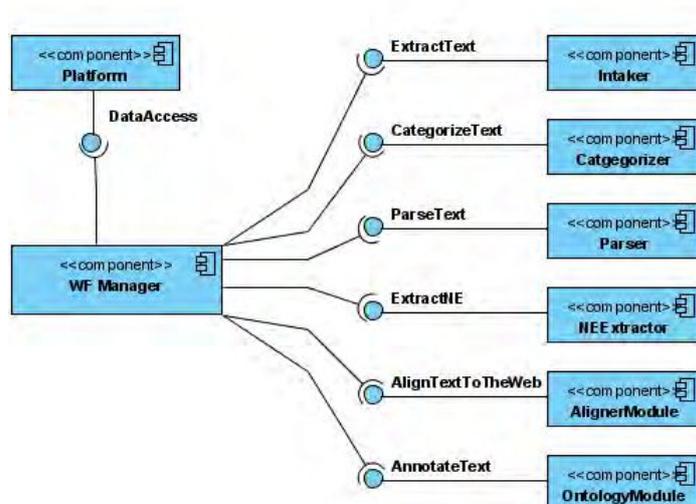


Figura 2 - Struttura del processore semantico per l’italiano, Ita SA GAMP

A differenza delle parole, i nomi propri (cioè le *Named Entities*) denotano individui concreti e non una classe o un membro generico di una classe. Mentre la semantica lessicale è di aiuto per riconoscere il significato delle parole, la comprensione di un nome proprio richiede una conoscenza del mondo non banale. L'annotazione semantica è quindi quel processo che assegna a tutti i nomi propri di un testo i legami appropriati con le loro descrizioni semantiche nell'ontologia.

Il processo di annotazione semantica è basato su un modello semplice delle classi di entità del mondo reale, cioè una ontologia e di una base di conoscenza estensiva. La rappresentazione delle annotazioni semantiche e dei metadati corrispondenti ha certi prerequisiti: (1) essi richiedono una ontologia (almeno una tassonomia) che definisca le classi delle entità (di interesse) (2) identificatori univoci che consentono alle entità di essere distinte e collegate alle loro descrizioni semantiche; infine, (3) una base di conoscenza che contenga la descrizione esplicita delle entità individuali. KIM si basa su due tipi di ontologia: una di tipo indipendente dal dominio, detta di *upper-level* (PROTON<sup>2</sup>), ed un'altra specializzata rispetto al dominio applicativo.

L'ontologia *upper-level* di PROTON codifica gli aspetti più comuni di una descrizione che può assumersi indipendentemente dalla specificità di un dominio (ad esempio, scienza, meteorologia o legge) e dalla specificità del *task* in esame, come ad esempio la classificazione dei film, l'accesso alle notizie d'agenzia o la descrizione dei temi di un documentario. PROTON è stato progettato per soddisfare il requisito di essere facilmente adattabile agli scopi della annotazione semantica su domini "aperti" (come il Web), così come per essere facilmente esteso ed adattato ad eventuali domini ristretti. Attualmente contiene circa 300 classi e 100 proprietà.

Per l'annotazione, l'indicizzazione ed il recupero dei documenti, KIM usa anche una KB di base. In questo contesto essa forma un insieme di assiomi formali riguardanti le entità, come strumento per le asserzioni di tipo non ontologico. Consiste di dati riguardanti le istanze, le descrizioni di entità e delle loro relazioni (individuali), alcune informazioni riguardanti la natura delle entità, come gli *alias*, che includono la denominazione principale (cioè il nome ufficiale o meglio conosciuto), ed infine la descrizione di attributi e relazioni. La KB di KIM fornisce la copertura di entità popolari nel dominio applicativo di interesse, molto comuni e ben conosciute, che non debbono essere introdotte esplicitamente nei documenti. Le più importanti ed usate entità nella KB sono, ad esempio, i nomi geografici e le organizzazioni. Queste rappresentano unità geografiche importate da *GNS* (*GEOnet Names Server*) o da altre sorgenti. Sono organizzate in modo da rappresentare istanze di *Location* (o sue sottoclassi) aventi la proprietà *subRegionOf* che sussiste tra *Continents*, *GlobalRegions*, *Countries*, e altre sottoclassi di *Location*. Alcuni sottotipi di *Location* contenute in KIM sono *Country*, *Province*, *County*, *CountryCapital*, *City*, *Ocean*, *Sea*. La rappresentazione ontologica delle entità in KIM avviene attraverso un URI (*Universal Resource Identifier*) che è, per sua natura, indipendente dalla lingua.

Le descrizioni ontologiche della KB forniscono una base molto utile per le interrogazioni ed il recupero automatico di dati di tipo cross-linguistico. Il ruolo di KIM in MAD è fornire una rappresentazione indipendente dalla lingua per le entità e questo tipo di metadati è un tipicamente condivisibile attraverso lingue diverse.

---

<sup>2</sup> PROTON: see <http://proton.semanticweb.org/>

Ad esempio, "*White House*" è tradotto in italiano con "*Casa Bianca*", che ne costituisce un alias. L'identificazione dell'entità in una notizia, piuttosto che il *pattern matching* con un *alias*, fornisce un criterio più sistematico per l'indicizzazione. Questo garantisce un supporto più espressivo per il trattamento di lingue diverse nell'interrogazione per l'accesso (multilinguale) dei dati multimediali: informazioni originate da lingue diverse potranno essere recuperate attraverso il recupero delle stesse entità.

#### 4 Il recupero dell'informazione in MAD

L'ampia varietà delle informazioni estratte dai GAMP impone requisiti alle funzionalità di *Information Retrieval* nella fase di pubblicazione. Anzitutto l'interfaccia utente dovrebbe fornire metodi di accesso dipendenti da funzionalità diverse ed integrate:

- ricerca *full-text* come generalmente applicata dai più popolari motori di ricerca;
- domande in linguaggio naturale;
- *browsing* semantico inteso come navigazione attraverso concetti, relazioni ed istanze dell'ontologia.

Tutte queste funzionalità sono intese come indipendenti dal linguaggio: le ricerche *full-text* debbono essere possibili in lingue differenti mentre le informazioni ontologiche, come la ricerca di *Named Entities* dovrebbero essere opportunamente presentate in modo che la variabilità e le ambiguità proprie di ogni lingua siano controllate. Inoltre tutte le funzionalità di ricerca dovrebbero essere fornite in modalità indipendente dalla lingua, poiché i dati multimediali disponibili negli archivi AV derivano da sorgenti spesso sovra-nazionali e quindi sono multilingui.

Una discussione dedicata delle soluzioni tecnologiche di supporto ai processi di IR sopra descritti viene riportata nelle sezioni successive, poiché queste hanno un impatto rilevante sull'accuratezza che tecnologie di *Cross-lingual Information Retrieval (CLIR)* possono raggiungere attraverso l'approccio proposto in PrestoSpace.

Le soluzioni applicabili al problema si riferiscono all'adozione di una rappresentazione dei metadati il più possibile indipendente dalla lingua, ed all'elaborazione delle *query* (attraverso la loro espansione e traduzione) che consenta il recupero di metadati specifici per una lingua *target* diversa dalla lingua *sorgente* che caratterizza una *query*.

##### 4.1 Recupero di informazioni attraverso l'ontologia

In MAD la piattaforma KIM [14] rende disponibile tutta l'estesa conoscenza ontologica relativa al dominio delle notizie e fornisce funzionalità di navigazione e ricerca.

KIM offre un'infrastruttura innovativa per la gestione dell'informazione e della conoscenza ed un certo numero di servizi per l'annotazione semantica automatica, l'indicizzazione ed il *retrieval* di dati e documenti non strutturati o semi-strutturati. La differenza rispetto ad altri sistemi ed approcci simili risiede principalmente nel fatto che, offrendo annotazioni semantiche, KIM fornisce anche supporto ai servizi di IR basati su tali informazioni. Inoltre, diverse interfacce utente sono rese disponibili in KIM attraverso API dedicate (KIM API's), che forniscono funzionalità ed infrastrutture per l'annotazione semantica, il *retrieval*, la documentazione e la navigazione attraverso le basi di conoscenza.

L'interfaccia Web di KIM (KIM Web *User Interface*, Figura 3) consente metodi di accesso ai tradizionali basati su parole chiave e metodi di accesso semantici, quali la ricerca di *entità* o di *pattern relazionali*.

Scopo della ricerca di tipo semantico è la richiesta di entità specifiche eventualmente ristrette da vincoli formali rispetto alle loro proprietà (come ad esempio “*Quali aziende del settore agro-alimentare agiscono nell’Europa dell’Ovest?*”). Questo avviene attraverso la compilazione di speciali schemi (detti *template*) che consentono di esplicitare facilmente l’insieme dei vincoli agenti su una certa classe di entità (Figura 3).

In tali casi, l’interfaccia garantisce l’accesso ad insiemi di entità (istanze) che soddisfano i vincoli oppure ad insiemi di documenti in cui *si parla di* tali entità. L’utente può accedere al contenuto di un documento recuperato, arricchito dei relativi metadati (es. titolo, data di pubblicazione o altre annotazioni semantiche).

Un *plug-in* per il browser Internet Explorer è disponibile per visualizzare l’annotazione semantica esplicita delle entità (e dei loro tipi) presenti e riconosciute nelle pagine Web mostrate dal browser. Un *tab* specifico viene usato per mostrare contemporaneamente la gerarchia dei tipi di entità (cioè un ramo dell’ontologia di KIM, in Sez. 3.2); uno specifico colore viene associato ad ogni tipo di entità, per enfatizzare le annotazioni di tale tipo nel *layout* della pagina del browser. *Check-box* associati ai tipi consentono all’utente di selezionare o nascondere i tipi di entità (e colori associati per l’evidenziazione) diversi. In questo modo l’utente può navigare il testo passando direttamente dalle annotazioni alle corrispondenti descrizioni ontologiche presenti nelle basi di conoscenza (KB) ad esse connesse. Attraverso questo meccanismo è possibile esplorare ulteriormente le KB, selezionando un’entità individuale e accedendo alle sue proprietà formali note o attraversando le classi di entità dell’ontologia PROTON. Ulteriori dettagli riguardanti la tecnologia KIM per il *retrieval* di informazioni attraverso l’ontologia PROTON sono discussi in [14, 13].

#### 4.2 Cross-Language Information Retrieval

Il trattamento di interrogazioni in linguaggio naturale per il recupero di metadati specifici per una lingua target  $L_2$ , diversa dalla lingua sorgente  $L$  che caratterizza l’interrogazione è detta *Cross-Language Information Retrieval (CLIR)*. Il CLIR è garantito nella piattaforma di pubblicazione del sistema MAD da un server specifico, detto *CLIR server* (Figura 4), che include diverse componenti:

- Il *Parser Grammaticale* che estrae nomi propri dalla query  $q$  nel lingua sorgente  $S$ ;



Figura 3 - Interrogazioni Ontologiche in KIM

- il *Generatore di pseudo-contesti* che produce, per ogni oggetto lessicale della query  $q$ , i termini più rilevanti ad esso tematicamente correlati nella lingua sorgente  $S$ ;
- il modulo di *Sense Detection* che elimina l'ambiguità dei nomi comuni del linguaggio sorgente  $S$  presenti nella query  $q$ ;
- il *Translator* che traduce i nomi comuni (ormai non più ambigui) di  $q$  dalla sorgente  $S$  alla lingua target  $T$ ;
- il *server KIM* che riconosce le entità ontologiche nella query  $q$ ;
- il *Classificatore Tematico* che classifica la query  $q$  in uno dei domini di interesse.

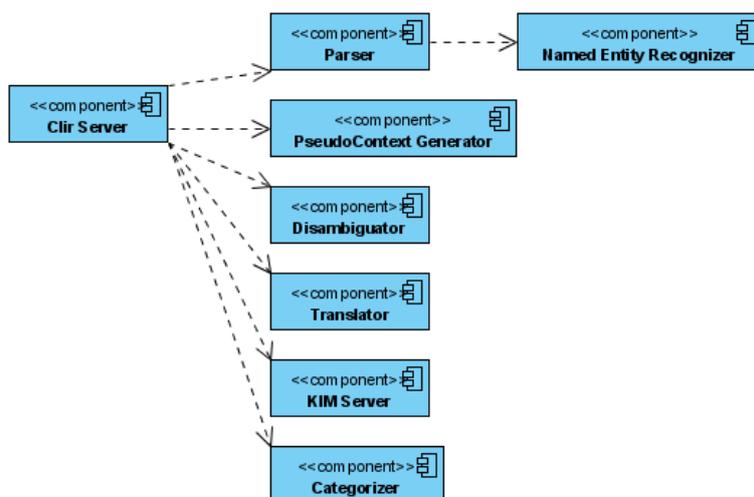
Il server CLIR comunica con queste componenti e gestisce il *workflow* interno come descritto in Figura 4. Il parser, KIM ed il classificatore tematico sono gli stessi servizi utilizzati nel processo di documentazione dal GAMP di analisi semantica (SA\_GAMP).

Un aspetto distintivo del server CLIR è la tecnica adottata nell'eliminazione delle ambiguità di senso e per la traduzione dell'interrogazione  $q$ . La traduzione di tutti i nomi comuni in  $q$  è necessaria poiché questi sono particolarmente significativi e debbono essere consistentemente riscritti in una rappresentazione *language-independent*. Per questo la nozione interlingua adottata è la nozione di *senso* (*word sense*), così come definita nella rete semantica Wordnet e adottata in numerosi progetti. L'algoritmo di disambiguazione del senso agente sulla rete Wordnet è presentato in [12].

L'approccio adottato per la disambiguazione è quello di (1) estendere automaticamente una query in  $S$  attraverso una tecnica di *text mining*, (2) determinare il senso in Wordnet per ogni nome comune presente in  $q$  e (3) utilizzare tale senso per scegliere la traduzione corretta in  $T$ . In questo modo l'opportuna espressione degli elementi salienti di una query in  $S$  viene riprodotta in  $T$ .

In particolare un processo di espansione della *query* viene inizialmente applicato attraverso un approccio di tipo *Latent Semantic Analysis* [15]. La *query* iniziale  $q$  è prima trasformata in uno spazio LSA (ottenuto precedentemente dall'analisi di collezioni di notizie in entrambe le lingue). Questo consente di associare a tutti i nomi in  $q$  i termini più vicini, e cioè un lessico  $dom(q)$  associato al dominio tematico di  $q$ . Tale fase è detta di *Pseudo Context Generation*. All'interno di questo lessico viene applicato un processo di disambiguazione del senso: una metrica di similarità  $n$ -aria [16] è utilizzata per ordinare i sensi di WordNet dei nomi in  $q$  a partire dal  $dom(q)$ . Poiché l'ambiguità di senso diminuisce all'interno di un dominio, il processo di eliminazione dell'ambiguità è molto efficace. I sensi preferiti vengono quindi utilizzati per generare traduzioni nella fase di *Translation*. Le interfacce multilinguali di WordNet che collegano i sensi delle diverse lingue alle loro espressioni lessicalizzate possono quindi essere sfruttate. I migliori sensi, cioè i *synset* scelti per i nomi di  $q$  in  $S$ , sono quindi selezionati e le traduzioni lessicali proposte da Wordnet nella lingua target  $T$  vengono derivati come traduzioni possibili.

La *query* è ottenuta dalla composizione delle *Named Entities*, riconosciute dal parser, della classificazione tematica della *query*, delle identità ontologiche riconosciute ed infine di tutti i sinonimi dei nomi originali (nel linguaggio sorgente  $S$ ) espressi nel linguaggio d'arrivo  $T$ . È importante notare che il metodo è applicabile in modo totalmente automatico, poiché l'analisi LSA, così come la metrica di similitudine, utilizzate per il processo di riconoscimento del senso (in  $L$ ) sono applicabili senza la richiesta esplicita di esempi o di annotazioni manuali.



**Figura 4 - Il Servizio di Cross-Language Information Retrieval**

L’algoritmo di riconoscimento combinato con LSA qui discusso è riportato in [12]. Ad esempio una *query* come la seguente:

*Blair calls on NATO member to contribute more troops to Afghan force*

è elaborata come riportato nella successiva Tabella. Le traduzioni individuali dei nomi comuni (inglesi) prodotte dal CLIR server sono mostrate in grassetto, distinguendoli dagli altri (potenzialmente molteplici) sensi attivati da WordNet per ogni nome. Il risultato dell’interrogazione viene mostrato nell’ultima riga, separando le entità ontologiche (ad es. Person:*Blair*) dalle traduzioni dei nomi comuni.

<b>Input Query</b>	<i>Blair calls on NATO member to contribute more troops to Afghan force.</i>		
<b>Parser NEs</b>	Blair [person] NATO [organisation]		
<b>KIM types</b>	NATO [mp7:OrganizationType] Blair [mp7:PersonType]		
<b>Nouns, Translations</b>	<b>Noun</b>	<b>Sensi in L</b>	<b>Sensi in L2</b>
	NATO	<b>North_Atlantic_Treaty_Organization, NATO</b>	<b>n.a.t.o., organizzazione_del_trattato_nordatlantico</b>
	<i>member</i>	member penis, phallus, member member extremity, appendage, member <b>member</b>	componente, membro asta, fallo, membro, membro_virile, pene, verga appartenente, componente, iscritto, membro arto, estremita', membro <b>Membro</b>
	<i>troops</i>	<b>military_personnel, soldiery, troops</b>	<b>Truppa</b>
	<i>force</i>	Force <b>military_unit, military_force, military_group, force</b> violence, force effect, force force, persone force, forcefulness, strength	Forza <b>Arma</b> forza, violenza effetto, forza forza, personale corpo, energia, forza, lena
<b>Query in L2</b>	Person:Blair & Organization:Nato & (n.a.t.o   “organizzazione del trattato nordatlantico”) & membro & truppa & arma		

È interessante notare l'efficacia della disambiguazione anche relativamente a parole (ad es. *member* o *force*) la cui polisemia è piuttosto alta (5 e 6 sensi diversi, rispettivamente). Inoltre, nel caso di *member* va notato come tra i due sensi più attinenti in Wordnet ("*appartenente, componente, iscritto, membro*" e "*membro (collettivo)*") rispettivamente in terza e quinta riga), il sistema scelga la nozione di *membership* collettiva, più adeguata per il "*NATO members*" usato nella query in inglese.

Un ulteriore esempio dall'Italiano all'Inglese è mostrato nella tabella seguente.

<b>Input Query</b>	<i>Berlusconi al parlamento sulla missione di guerra in Iraq.</i>		
<b>Parser NEs</b>	Berlusconi [person] Iraq [paese]		
<b>KIM NEs</b>	Iraq [mp7:CountryType]		
<b>Nouns, Translations</b>	<b>Noun</b>	<b>Sensi in L</b>	<b>Sensi in L2</b>
	<i>parlamento</i>	<b>parlamento</b>	<b>Parliament</b>
	<i>missione</i>	<b>delegazione, deputazione, missione, rappresentanza</b> missione	<b>deputation, commission, delegation, delegacy, mission</b> mission, military_mission
	<i>guerra</i>	guerra battaglia, combattimento, conflitto, guerra, lotta, scontro discordia, disunione, guerra, zizzania <b>guerra</b>	war, warfare battle, conflict, fight, engagement  discord, strife  <b>strife</b>
<b>Query in L2</b>	Person:Berlusconi & Location:Iraq & parliament & (deputation   commission   delegation   delegacy   mission) & strife		

Il processo di CLIR qui proposto, è completamente automatico e rappresenta uno strumento di notevole efficacia, grazie alla sua robustezza ed alla precisione raggiungibile attraverso l'uso combinato con i altri vincoli (ad esempio, canale di trasmissione, data e ora del programma) spesso rilevanti nel recupero cross-linguistico di notizie d'agenzia. Inoltre la sua applicazione combinata con la navigazione ontologica di KIM fornisce per la piattaforma di pubblicazione di MAD uno strumento molto innovativo e flessibile per l'accesso intelligente ai dati multimediali, come vedremo nella prossima sezione.

#### 4.3 Navigazione utente in MAD

La piattaforma di pubblicazione in MAD fornisce funzionalità di *retrieval* e *browsing*. Essa gestisce le istanze dei documenti che si conformano al formato MAD dei metadati e li rende disponibili in una rappresentazione basata su Web. Inoltre fornisce l'accesso ai materiali multimediali in modo che possano essere esportati dalla piattaforma centrale (Core Platform). L'architettura del sistema di pubblicazione è basata su una applicazione Web per l'interfaccia utente, su un DBMS per la memorizzazione delle informazioni disponibili riguardo ai programmi, e sul motore di indicizzazione e ricerca KIM. La Piattaforma di Pubblicazione offre due principali funzionalità:

- *il data import*, che consente di sottomettere ed indicizzare i materiali provenienti dalla documentazione;
- *la ricerca e navigazione* richiesti nell'applicazione Web.

L'interfaccia per la ricerca supporta diversi approcci di retrieval, descritti nella sezione 4.1. L'utente può scegliere la propria unità informativa (un programma o una notizia) da recuperare. Le interrogazioni possono insistere su dati strutturati, quali il titolo, la data di trasmissione ed il servizio, i contributi (autori, giornalisti o registi), la classificazione tematica, una descrizione testuale.

Quando l'utente seleziona un'unità informativa, una nuova finestra di navigazione viene aperta, presentandone i dettagli. La finestra è composta di quattro frame: una video *preview*, l'albero delle componenti editoriali, la finestra dei *frame* chiave per la visualizzazione rapida ed un *frame multi-tab* estensibile; in ciascuno di essi viene rappresentato il risultato di una specifica elaborazione. Durante l'interazione utente il contenuto di tutti i frame viene sincronizzato.



**Figura 5 - Piattaforma di Pubblicazione: la visualizzazione delle trascrizioni audio**

Il frame multi-tab contiene le seguenti implementazioni:

- *Info*. Questo tab mostra i metadati generali riguardanti un programma (per esempio il titolo, sottotitolo ...).
- *Transcriptions* (Figura 5) Questo tab mostra i risultati del GAMP responsabile della trascrizione della voce in testo. Il testo è suddiviso in segmenti che rappresentano singole notizie. L'interfaccia consente di selezionare frammenti specifici di testo.
- *Semantic Analysis*. Questo tab mostra un albero navigabile esplorato interattivamente (Figura 6). Esso rende accessibili le entità riconosciute dal GAMP semantico ed i metadati riguardo i *link* esterni che rappresentano associazioni a sorgenti informative esterne (per esempio giornali su Web).

- *Content analysis.* Questo tab consente la visualizzazione di informazioni di sintesi visuale della notizia (*stripe images*) o informazioni relative ai movimenti della camera su una barra temporale (*timeline*).

Il video *preview* utilizza la tecnologia Windows Media Player e permette la sincronizzazione di tutti i tab disponibili.

L'interfaccia di pubblicazione consente diversi approcci alla navigazione. È possibile navigare una notizia usando l'albero della classificazione posto al di sotto del visualizzatore, o anche selezionando i *key frames* in fondo alla pagina. Selezionando i tab corrispondenti è infine possibile vedere i diversi metadati relativi ad una notizia o di un programma, cioè vedere le trascrizioni audio allineate con la *timeline* o osservare metadati di tipo semantico legati al contenuto. Nell'immagine un TG regionale.



**Figure 6 - Publication Platform: Risultati della Analisi Semantica**

Nella Figura 7 viene riportato il risultato che la piattaforma di pubblicazione produce in risposta alla query in italiano “*La dichiarazione di Chirac riguardo alla assegnazione delle Olimpiadi a Londra*”. Nella prima parte viene riportata la notizia come scena dal TG2 italiano del 6 luglio 2005: questa costituisce la seconda proposta del sistema. Nella seconda parte della figura viene riportata la prima proposta del sistema che corrisponde infatti ad una notizia della BBC, nella stessa data. Le due risposte sono corrette e corrispondono a flussi di notizie video in lingue diverse: è evidente che l'efficacia del retrieval non dipende dalla lingua della interrogazione, grazie al processo di CLIR applicato.



Figura 7 - Piattaforma di pubblicazione: i risultati del CLIR per una ricerca in italiano

## 5 Conclusioni

Le applicazioni dell'Intelligenza Artificiale affrontano un nodo cruciale nella gestione ed nell'accesso ai materiali multimediali e debbono fronteggiare il "semantic gap" tra i contenuti audiovisivi di base e le concettualizzazioni proprie delle comunità utenti.

In questo lavoro è stato sintetizzato l'approccio all'indicizzazione dei materiali multimediali negli archivi radiotelevisivi di vasta scala realizzato nel progetto PrestoSpace. I contributi di tale approccio spaziano dall'arricchimento e la generalizzazione del materiale audiovisivo di base fornito in input attraverso la fase di documentazione, sino all'abilitazione di forme avanzate di *information retrieval*, quali la navigazione attraverso modelli ontologici, e l'elaborazione di *query* in scenari d'uso multilinguali.

Le tecnologie presentate sono state individualmente oggetto di valutazioni quantitative che attestano la robustezza e la qualità dei singoli componenti impiegati (ad es. [8,10,12,18]). Un'analisi quantitativa complessiva non è ad oggi disponibile, sebbene l'iniziale analisi qualitativa conferma un'accuratezza generale molto buona. L'analizzatore semantico dei materiali audiovisivi in lingua italiana è oggi in uso sperimentale per la documentazione assistita dei notiziari quotidiani.

Il *framework* tecnologico offerto da PrestoSpace apre la strada ad una molteplicità di applicazioni, che includono il trattamento di dati audiovisivi diversi dalle notizie TV, e l'automazione dell'estrazione di informazioni complesse (quali gli eventi, ossia le relazioni tra gli individui e i concetti rilevati nell'input AV). La crescente disponibilità di archivi digitalizzati e le possibilità offerte al loro arricchimento abiliteranno ulteriori estensioni legate, ad esempio, allo sfruttamento di proprietà cross-mediali. Inoltre, le applicazioni allo studio ed alla specializzazione di materiali AV nella promozione di reti sociali o nella creazione di strumenti proattivi di condivisione per le comunità virtuali sono evidentemente altri concreti sviluppi di indagine aperti dalla ricerca qui discussa.

## RINGRAZIAMENTI

Questo lavoro sarebbe stato impossibile senza il contributo complessivo del gruppo di lavoro MAD di Prestospace che ha permesso la realizzazione di un così vasto insieme di servizi e metodologie. Un ringraziamento particolare va certamente a Marco Cammisa per il suo entusiasmo ed il suo insostituibile contributo al progetto ed allo sviluppo dell'analizzatore semantico dei programmi in lingua italiana.

## 6 Bibliografia

- [1] R. DEL PERO, G. DIMINO, AND M. STROPPIANA, "Multimedia Catalogue – the RAI experience", *EBU Technical Review nr. 280*, European Broadcasting Union, Geneva, Summer 1999, pp. 1-13.
- [2] A. MESSINA, AND D. AIROLA GNOTA, "Automatic Archive Documentation based on Content Analysis", *IBC 2005 Conference Publication, International Broadcasting Convention*, Amsterdam, September 2005, pp. 278-286.
- [3] A. MESSINA, "Documenting the Archive using Content Analysis Techniques", *EBU Technical Review nr. 305*, European Broadcasting Union, Geneva, January 2006.
- [4] ISO/IEC 15938, *Multimedia Content Description Interface*.
- [5] EBU Tech3295, European Broadcasting Union (EBU) P\_META Metadata Exchange Scheme.

- [6] BRUGNARA, F., CETTOLO, M., FEDERICO, M., AND GIULIANI, D. (2000), A system for the segmentation and transcription of Italian radio news. In *Proceedings of RIAO, Content-Based Multimedia Information Access*, Paris, France.
- [7] W. BAILER, F. HÖLLER, A. MESSINA, D. AIROLA, P. SCHALLAUER, M. HAUSENBLAS, State of the Art of Content Analysis Tools for Video, Audio and Speech, *Deliverable 15.3 of the IST PrestoSpace project*, March 2005.
- [8] W. BAILER, P. SCHALLAUER G. THALLINGER, “Joanneum Research at TRECVID 2005 – Camera Motion Detection”, *Proc. of TRECVID Workshop*, Gaithersburg, MD, USA, Nov. 2005.
- [9] W. BAILER, H. MAYER, H. NEUSCHMIED, W. HAAS, M. LUX, W. KLIEBER, “Content-based video retrieval and summarization using MPEG-7”, *Proc. Internet Imaging V*, San Jose, CA, USA, Jan. 2004, pp. 1-12.
- [10] BASILI R., F.M. ZANZOTTO, Parsing Engineering and Empirical Robustness, 8 (2/3) 97-120, *Journal of Language Engineering*, Cambridge University Press, 2002
- [11] ROBERTO BASILI, MARCO CAMMISA, EMANUALE DONATI, RitroveRAI: A Web Application for Semantic Indexing and Hyperlinking of Multimedia News, in *International Semantic Web Conference*, Y. Gil, E. Motta, V.R. Benjamins, M.A. Musen Eds., Springer-Verlag, Lecture Notes in Computer Science, LN 3279, 97-111, 2005.
- [12] R. BASILI M. CAMMISA, A. GLIOZZO, Integrating Domain and Paradigmatic Similarity for Unsupervised Sense Tagging, *Proceedings of the European Conference on Artificial Intelligence*, Riva del Garda, (Italy), 2006.
- [13] A. KIRYAKOV, B. POPOV, D. OGNJANOFF, D. MANOV, A. KIRILOV, M. GORANOV, Semantic Annotation, Indexing, and Retrieval. *Elsevier's Journal of Web Semantics*, Vol. 2, Issue (1), 2005.
- [14] B. POPOV, A. KIRYAKOV, D. OGNJANOFF, D. MANOV, A. KIRILOV, KIM - a semantic platform for information extraction and retrieval, *Journal of Natural Language Engineering*, Vol. 10, Issue 3-4, Sep 2004, pp. 375-392, Cambridge University Press.
- [15] BERRY, M.W., DUMAIS, S.T., O'BRIEN, G.W. Using linear algebra for intelligent information retrieval, *SIAM Review*, Vol. 37, No. 4, pp. 573-595, December 1995.
- [16] R. BASILI, M. CAMMISA, F.M. ZANZOTTO, A semantic similarity measure for unsupervised semantic disambiguation, *Proceedings of the Language, Resources and Evaluation LREC 2004 Conference*, Lisbon, Portugal, 2004.
- [17] M. DOWMAN, V. TABLAN, H. CUNNINGHAM AND B. POPOV. Web-Assisted Annotation, Semantic Indexing and Search of Television and Radio News. *14th International World Wide Web Conference*. Chiba, Japan, 2005.
- [18] ROBERTO BASILI, ALESSANDRO MOSCHITTI AND MARIA TERESA PAZIENZA, NLP-driven IR: Evaluating performance over a text classification task, in *Proceedings of the 17th International Joint Conference of Artificial Intelligence (IJCAI 2001)*, August, Seattle, Washington, USA, 2001.
- [19] ROBERTO BASILI AND ALESSANDRO MOSCHITTI, Automatic Text Categorization: from Information Retrieval to Support Vector Learning. *Aracne editrice*, Rome, Italy, 2005.



# VISUAL INFORMATION RETRIEVAL

## INDICIZZAZIONE E REPERIMENTO DI DATI PITTORICI

*Gianluigi Ciocca, Raimondo Schettini*

Imaging and Vision Laboratory  
Dipartimento di Informatica Sistemistica e Comunicazione (DISCo)  
Università degli studi di Milano-Bicocca  
Viale Sarca, 336 - 20126 Milano, Italy

ciocca@disco.unimib.it, schettini@disco.unimib.it  
<http://www.ivl.disco.unimib.it>

### 1 Introduzione

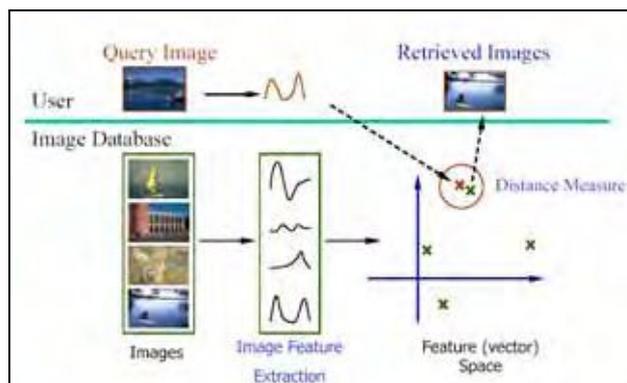
La ricerca per contenuto su database di immagini e video ha avuto negli ultimissimi anni un notevole impulso. Vi sono diversi prodotti commerciali e sistemi prototipali attualmente disponibili (ad esempio [1][2][3][4]), questi però sono spesso inadeguati in contesti applicativi quali i beni artistici e culturali. In generale, il funzionamento di questi sistemi si basa su tecniche di indicizzazione che estraggono automaticamente dai dati visuali caratteristiche numeriche o simboliche che ne codificano il contenuto. Durante un'interrogazione al database le descrizioni sono confrontate fra loro secondo una metrica predefinita per reperire i dati pittorici richiesti.

L'efficacia di questi sistemi è dipendente da diversi fattori:

1. La flessibilità data all'utente di eseguire ricerche con diversi obiettivi, quali:
  - a) *target search*: l'utente vuole reperire una particolare dato pittorico che conosce a priori e di cui non ha necessariamente un esemplare approssimato con il quale iniziare la ricerca;
  - b) *similarity search*: l'utente vuole reperire dei dati visuali, simili per una o più caratteristiche pittoriche (colore, forme ecc...), ad una o più immagini di esempio;
  - c) *category search*: l'utente vuole reperire tutti i dati pittorici che appartengono ad una specifica categoria (ritratti di persone, ad esempio).
2. L'insieme delle caratteristiche (pittoriche e/o testuali) usate per indicizzare i dati visuali;
3. La metrica adottata per valutare la similarità o la corrispondenza tra le descrizioni pittoriche o testuali dei dati visuali;
4. La flessibilità con cui il sistema può adattarsi alle specifiche esigenze dell'utente e quindi offrire strumenti atti a migliorare i risultati delle ricerche o la loro fruizione.

L'idea alla base dei sistemi di *content-based retrieval* basati su descrittori pittorici multipli e su metriche di similarità è schematizzata in Figura 1.

Da ogni immagine sono estratti automaticamente uno o più descrittori pittorici (*feature*) che sono memorizzati nel database. L'insieme di questi descrittori (di solito vettori numerici), forma lo spazio dei descrittori (*feature space*). Data una interrogazione (*query*), il descrittore corrispondente è estratto e confrontato con tutti quelli presenti nel database, il sistema ritorna le immagini a distanza minore dalla query.



**Figura 1 - Principio di funzionamento di un sistema di *content-based retrieval* su immagini**

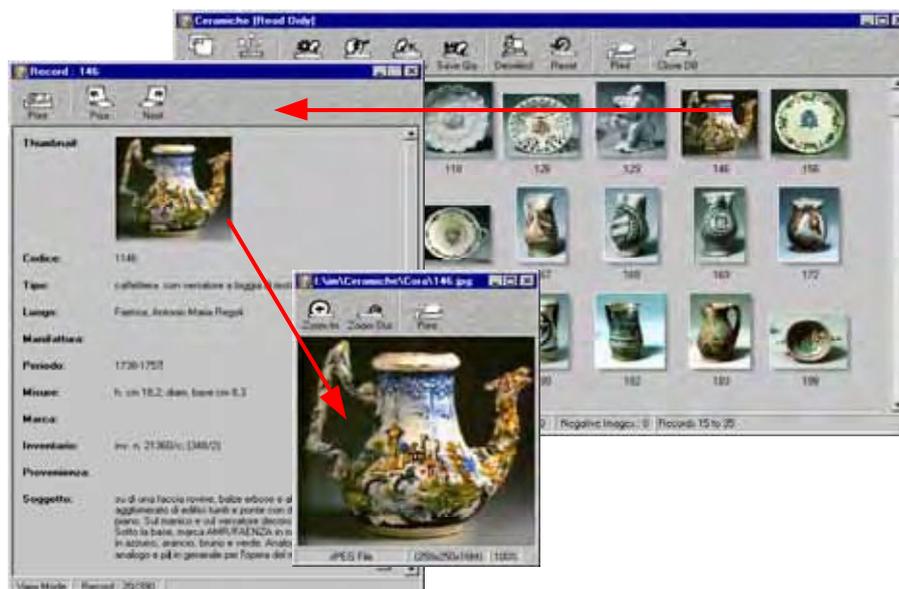
In questo paradigma il controllo del processo di ricerca può diventare molto difficoltoso se le caratteristiche usate nell'indicizzazione sono pre-definite e fissate per specifiche classi di immagini o se, in un sistema di tipo *general-purpose*, l'utente è forzato a scegliere interattivamente quali caratteristiche visuali o testuali usare nel reperimento dei dati pittorici. In sistemi di ricerca come Virage [5], QBIC [6], JACOB [7], VisualSeek [8] ed altri, all'utente è richiesto di costruire la query e di modellare l'obiettivo della ricerca esplicitandola in termini di caratteristiche visuali elementari (colore, forma, tessitura) che sono difficilmente descrivibili singolarmente ed indipendentemente. Per l'utente è più semplice indicare uno o più esempi di dati visuali che sono simili o corrispondenti fra loro, piuttosto che definire esplicitamente la natura di tale similarità. Un sistema di ricerca evoluto dove quindi essere flessibile nella costruzione della query e deve evincere in modo automatico l'insieme di descrittori (ed eventualmente di misure di similarità) più adatto per un determinato obiettivo di ricerca [1].

Per questo motivo, è stato studiato per le immagini un paradigma di ricerca già utilizzato nei sistemi di ricerca di documenti testuali basato su tecniche di *relevance feedback*. In un sistema di *information retrieval*, il *relevance feedback* è il meccanismo di raffinamento automatico di una query esistente, effettuato usando le informazioni fornite dall'utente sulla rilevanza dei documenti precedentemente reperiti, in modo tale che la nuova query è un'approssimazione migliore di ciò che intende reperire.

L'utente è quindi sollevato dal compito di costruire esplicitamente la query nei termini usati per l'indicizzazione e può limitarsi a fornire esempi di ciò che considera rilevante e/o non rilevante. Diverse tecniche di *relevance feedback* sono state proposte ed applicate ai dati visuali [9][10][11][12].

In questo contributo, mostreremo le caratteristiche principali del sistema di *content-based retrieval* QuickLook che, sviluppato inizialmente per archivi di immagini, è in fase di estensione per gestire archivi video.

La Figura 2 mostra le informazioni che sono gestite da QuickLook e che possono essere usate per interrogare il database. Il contenuto del database è visualizzato in forma tabellare tramite miniature (*thumbnail*). A ciascuna immagine può essere associata una scheda testuale contenente parole chiave e una o più annotazioni in formato libero.



**Figura 2 - Informazioni gestite dal sistema QuickLook**

Dati pittorici e testuali possono essere utilizzati in combinazione tra loro per eseguire:

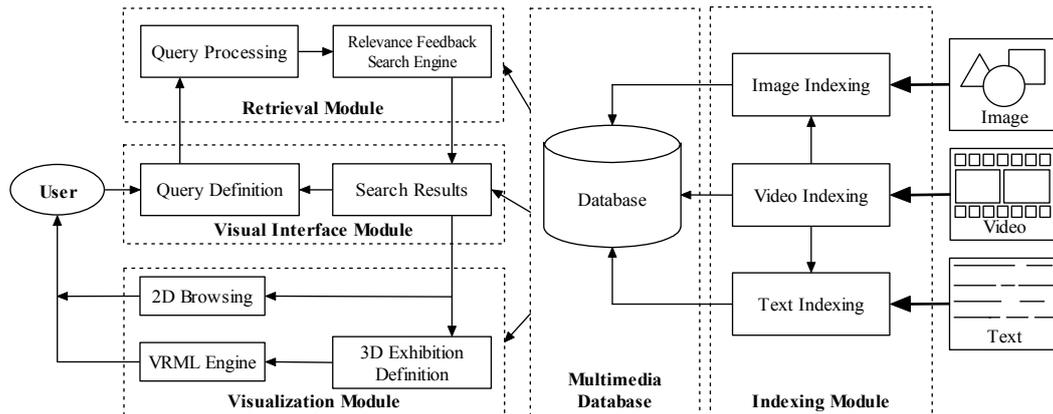
- interrogazioni tradizionali per mezzo di parole chiave associate alle immagini;
- interrogazioni per similarità pittorica fra le immagini;
- interrogazioni per similarità testuale fra le eventuali annotazioni in formato libero;
- una qualsiasi combinazione delle interrogazioni sopra riportate.

Dopo la prima risposta del sistema, l'utente può raffinare progressivamente l'interrogazione fornendo ulteriori esempi di dati rilevanti e non rilevanti. Il meccanismo di *relevance feedback* esegue un'analisi statistica sui descrittori (pittorici e/o testuali) delle immagini e permette di identificare quali descrittori sono da considerarsi rilevanti (e quanto) nel modellare l'idea di similarità che era implicita nella interrogazione dell'utente.

## 2 Panoramica del sistema QuickLook

Il sistema è composto di cinque moduli (Figura 3):

- *Visual Interface Module*, è il modulo principale che permette all'utente di interagire con il sistema;
- *Retrieval Module*, si occupa del processamento delle query e di reperire i dati che le soddisfano;
- *Multimedia Database* raccoglie e gestisce i dati pittorici e le informazioni ad essi associate;
- *Indexing Module* si occupa dell'estrazione automatica, dai dati pittorici, delle informazioni necessarie per poter eseguire le interrogazioni nelle varie modalità;
- *Visualization Module* permette di mostrare i risultati di una ricerca collocati in un ambiente virtuale.



**Figura 3 - L'architettura del sistema QuickLook**

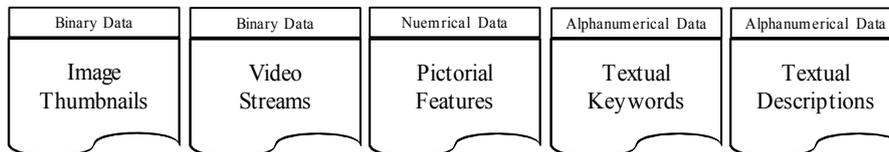
*Visual Interface Module:* Questo modulo consente all'utente di vedere il contenuto del database per mezzo di una serie di immagini in miniatura. In ogni momento l'utente può visualizzare le immagini corrispondenti in alta risoluzione. Inoltre è possibile accedere alle componenti testuali associate ai dati attraverso delle schede che contengono informazioni strutturate e informazioni non strutturate (descrizioni in testo libero). L'utente può eseguire una tradizionale interrogazione mediante i campi testuali (ricerca per keyword e similarità testuale), eseguire una ricerca per similarità selezionando una o più immagini che approssimano la query che l'utente vuole eseguire, o entrambe.

*Retrieval Module:* Questo modulo si occupa della definizione e del processamento della query (in particolare quando la query è effettuata secondo la modalità di ricerca per similarità). Una query può essere composta da dati visuali e/o dati testuali (*sub-query*). Ogni sub-query è processata separatamente e i risultati sono poi combinati tra loro per mezzo di una funzione di similarità che associa ad ogni elemento uno score di rilevanza. I risultati così valutati sono poi ordinati secondo lo score in ordine decrescente. Se l'utente non è soddisfatto del risultato ottenuto, può raffinare la ricerca aggiungendo vincoli testuali e/o selezionando esempi rilevanti e non rilevanti di ciò che sta cercando. La funzione di similarità complessiva viene modificata adattivamente secondo le informazioni fornite dall'utente variando il contributo di ciascuna informazione pittorica e testuale attraverso l'algoritmo di *relevance feedback*.

*Multimedia Database:* Il Database contiene tre tipi di dati (Figura 4):

- *Binary data:* le miniature (thumbnail) che rappresentano le immagini o fotogrammi estratti dalle sequenze video (*key frame*), le immagini originali e gli stream video.
- *Numerical data:* informazioni relative alle caratteristiche pittoriche delle immagini come colore, tessitura (*texture*), forma, e distribuzione delle regioni. Queste informazioni (feature o descrittori) sono estratte automaticamente senza l'intervento di un operatore e permettono di descrivere il contenuto visuale delle immagini. La loro rappresentazione è generalmente ma non necessariamente costituita da vettori numerici.
- *Alphanumerical data:* informazioni testuali associate alle immagini. Possono essere suddivise in: i) dati indipendenti dal contenuto delle immagini (*content-independent data*). Tali dati non sono direttamente legati al contenuto delle immagini ma sono in un qualche modo relazionati ad esse come il nome dell'autore, la data di creazione, luogo dove si trova l'oggetto ecc... Questi dati sono anche definiti come *keyword*. ii) Dati che

descrivono il contenuto delle immagini (*content-descriptive data*) cioè dati che, utilizzando un testo libero scritto in linguaggio naturale, permettono di conoscere il contenuto delle immagini ad un più alto livello semantico rispetto ai descrittori pittorici numerici; iii) Annotazioni automaticamente estratte dalle immagini come ad esempio didascalie, titoli o sottotitoli che compaiono nelle sequenze video.



**Figura 4 - I dati contenuti nel database**

*The Indexing Module:* data la loro diversa natura, testo, immagini e video sono processati separatamente e le relative informazioni sono memorizzate nel Database. La descrizione del modulo di indicizzazione è data nella sezione seguente.

*Visualization Module:* la maggior parte dei sistemi di content-based retrieval, ed in particolare quelli di dedicati alle immagini, utilizzando una rappresentazione tabellare 2D per mostrare i risultati di una interrogazione. Il sistema QuickLook, pur mantenendo questo tipo di rappresentazione per permettere un efficiente browsing del Database, è in grado di visualizzare i risultati di una ricerca in un ambiente 3D virtuale. L'ambiente scelto è quello di una esposizione dove i risultati (immagini e altro) sono collocati nello stesso ordine di rilevanza con cui sono stati reperiti dal sistema. L'utente è così in grado di muoversi all'interno delle varie stanze che compongono l'esposizione e, se lo desidera, recuperare le informazioni associate a ciascun elemento inserito. L'ambiente è stato realizzato utilizzando la tecnologia VRML 2.0 per garantirne la fruibilità sia off-line che via Web.

### 3 Indicizzazione dei dati

#### 3.1 Content-independent data: Keyword

I dati testuali semplici (keyword) non riguardano direttamente il contenuto visuale delle immagini ma rappresentano informazioni ancillari non derivabili direttamente da esse. Le *keyword* sono trattate con le tradizionali tecniche dei sistemi DBMS. Le ricerche sono effettuate attraverso interrogazioni in linguaggio SQL che l'utente compone con l'ausilio di maschere e i risultati reperiti dal sistema sono tutti quelli che soddisfano i vincoli imposti sulle keyword (combinare tra loro per mezzo di operatori di aggregazione) scelte dall'utente.

#### 3.2 Content-descriptive data: Annotazioni Testuali

I dati possono essere accompagnati da annotazioni testuali che ne descrivono il contenuto visuale e semantico. Ad esempio, i fotogrammi video possono essere corredati dalla trascrizione audio del dialogo che avviene in una certa sequenza; le immagini possono essere accompagnate da una descrizione del contenuto fatta da qualche operatore o da esperti archivisti. Queste annotazioni possono essere usate per eseguire delle interrogazioni avanzate. Nell'indicizzazione i termini più significativi sono estratti da ogni annotazione e impiegati per costruire un dizionario linguistico di riferimento.

In QuickLook, il dizionario (Italiano) è automaticamente costruito ed è composto da tutti i termini che compongono le annotazioni testuali tranne quelli che fanno parte di una speciale stop-list (congiunzioni, articoli ecc...). Nessuna procedura di *stemming* (identificazione delle radici linguistiche) è utilizzata in quanto non esiste un algoritmo efficace per la lingua Italiana. Alcune variazioni dei termini (singolare/plurale, maschile/femminile,...) sono automaticamente rimosse. Ad ogni termine del documento testuale in esame è assegnato un peso di rilevanza  $TW$  che riflette la sua importanza all'interno dell'intera collezione dei documenti testuali. Ogni annotazione testuale è quindi indicizzata attraverso l'insieme dei suoi termini rilevanti e dai corrispondenti pesi  $TW$ .

Il dizionario è utilizzato per confrontare due annotazioni testuali [13]. Detto  $T_i$  l'insieme dei termini del documento  $i$ -esimo, la similarità testuale  $TS$  tra due insiemi  $T_i$  e  $T_j$  è definita come segue:

$$TS(T_i, T_j) = \frac{\sum_{k \in (T_i \cap T_j)} (TW_{ik} TW_{jk})}{\sqrt{\sum_{k \in T_i} (TW_{ik})^2 \sum_{k \in T_j} (TW_{jk})^2}} \quad (1)$$

dove  $TW_{ik}$  è la rilevanza del termine  $k$  del documento  $i$ -esimo.  $TS$  può assumere qualunque valore nell'intervallo  $[0,1]$ . Maggiore è il valore di  $TS$  e maggiore è la similarità testuale tra due annotazioni.

### 3.3 Content-dependent data: Descrittori Pittorici

I descrittori pittorici, come già introdotto in precedenza, sono estratti automaticamente e riguardano strettamente il contenuto visuale dei dati. Descrittori pittorici tipici sono: il colore, la tessitura, le relazioni spaziali tra regioni omogenee... Si tratta quindi di informazioni di basso livello (non semantiche) difficilmente descrivibili attraverso il linguaggio naturale.

È necessario quindi definire algoritmi specifici che siano in grado di determinare, automaticamente, l'entità di questi attributi e di fornirne una rappresentazione compatta. La scelta dei descrittori pittorici è fondamentale per la realizzazione di un sistema di content based retrieval efficace ed efficiente [14]. Diversi contesti applicativi spesso necessitano di insiemi particolari di descrittori (ad esempio [15][16]).

Poiché QuickLook è stato progettato come sistema *general purpose*, è stata definita un'ampia libreria di descrittori pittorici di basso livello. Questa libreria può però essere estesa in qualunque momento introducendo descrittori specifici. I descrittori utilizzati di *default* dal sistema sono i seguenti: *Color Histogram* [17], *Color Coherence Vector (CCV)* [18], *HSV Moments* [19], *Color Transitions* [20], *Edge Direction Histogram* [20], *Wavelet Statistics* [20], *Color Region Composition* [22], *Spatial Chromatic Histogram (SCH)* [23]. Questi descrittori sono calcolati sia sull'immagine nella sua interezza che su sotto-immagini ottenute suddividendo l'originale in diverse regioni secondo diverse strategie (Figura 5).

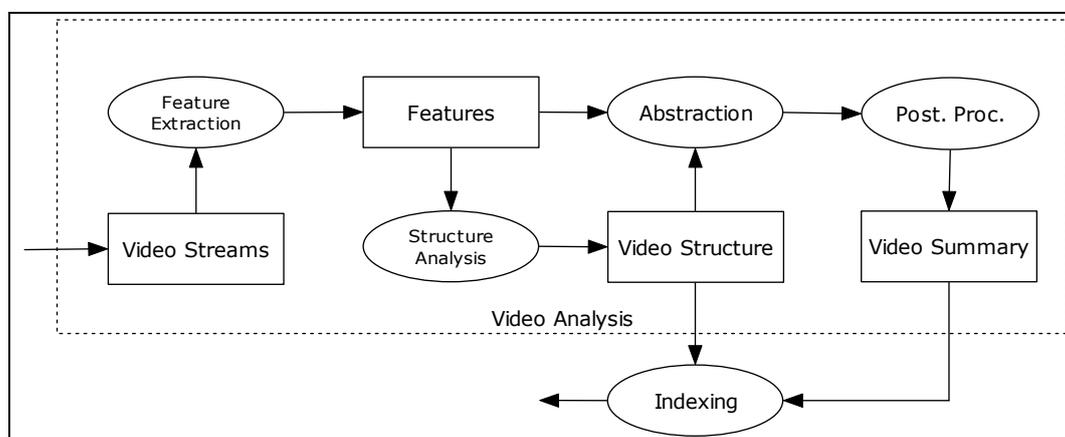
L'insieme dei valori dei descrittori forma l'indice pittorico dell'immagine. Ad ogni descrittore è associata una specifica misura di similarità. Le singole misure sono poi combinate fra loro mediante somma pesata. I pesi sono inizialmente uguali e sono poi automaticamente aggiornati dall'algoritmo di relevance feedback descritto nella Sezione 4.1. Per risolvere il problema delle distanze definite su intervalli di valori differenti è utilizzato un procedimento di normalizzazione Gaussiano [22][24].



**Figura 5 - Esempi di possibili suddivisioni dell'immagine in sotto-regioni per l'estrazione di descrittori pittorici. QuickLook utilizza la suddivisione a sinistra: oltre all'immagine intera, ci sono quattro regioni distinte e una regione centrale sovrapposta**

### 3.4 Analisi e indicizzazione video

L'indicizzazione dei video può richiedere l'analisi sia della componente visuale sia di quella audio. Algoritmi di *speech recognition* [25] permettono di creare trascrizioni testuali dei dialoghi all'interno delle sequenze video [26]. La fedeltà e completezza delle trascrizioni dipendono dalla tipologia dei video analizzati: i risultati migliori si ottengono nel caso di notiziari o documentari dove la componente parlata è predominante rispetto ad altre componenti audio (ad esempio, musica o rumori di sottofondo) [27]. Per quanto riguarda i contenuti visuali, è possibile indicizzare i video analizzando i singoli fotogrammi con le stesse tecniche utilizzate per le immagini. Dato che i fotogrammi, all'interno di una sequenza video, sono ridondanti dal punto di vista pittorico, non è necessario analizzarli tutti ma solo quelli più rappresentativi. L'insieme di questi fotogrammi (chiamati *key frame*) costituiscono il sommario visuale (*story-board*) del video e ne rappresentano, in una forma estremamente compatta e rapidamente accessibile, il contenuto visuale. La creazione di un sommario visuale è un processo complesso che coinvolge diverse fasi di elaborazione [28]:

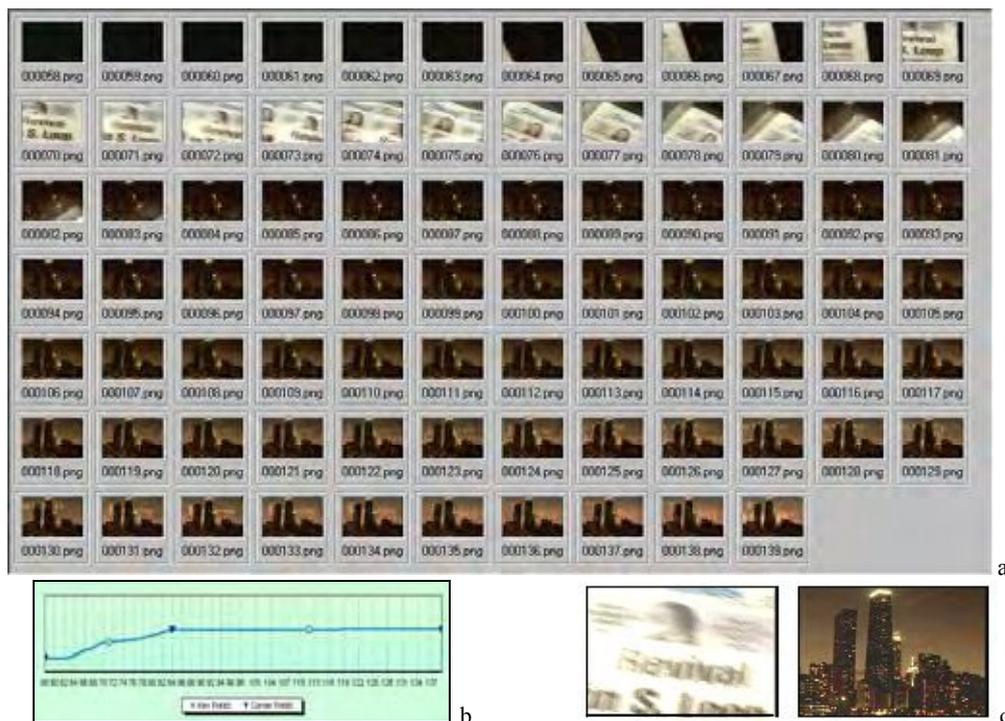


**Figura 6 - Schema dell'algoritmo di analisi video**

- *Definizione di una opportuna rappresentazione del contenuto visuale.* L'estrazione di descrittori sintetici è un passo fondamentale per un'analisi efficiente del contenuto di un video. Attributi visuali e informazioni temporali possono essere determinati in modo automatico e permettono di descrivere il contenuto dei fotogrammi. Le tecniche di estrazione dei descrittori si possono suddividere in due gruppi fondamentali: tecniche basate su dati compressi (generalmente con codifica MPEG) e tecniche basate su dati non compressi. Nel dominio compresso i descrittori sono estratti direttamente dal flusso

codificato dei dati e comprendono, ad esempio, statistiche sui coefficienti DCT (*Discrete Cosine Transform*) e sul grado di compressione dei singoli fotogrammi [29]. Questi metodi dipendono strettamente dall'uso della codifica MPEG utilizzata. Nel dominio non compresso, le tipologie di descrittori che possono essere utilizzate sono in numero maggiore ed è possibile ricorrere alle stesse metodologie impiegate per descrivere il contenuto pittorico delle immagini in QuickLook.

- *Segmentazione del video in unità elementari.* Le unità elementari di informazione all'interno di un video (dopo i singoli fotogrammi), sono gli *shot*. Uno *shot* è una sequenza video omogenea, ininterrotta, che è stata ripresa da una camera in un certo intervallo di tempo. Identificare all'interno del video i singoli *shot* permette di conoscere la struttura narrativa. Nella costruzione di un video, gli *shot* sono collegati tra loro per mezzo di effetti di editing come *cut* (cambio repentino tra uno *shot* ed un altro), *fade* (transizioni graduali tra uno *shot* e un fotogramma di colore uniforme), dissolvenze (transizioni graduali tra due *shot*) e altri (*zoom* ed effetti più complessi). Nell'implementazione attuale, QuickLook è in grado di riconoscere *shot* delimitati da *cut* e *fade* dato che questi sono gli effetti di editing più comuni. In questa fase, non tutti gli *shot* riconosciuti sono significativi. Ad esempio, in uno *shot* delimitato da un *fade-out* ed un *fade-in*, i fotogrammi “sfumano a nero” e successivamente da “nero” verso un fotogramma normale. Questo *shot*, non contenendo informazioni utili, deve essere riconosciuto ed eliminato dalle successive fasi di analisi. Un algoritmo di riconoscimento di dissolvenze è attualmente in fase di sviluppo e una volta completato, verrà integrato nel sistema.
- *Creazione del sommario visuale.* Esistono diverse modalità di creazione di un sommario visuale: sommari statici, cioè formati da un insieme di fotogrammi estratti dalla sequenza video e sommari dinamici, cioè una breve sequenza video costruita a partire dall'originale [30]. Qualunque sia il metodo utilizzato, il sommario deve mantenere le informazioni rilevanti del video in modo compatto, eliminando quindi tutte le informazioni ridondanti. La creazione di sommari dinamici pone dei problemi di gestione e sincronizzazione dell'audio. In particolare, è necessario editare l'audio in modo tale da fornire un video senza interruzioni o tagli che possono rendere poco fruibile il sommario. Ciò è realizzabile efficacemente solo nel caso di particolari tipologie di video (notiziari, documentari ed eventi sportivi) che hanno una struttura particolarmente definita e codificata tale da permettere l'utilizzo di regole ad-hoc per la costruzione dei sommari. Per questi motivi, QuickLook estrae sommari statici che non dipendono dalla tipologia del video da elaborare. L'algoritmo implementato seleziona dinamicamente i fotogrammi che devono fare parte del sommario, analizzando la complessità visuale di ogni *shot* individuato nella fase precedente. Per determinare la complessità visuale, le differenze pittoriche tra coppie di fotogrammi consecutivi sono accumulate nel tempo, il grafico dei valori così ottenuti descrive come i vari fotogrammi cambiano lungo la sequenza video. Discontinuità nel grafico indicano fotogrammi che si differenziano visivamente in modo significativo. Ciò può essere dovuto a movimenti di oggetti o persone, movimenti della camera o al verificarsi di particolari eventi legati alla storia narrata nel video. Questi punti sono considerati come “punti di interesse” [31] e sono utilizzati per identificare i *key frame* che costituiranno il sommario finale. In Figura 7 è mostrato un esempio di estrazione di *key frame*: i punti di interesse sono quelli identificati dai triangoli e corrispondono a punti ad alta curvatura sul grafico; i punti identificati dai rombi corrispondono ai *key frame* selezionati.



**Figura 7 - Esempio di costruzione di un sommario visuale. a) shot. b) grafico cumulativo delle differenze. Sono riportate le posizioni corrispondenti ai punti di interesse (triangoli) e ai key frame (rombi).** d) I fotogrammi corrispondenti ai key frame

Se la sequenza non presenta un andamento dinamico, non vi saranno punti ad alta curvatura e quindi un solo key frame sarà selezionato al centro della sequenza. Un altro esempio di sommario, applicato questa volta ad un video complesso contenente sequenze di un notiziario intervallate da spot pubblicitari, è mostrato in Figura 8. In questo esempio, l’algoritmo ha estratto un numero maggiore di key frame dalle sequenze pubblicitarie rispetto alle sequenze del notiziario, data la natura maggiormente dinamica delle prime rispetto le seconde.

- *Post-Processing dei sommari.* L’insieme dei key frame ottenuto può essere ulteriormente raffinato ed elaborato [32]. Il sommario potrebbe contenere dei fotogrammi che non sono significativi perché privi di sufficienti informazioni (molto scuri o sfocati), fotogrammi duplicati a causa del falso riconoscimento di shot o della bassa qualità delle riprese video. E’ necessario quindi procedere ad una fase di identificazione e rimozione di questa tipologia di fotogrammi come descritto in [32]. Nel caso in cui il sommario contenga un numero molto elevato di key frame, è possibile creare una gerarchia di sommari ciascuno dei quali è formato da un numero decrescente di key frame. Questa gerarchia struttura i key frame in modo tale da fornire sommari visuali a diversi livelli di dettaglio.
- *Definizione delle informazioni da indicizzare.* Ad ogni passo nel processo di analisi, possono essere raccolte diverse informazioni sul video come, ad esempio, la posizione temporale degli shot, la loro durata, l’insieme dei key frame, la struttura gerarchica del sommario. Altre informazioni si possono derivare a partire da queste: descrizioni pittoriche dei key frame; trascrizione dell’audio estratto da una porzione del video localizzata nell’intorno dei key frame [33]; estrazione dei testi sovra-impresi ai

fotogrammi [34] ecc... Nel caso specifico di QuickLook, i key frame sono memorizzati ed indicizzati come delle normali immagini a cui sono associate anche le informazioni relative alle sotto-sequenze video da cui sono stati estratti. Questo permette all'utente di accedere direttamente alla porzione video il cui contenuto è rappresentato dal key frame reperito.



**Figura 8 - Key frame estratti da un notiziario video con pubblicità. Il video dura 3 minuti e 39 secondi (6.561 fotogrammi e un frame rate di 30 frame al secondo). Il video è stato riassunto in 47 fotogrammi e 41 shot sono stati identificati**

#### 4 Interrogazione del database multimediale

QuickLook mette a disposizione dell'utente diverse modalità di interrogazione del database multimediale:

*Textual Search by Keywords:* l'utente può reperire immagini e/o video che soddisfano delle precise condizioni come: tutti i manufatti che sono stati realizzati da un particolare artista in un dato periodo; tutti i video riguardanti le Olimpiadi; ecc....

*Query by Sample:* l'utente fornisce un prototipo, esterno al database, di ciò che vuole reperire. Sono possibile tre diverse tipologie di query:

- *Pictorial Query by Sketch:* l'utente realizza uno sketch dell'immagine che vuole reperire. Lo sketch viene indicizzato e utilizzato dal sistema come query. Questa tipologia di interazione è utile quando non si ha a disposizione un esempio di query e si vuole iniziare la ricerca con informazioni approssimate.
- *Pictorial Target Search:* l'utente importa nel sistema una immagine che rappresenta la sua query. Questa tipologia di interrogazione è utile quando si deve verificare la presenza di una immagine nel database (cerca duplicati), oppure si vogliono reperire le informazioni relative ad una specifica immagine indicizzata dal sistema.
- *Query by Textual Sample:* è simile alle precedenti, ma riguarda i dati testuali ed è possibile quando sono disponibili annotazioni testuali sul contenuto delle immagini.

*Query by Example*: l'utente fornisce uno o più esempi, selezionati all'interno del database, di ciò che sta cercando. Anche in questo caso sono disponibili tre diverse tipologie di query:

- *Query by Pictorial Example*: l'utente seleziona una o più elementi del database focalizzando l'attenzione solo sulle proprietà pittoriche. Durante le fasi di ricerca, le misure di similarità utilizzate per valutare il contenuto pittorico degli elementi e la definizione della query rappresentativa degli esempi sono automaticamente gestite dall'algoritmo di relevance feedback.
- *Query by Textual Example*: l'utente seleziona uno o più elementi del database focalizzandosi solo sulle descrizioni testuali. Le descrizioni sono usate nelle fasi di ricerca per costruire la query analogamente a quanto è fatto per i dati pittorici.
- *Query by Multimedia Object Example*: l'utente seleziona uno o più elementi del database considerando sia le informazioni testuali che quelle pittoriche. Tutte le informazioni disponibili sono utilizzate durante la fase di ricerca e ad esse viene applicato l'algoritmo di relevance feedback.

#### 4.1 Relevance Feedback e raffinamento della query

La metrica finale per valutare la similarità tra due elementi visuali  $i$  e  $j$  è definita come:

$$\text{Dist}(\mathbf{X}^i, \mathbf{X}^j) = \frac{1}{p} \sum_{h=1}^p w_h d_h(\mathbf{X}_h^i, \mathbf{X}_h^j) + w_T d_T(T_i, T_j) \quad (2)$$

dove  $h$  è l'indice del descrittore pittorico (ricordiamo che possono essere utilizzati diversi descrittori pittorici per descrivere il contenuto visuale);  $d_h$  è la misura di similarità (normalizzata) per il descrittore  $h$ -esimo;  $T^i$  è la descrizione testuale (eventualmente) associata all'elemento  $i$ -esimo;  $d_T$  è la funzione di similarità tra due testi. Tutti i pesi associati ai vari termini  $w_T$  e  $w_h$  sono determinati automaticamente dall'algoritmo di relevance feedback. L'algoritmo di relevance feedback è descritto in dettaglio in [22] e [24]. L'idea di base è l'analisi statistica dei descrittori pittorici e testuali degli elementi rilevanti e non rilevanti selezionati dall'utente.

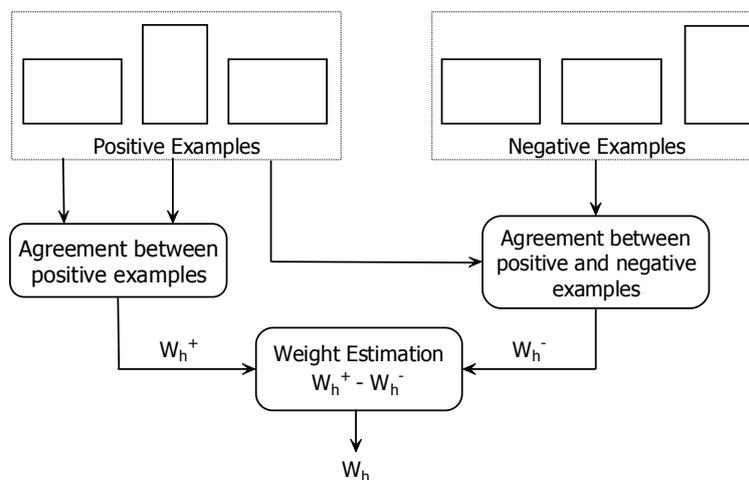


Figura 9 - Schema per il calcolo dei pesi di rilevanza dei descrittori

L'algoritmo evince quali caratteristiche sono comuni tra i vari esempi e che quindi rappresentano le informazioni importanti per l'utente. Il livello d'importanza è espresso per mezzo dei pesi associati ai vari descrittori. La Figura 9 mostra le fasi fondamentali per il calcolo dei pesi di rilevanza: descrittori comuni ai diversi esempi rilevanti (positivi) avranno associati pesi maggiori rispetto agli altri, a meno che questi descrittori non siano comuni anche agli esempi non rilevanti (negativi). In tal caso i pesi saranno abbassati di conseguenza. Durante le diverse sessioni di ricerca, gli utenti tendono a selezionare gli elementi rilevanti non perché sono globalmente simili tra loro ma perché questi hanno in comune certe caratteristiche con le immagini effettivamente cercate. Ad esempio, durante le fasi iniziali dell'interrogazione, un'immagine può essere selezionata come rilevante perché contiene gli stessi colori dell'immagine cercata, mentre un'altra immagine può essere selezionata perché ha la forma (ma non il colore) simile a quella voluta. Per questo motivo il nostro sistema adotta un nuovo algoritmo di raffinamento della query che sintetizza le caratteristiche parzialmente comuni alle immagini selezionate [24]. Un esempio del funzionamento dell'algoritmo è mostrato schematicamente in Figura 10 dove la query finale ha le caratteristiche evinte da diverse immagini considerate rilevanti. Un approccio simile è definito anche per le annotazioni testuali.

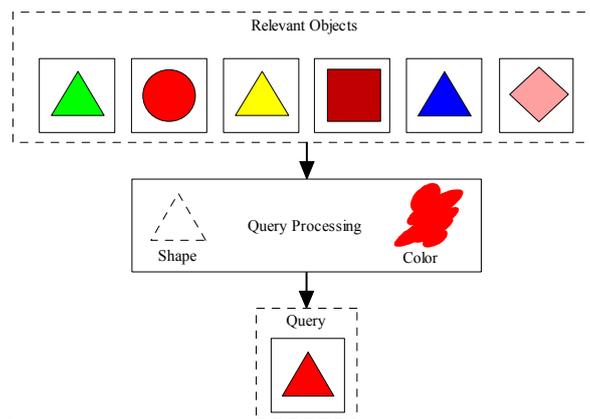


Figura 10 - Un esempio visivo del raffinamento della query

#### 4.2 Filtraggio dei dati

L'analisi di similarità tra una query  $Q$  e ogni immagine  $I$  nel database è un'operazione che può richiedere calcoli e tempi non trascurabili, è necessario quindi predisporre strutture dati opportune, oppure artifici computazionali, per accelerare le fasi di ricerca. Nel nostro sistema si è optato per l'utilizzo di un algoritmo di filtraggio dei dati basato sulle proprietà delle metriche di similarità. Questo algoritmo è in grado di eliminare dalla fase di ricerca tutte quelle immagini che, se si calcolasse la loro distanza pittorica con la query, risulterebbero avere similarità molto bassa. Basato su una variante del metodo proposto da Berman e Shapiro [35], l'algoritmo si può descrivere come segue: per ogni terna di immagini  $I$ ,  $Q$  e  $K$ , e una misura di distanza  $d$  (che sia una metrica), sono sempre valide le due seguenti disuguaglianze triangolari

$$d(I, Q) + d(Q, K) \geq d(I, K)$$

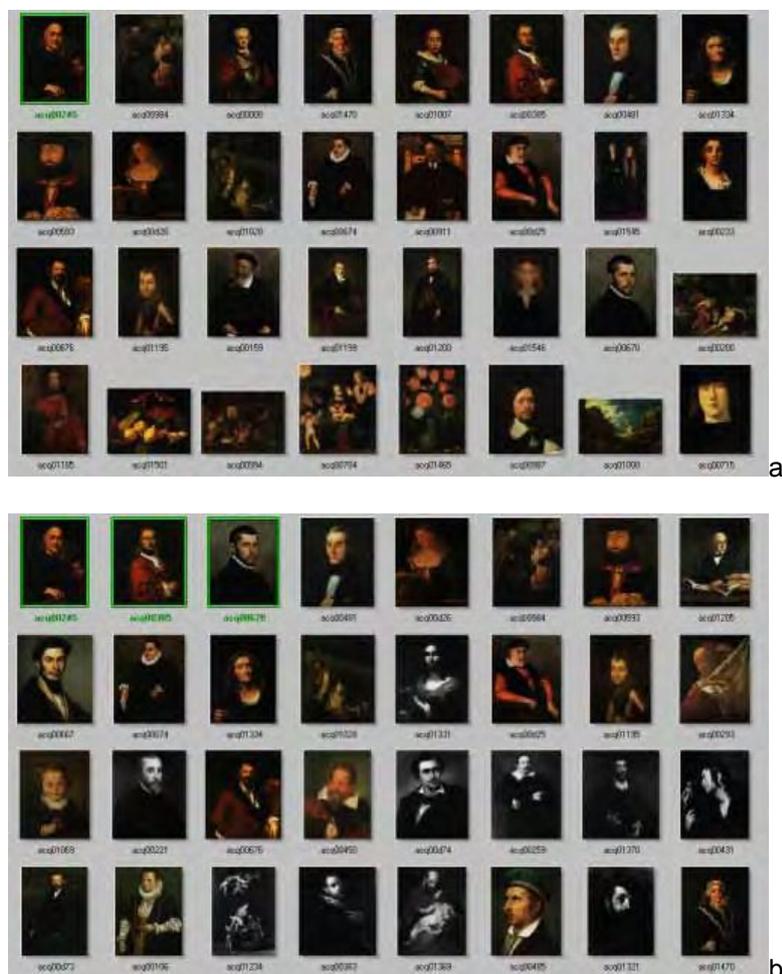
$$d(I, Q) + d(I, K) \geq d(Q, K)$$

Queste disuguaglianze possono essere combinate per ottenere il limite inferiore della distanza tra  $I$  e  $Q$ :

$$d(I, Q) \geq |d(I, K) - d(Q, K)|$$

Quindi, stabilito il limite inferiore della distanza  $d(I, Q)$ , è possibile eliminare tutte le immagini che soddisfano l'equazione precedente, dato che avranno una distanza superiore a quella di minima richiesta. L'immagine  $K$  è detta immagine chiave e le distanze tra questa immagine e le immagini nel database sono pre-calcolate e memorizzate nel database per poter essere immediatamente utilizzate durante le fasi di ricerca. Il metodo è stato esteso per supportare più immagini chiave e più distanze [24].

La Figura 11 mostra un esempio di ricerca eseguita sfruttando il meccanismo di relevance feedback. Inizialmente l'utente ha selezionato una sola immagine e il sistema ha reperito le immagini mostrate in a). Successivamente l'utente ha selezionato ulteriori immagini rilevanti e non rilevanti. Il risultato dopo l'utilizzo del relevance feedback è mostrato in b).



**Figura 11 - Esempio di ricerca di immagini mediante relevance feedback:**  
**a) mostra il risultato della ricerca dell'immagine in alto a sinistra in un database di circa 2000 dipinti**  
**b) mostra il risultato dopo la selezione di alcune immagini rilevanti e non rilevanti**

## 5 Visualizzazione dei risultati

Il modulo di visualizzazione consente ad un utente di creare un ambiente virtuale 3D dove mostrare i risultati delle sue interrogazioni (Virtual Museum) [36]. L'ambiente di riferimento è quello di un allestimento museale in quanto consente la disposizione di diversi elementi (immagini, testi, filmati ecc...) in un contesto coerente e familiare. È possibile rendere accessibili le riproduzioni delle opere ospitate nell'allestimento secondo criteri di organizzazione logica dei contenuti (mostre a tema, opere di uno stesso autore ecc...), oppure secondo l'ordine di rilevanza legato ad una precedente interrogazione.

Dal punto di vista tecnico il museo virtuale è realizzato mediante la tecnologia VRML (*Virtual Reality Modeling Language*), ciò ha permesso di costruire un ambiente virtuale che dà all'utente la sensazione di entrare in uno spazio 3D reale. Durante l'esplorazione, il fruitore del museo virtuale percepisce lo spazio, le proporzioni e le dimensioni dell'ambiente espositivo e degli oggetti in esso contenuti. Questo spazio risulta totalmente navigabile poiché strutturato in modo tale che un utente possa muoversi ed interagire con esso nella massima libertà. L'utente è coadiuvato da un'interfaccia grafica che permette di visualizzare il nome dell'opera selezionata o di eseguire una serie di azioni dipendenti dalla tipologia dell'oggetto (accesso alle schede informative, la lettura di un testo, la riproduzione di un filmato...).

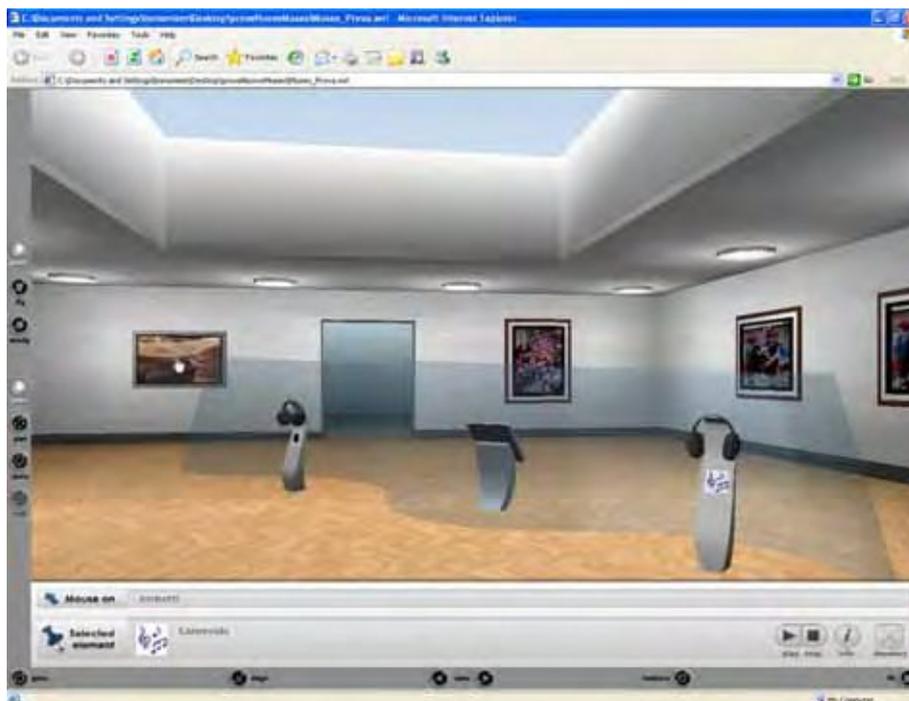
Dato che i contenuti che vengono visualizzati nelle sale del museo possono essere di tipo e di genere diversi (immagini, video, audio, testi e oggetti tridimensionali), essi possono essere rappresentati in modo diretto o indiretto.

Il primo caso riguarda le immagini e i video che l'utente può vedere senza intermediazioni.

Il secondo caso riguarda testi, audio ed oggetti 3D che, a causa della loro astrattezza o della forma insolita, risultano difficili da rappresentare direttamente. Questi dati vengono mostrati tramite un opportuno "avatar", ovvero una rappresentazione visuale semplificata dell'oggetto in questione. Ogni oggetto può possedere un link ad una scheda di presentazione contenente informazioni di supporto per la sua comprensione.

Le sale che costituiscono il museo sono in numero variabile: quando in una stanza non è più possibile inserire oggetti, ne è generata automaticamente una nuova.

Le varie stanze sono collegate tra loro mediante porte virtuali che, come un normale "hyperlink", trasportano l'utente nella nuova stanza (Figura 12).



**Figura 12. La finestra di navigazione del museo virtuale. Si possono vedere i quadri inseriti alle pareti così come gli avatar che rappresentano elementi audio e testuali e la porta che conduce alla stanza successiva. L'ambiente è liberamente navigabile senza restrizioni**

## 6 Conclusioni

In questo contributo abbiamo presentato le diverse problematiche inerenti all'indicizzazione ed al reperimento di dati pittorici. Lo sviluppo di sistemi di content-based retrieval, continua ad avere un notevole interesse applicativo. Attraverso la descrizione del sistema QuickLook, sono state descritte varie tecniche di indicizzazione per le immagini ed i video, sono stati quindi introdotti i più recenti paradigmi di interrogazione per i database multimediali e le problematiche ad essi associate. Dettagli sull'efficacia di QuickLook nel gestire database di immagini in diversi contesti applicativi si possono trovare nelle pubblicazioni [20][22][24]. L'estensione di QuickLook ai video è in una fase di sviluppo avanzato. Risultati preliminari sono riportati in [31][32]. E' stato infine presentato Virtual Museum, un nuovo modello di fruizione dei dati visuali che consiste in una ricostruzione 3D di un'esposizione liberamente navigabile dall'utente.

È possibile testare il funzionamento del sistema QuickLook per il reperimento di immagini:

- all'indirizzo <http://hyperion.sal.disco.unimib.it/qlook/>,
- all'interno dell'Archivio di Etnografia e Storia Sociale (AESS) della Regione Lombardia: <http://aess.itc.cnr.it>. Allo stesso indirizzo è anche possibile visitare l'esposizione virtuale "Luoghi e tradizioni del lavoro" realizzata sfruttando le potenzialità del museo virtuale qui presentato.

## 7 Ringraziamenti

Il sistema di content-based retrieval QuickLook, il modulo di analisi e indicizzazione video e il modulo di visualizzazione 3D (Virtual Museum), sono stati realizzati con il contributo di:

- Regione Lombardia, Direzione Generale Culture, Identità e Autonomie della Lombardia, nell'ambito dei progetti "Sistemi Interattivi per la fruizione di archivi multimediali" e IVIS-Intelligent Video Indexing and Summary- (INTERNUM – Programma INTERREG IIIB MEDOCC);
- Fondazione Cariplo e Regione Lombardia nell'ambito del progetto GIO-CAR Giochi d'archivio. Percorsi interattivi della ritualità e dello spettacolo tradizionale (Creazione di cataloghi informatizzati dei documenti audio-visuali);
- Università degli Studi di Milano-Bicocca, Comitato d'Area 6, Settore disciplinare INF 01, progetto: Algoritmi per l'interrogazione per contenuto di archivi fotografici;
- Progetto FIRB - MAIS (Multi-channel Adaptive Information Systems) [37];
- Consiglio Nazionale delle Ricerche (ITC-Sezione Milano).

## 8 Bibliografia

- [1] AIGRAIN O., ZHANG H., PETKOVIC D., Content-based representation and retrieval of visual media: A state-of-the-art review, *Multimedia Tools and Applications*, Vol. 3, pp. 179-182, 1996.
- [2] RUI Y., HUANG T.S., Image retrieval: current technologies, promising directions, and open issues", *Journal of Visual Communication and Image Representation*, Vol. 10, pp. 39-62, 1999.
- [3] GUDIVADA V.N, RAHAVAN V.V., Modeling and retrieving images by content, *Information Processing and Management*, 33, pp. 427-452, 1997.
- [4] DEL BIMBO A., Visual Information Retrieval, Morgan Kaufmann Publishers, San Francisco, California, 1999.
- [5] BACH J.R., FULLER C., GUPTA A., HAMPAPUR A., HOROWITZ B., HUMPHREY R., JAIN R., AND SHU C.F., The Virage image search engine: An open framework for image management, *Proc. SPIE Storage and Retrieval for Still Image and Video Databases IV*, Vol. 2670, pp. 76–87, 1996.
- [6] FLICKNER R., SAWHNEY H., NIBLACK W., et al., Query by image and video content: The QBIC system, *IEEE Computer*, pp. 23-31, 1995.
- [7] LA CASCIA M.; ARDIZZONE E., JACOB: just a content-based query system for video databases, *Proc. ICASSP-96 Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1216-1219, 1996.
- [8] JOHN R. SMITH, SHIH-FU CHANG, VisualSEEk: a Fully Automated Content-Based Image Query System, *Proc. ACM Multimedia*, pp. 87-98, 1997.
- [9] ZHOU X.S., HUANG T.S., Relevance feedback in image retrieval: A comprehensive review, *J. of Multimedia Sys.*, Vol. 8, No. 6, pp. 536-544, 2003.

- [10] COX I.J., MILLER M.L., MINKA T.P., PAPATHOMAS T.V., YIANILOS P.N., The Bayesian image retrieval system, PicHunter: theory, implementation and psychological experiments, *IEEE Trans. on Image Proc.*, Vol. 9, No. 1, pp. 20-37, 2000.
- [11] ORTEGA B.M., MEHROTRA S., Relevance feedback techniques in the MARS image retrieval system, *Multimedia Systems*, Vol. 9, No. 6, pp. 535-547, 2004
- [12] BRUNELLI R., MICH O., Image retrieval by examples, *IEEE Transactions on Multimedia*, Vol. 2(3), pp. 164–171, 2000.
- [13] MC GILL M.J., SALTON G., *Introduction to modern Information Retrieval*, McGraw-Hill, 1983.
- [14] SMEULDERS A.W.M., WORRING M., SANTINI S., GUPTA A., JAIN R., Content-based image retrieval at the end of the early years, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 2, No. 12, pp. 1349-1380, 2000.
- [15] TANGELDER J.W.H., VELTKAMP R.C., A survey of content based 3D shape retrieval methods, *Proc. of the Shape Modeling Apps.*, pp. 146-156, 2004
- [16] ZHANG D., PHAM B., YUEFENG LI, Modelling traditional chinese paintings for content based classification and retrieval, *Proc. Multimedia Modelling Conference*, pp. 258-264, 2004.
- [17] STRICKER M, SWAIN M, The capacity of color histogram indexing, *IEEE Int. Conf. On Computer Vision and Pattern recognition*, pp. 704-708, 1994.
- [18] PASS G., ZABIH R., MILLER J., Comparing Images Using Color Coherence Vectors, *Proc. Fourth ACM Multimedia 96 Conference*, pp. 65-73, 1996.
- [19] STRICKER M. AND M. ORENGO, Similarity of color images, *Proc. Storage and Retrieval for Image and Video Databases III*, pp. 381-392, 1995.
- [20] CIOCCA G., SCETTINI R., Content-based similarity retrieval of trademarks using relevance feedback, *Pattern Recognition*, Vol. 34, pp. 103-199, 2001.
- [21] SCHEUNDERS P., LIVENS S., VAN DE WOUWER G., VAUTROT P., VAN DYCK D., Wavelet-based texture analysis, *Journal on Computer Science and Information Management*, Vol.1, No. 2, pp. 22-34, 1998.
- [22] CIOCCA G., SCETTINI R., A relevance feedback mechanism for content-based image retrieval, *Inf. Proc. and Management*, Vol. 35, pp. 605-632, 1999.
- [23] CINQUE L., LEVIALDI S., AND PELLICANO' A., Color-based image retrieval using Spatial-Chromatic Histograms, *IEEE Multimedia Systems 99, IEEE Computer Society*, II, pp. 969-973, 1999.
- [24] CIOCCA G., GAGLIARDI I., SCETTINI R, Quicklook<sup>2</sup>: an integrated multimedia system”, *J. of Visual Lang. and Computing*, Vol. 12, pp. 81-103, 2001.
- [25] RUDNICKY I. A., HAUPTMANN A. G., Lee KF., Survey of current speech technology, *Communications of the ACM*, Vol. 37, No. 3, pp. 52-57, 1994.
- [26] LU GOUJUN, Indexing and retrieval of audio: a survey, *Multimedia Tools and Applications*, Vol. 15, No. 3, pp. 269-290, 2001.

- [27] GAUVAIN J.L., LAMEL L., ADDA G., Transcribing broadcast news for audio and video indexing, *Communications of the ACM*, Vol. 43, No. 2, pp. 43-70, 2000.
- [28] DIMITROVA N., ZHANG H.J. , SHAHRARAY B., IBRAHIM S.M., HUANG T., ZAKHOR A., Applications of video-content analysis and retrieval, *IEEE MultiMedia*, Vol. 9, No. 3, pp. 42-55, 2002.
- [29] ZHANG H.J., LOW C.Y., SMOLIAR S.W., Video parsing and browsing using compressed data, *Multimedia Tools Applications*, Vol. 1, pp. 91-113, 1995.
- [30] TRUONG B.T., VENKATESH S., Video abstraction: a systematic review and classification, *ACM Trans. on Multimedia Computing, Communications, and Applications*, Vol. 3, No. 1, pp. 1-37, 2007.
- [31] CIOCCA G., SCETTINI R. , An innovative algorithm for key frame extraction in video summarization, *J. of Real-Time Image Proc.*, Vol. 1, No. 1, pp. 69-88, 2006.
- [32] CIOCCA G., SCETTINI R., Supervised and unsupervised classification post-processing for visual video summaries, *IEEE Transactions on Consumer Electronics*, Vol. 2, No. 52, pp. 630-638, 2006.
- [33] DUYGULU P., WACTLAR H.D., Associating video frames with text, *ACM SIGIR Conference*, 2003.
- [34] CRANDALL D., ANTANI S., KASTURI R., Extraction of special effects caption text events from digital video, *Int. Journal on Document Analysis and Recognition*, pp. 138-157, Vol. 5, 2003.
- [35] BERMAN A.P., SHAPIRO L.G., A flexible image database system for content-based retrieval, *Computer Vision and Image Understanding*, Vol. 75, pp. 175-195, 1999.
- [36] BONOMI E., COLOMBO A., RIZZINI D., Virtual Museum, <http://www.ivl.disco.unimib.it/docs/museo.pdf> , 2006.
- [37] MAIS Consortium, Mais: Multichannel Adaptive Information Systems. <http://black.elet.polimi.it/mais/>.

## OMERO

### UN SISTEMA PER PERMETTERE DI “TOCCARE” LA REALTÀ VIRTUALE

*Fabio De Felice, Floriana Renna, Giovanni Attolico, Arcangelo Distante*

Istituto di Studi sui Sistemi Intelligenti per l'Automazione (ISSIA)

Consiglio Nazionale delle Ricerche (CNR)

Via Amendola, 122/D - 70126 Bari, Italy

attolico@ba.issia.cnr.it

<http://www.issia.cnr.it/>

#### 1 Abstract

Questo capitolo descrive lo sviluppo di OMERO, un sistema che si propone come una piattaforma multi-modale aperta per la fruizione di contenuti informativi descritti tramite scene virtuali tridimensionali. La piattaforma è *multi-modale* in quanto utilizza l'interazione tattile per estendere le esperienze visive ed uditive normalmente utilizzate nelle applicazioni di realtà virtuale ed è *aperta* in quanto basata su un'architettura che permette l'aggiunta di nuove modalità d'interazione.

La modalità tattile permette un'efficace restituzione d'informazioni e dati di natura spaziale che ordinariamente sono percepibili principalmente attraverso il canale visivo: questo permette di realizzare applicazioni che offrano i vantaggi della realtà virtuale anche ad utenti che per le proprie limitazioni sensoriali (cecità o ipovisione) non possono accedere in modo soddisfacente alle abituali comunicazioni grafiche o pittoriche. Per rendere più flessibile e funzionale alla conoscenza l'esperienza dei modelli tridimensionali, il sistema prevede la possibilità di comporre modelli multi-livello (che permettono un accesso progressivo all'informazione), con componenti attive (per trasmettere l'informazione in modo tempestivo ed efficiente all'utente) e percorsi guidati di esplorazione (per migliorare l'esperienza in termini di sequenza temporale e completezza delle percezioni sensoriali). Globalmente il sistema si propone quindi, attraverso il polimorfismo e la ridondanza, di consentire a ciascun utente di selezionare in base alle proprie caratteristiche i canali comunicativi migliori ed offrire un'interazione più soddisfacente e significativa.

#### 2 Introduzione

Apprendere un concetto astratto, riconoscere le fattezze di una forma, ricordarsi la dislocazione di vari ambienti connessi fra loro sono compiti che richiedono la costruzione di uno schema mentale efficiente che copra tutte le caratteristiche della porzione di realtà o del concetto che è in esame. Normalmente possiamo dire che un contenuto informativo (concetto, forma, mappa ecc.) ci è chiaro in mente se riusciamo a “vederlo” in mente, ossia se abbiamo sempre nella nostra memoria un modello che lo descriva.

Questo modello mentale viene costruito mediante le informazioni che riceviamo tramite i nostri sensi (vista, udito, tatto, olfatto, gusto) e mediante il nostro intervento nell'ambiente circostante (gestualità, voce, movimento). Non sempre un contenuto informativo può essere offerto ai nostri sensi direttamente, si pensi a come un concetto astratto abbia bisogno di essere descritto testualmente e spiegato oralmente ma principalmente graficamente mediante modelli bidimensionali (grafici, schemi ecc.), o anche con modelli n-dimensionali a seconda dei parametri in gioco.

Un edificio non sempre può essere esplorato direttamente (per motivi di ubicazione, dimensione, sicurezza) e quindi abbiamo bisogno di foto, mappe, modellini in scala. Possiamo dunque dire che in generale ricorriamo ad una rappresentazione intermedia per apprendere un determinato contenuto informativo (porzione di realtà o concetto astratto).

Questa rappresentazione intermedia può essere fruita diversamente a seconda della tipologia di utenza a cui viene offerto: un non vedente farà più affidamento sul tatto, sulla voce e sull'udito, un ipovedente si affiderà, a vari gradi, anche alla vista, per un audio lesa sarebbero inutili le descrizioni vocali.

Per questo motivo tradizionalmente esistono vari supporti costruiti ad hoc per i vari tipi di utenza. Ognuno di questi supporti ha degli svantaggi che rendono l'accesso molto difficoltoso da parte di utenti con particolari disabilità sensoriali. Come esempio di queste limitazioni verrà riportato il caso dei disabili visivi, e inoltre verranno descritte le problematiche che coinvolgono anche i vedenti.

### **2.1 Problematiche dei non vedenti**

I disabili visivi sfruttano intensivamente i sensi quali l'udito e, principalmente, il tatto. La principale difficoltà dell'esperienza tattile è legata alla sua natura *seriale* rispetto a quella *parallela* dell'esperienza visiva. Osservando una fotografia la vista coglie nell'insieme quanto osservato e fornisce in modo immediato all'osservatore informazioni fondamentali: contesto, disposizione spaziale e identità dei singoli oggetti in essa presenti. Acquisite queste informazioni, i dettagli vengono colti in seguito focalizzando l'attenzione sui diversi particolari della scena e trovano il loro posto in un contesto generale ormai chiaro.

Se la medesima fotografia fosse osservata attraverso un piccolo foro fatto scorrere lungo di essa (che renda visibili in sequenza solo piccole parti dell'immagine) le percezioni, scollegate dall'insieme, verrebbero comprese ed interpretate con difficoltà molto maggiori e potrebbero addirittura risultare incomprensibili. Questa esperienza è l'esperienza ordinaria di un non vedente che tocca un oggetto e ne esplora in sequenza le diverse parti, dovendo compiere uno sforzo nella propria mente per derivare un'idea complessiva completa ed organizzata, uno schema mentale, di quello che sta toccando e delle relazioni tra i diversi particolari percepiti.

La scala alla quale i dettagli possono essere percepiti costituisce un altro limite. Avvicinandosi o allontanandosi da una scena è possibile renderne evidenti o nascondere i dettagli ai propri occhi. Ciò che può o non può essere percepito al tatto invece dipende dalla dimensione relativa dei particolari e dei polpastrelli della mano, nessuno dei quali può essere modificato nella realtà. Queste limitazioni riguardano l'interazione con qualunque oggetto reale, compresi gli ausili tradizionali, generalmente modellini in scala, in quanto legate alla natura intrinseca del tatto ed alla staticità di questi supporti.

Inoltre per un non vedente muoversi consapevolmente in ambienti indoor o outdoor, pianificare un percorso, acquisire padronanza della forma e dell'ubicazione di regioni geografiche sono bisogni che spesso non vengono risolti in maniera soddisfacente. Generalmente un non vedente accede a informazioni spaziali tramite l'ascolto della descrizione che ne fa un accompagnatore vedente o mediante messaggi acustici pre-registrati come gli audio libri o le audio guide, oppure utilizzando mappe in rilievo integrate con testi in Braille o, ancora, esplorando direttamente l'ambiente con l'assistenza di un bastone o di un cane guida. Con questi mezzi non è sempre possibile costruire in modo facile e rapido uno schema mentale efficiente dell'ambiente esplorato o di oggetti toccati.

## *2.2 Problematiche dei vedenti*

Per un utente normodotato la possibilità di fruire direttamente di informazioni può essere influenzata da vari fattori, primo fra tutti l'ubicazione geografica della stessa. Per conoscere una statua o un quadro bisogna essere sul luogo, perché anche se ci si può avvalere di foto, si può fare affidamento a modellini in scala o si può leggere a riguardo ma in generale questi canali indiretti non permettono mai una esperienza sensoriale completa e integrata e quindi possono portare a una conoscenza parziale. Altri tipi di ambienti possono essere inaccessibili per ragioni di sicurezza, sia per l'operatore umano sia per l'ambiente stesso, nel primo caso si pensi ad una situazione in cui le informazioni sono in un ambiente sconosciuto e potenzialmente ostile, in questi casi applicazioni multimodali aiutano a pilotare a distanza un attuatore (robot, veicolo) salvaguardando l'incolumità dell'operatore umano. Nel secondo caso possiamo pensare a ambienti delicati dal punto di vista dell'ecosistema come possono essere siti archeologici contenenti preziose informazioni storiche, in cui la sola presenza umana può essere fonte di danneggiamento, anche in questo caso non è possibile un accesso diretto se non a patto di una parziale perdita d'informazione dovuta all'intervento dell'uomo.

## *2.3 L'uso della realtà virtuale*

Il processo conoscitivo normalmente nasce da un'interazione tra il soggetto e la realtà attraverso i sensi. Le informazioni raccolte sono elaborate, interpretate e manipolate per costruire delle rappresentazioni mentali che raccolgono il senso di quanto conosciuto. La realtà virtuale si propone di offrire ai sensi del soggetto l'esperienza di modelli che sono sintetizzati all'interno dell'elaboratore. Un vantaggio banale derivante dall'uso della realtà virtuale è la possibilità di rendere l'esperienza indipendente dalla compresenza fisica del soggetto e dell'oggetto del processo conoscitivo: realtà che per la loro natura o ubicazione geografica o pericolosità o vulnerabilità non potrebbero essere esperite nel mondo fisico diventano conoscibili attraverso una loro modellazione opportuna. La modellazione della realtà è una fase delicata perché nella realtà virtuale (a differenza di quanto avviene nella realtà aumentata) ciò che il soggetto può percepire è solo e soltanto quanto della realtà è stato trasferito e rappresentato nel modello. Gli strumenti attualmente disponibili per la creazione di modelli 3D da oggetti reali (Laser 3D) o attraverso una creazione originale e creativa (Auto CAD, Maya) possono offrire molteplici possibilità: rappresentare la componente volumetrica degli oggetti e la loro rappresentazione pittorica superficiale non presenta particolari problemi.

Le applicazioni di realtà virtuale devono invece cogliere la sfida di rappresentare la realtà in modo da facilitarne la percezione e la comprensione. Questo può richiedere l'integrazione delle informazioni ordinarie sulla scena con dati rilevati attraverso strumenti specialistici: il modello aggiunge quindi alle informazioni di forma o di finitura pittorica delle superfici, delle misure acquisite attraverso strumentazione particolare o rappresentazioni di analisi interpretative svolte sulla realtà fisica. L'esperienza virtuale non è quindi più un'approssimazione, buona quanto si vuole, della realtà corrispondente ma diventa un complemento dell'esperienza reale, permettendo una conoscenza ed una comprensione superiori e personalizzati in dipendenza degli interessi e delle capacità dell'utente.

In questa ottica la realizzazione di un sistema che permetta, attraverso la multi-modalità, un'esperienza di realtà virtuale anche ad utenti con deficit sensoriali non diventa un semplice strumento di integrazione di questi utenti verso applicazioni di informatica di frontiera, ma permette di realizzare ed offrire nuove opportunità per dischiudere orizzonti di apprezzamento e comprensione della realtà che sarebbero altrimenti riservati agli utenti dotati della vista.

Sotto questo punto di vista i modelli digitali possono permettere il superamento delle limitazioni degli artefatti materiali (modelli in scala): sono fruibili senza necessità di spostare oggetti fisici e possono essere modificati (in dimensione, livello di dettaglio, informazioni rappresentate) con relativa semplicità ed economia di costo. La loro dinamicità permette con più facilità di rappresentare fenomeni che evolvono nel tempo.

Nuove tecnologie stanno offrendo nuove possibilità di interazione tra gli utenti ed i sistemi digitali. Le interfacce aptiche [1] permettono all'utente di percepire i modelli virtuali 3D anche attraverso il tatto, applicando alla sua mano ritorni di forza che simulano realisticamente il contatto con geometrie reali. Sistemi di TextToSpeech e di riconoscimento vocale automatico rendono possibili canali comunicativi più naturali rispetto a tastiera, mouse e monitor. Essi possono incrementare la quantità e il tipo d'informazione che può essere offerta all'utente. Le applicazioni multi-modali sfruttano differenti canali informativi (visivo, tattile, uditivo ecc.) in un modo integrato e ridondante [2]. La ridondanza traduce la rappresentazione della stessa informazione verso diverse modalità di restituzione (in maniera polimorfa quindi) in modo che l'utente possa configurare l'interazione che più si adatta alle sue specifiche capacità, permettendo anche agli utenti con limitazioni sensoriali di accedere ad applicazioni della realtà virtuale che possono semplificare molti aspetti del vivere quotidiano.

#### *2.4 Riferimenti a lavori simili e commenti*

Le interfacce aptiche hanno trovato numerose applicazioni anche verso utenti vedenti ed in diversi contesti: fruizione/conservazione di beni culturali, addestramento (telecontrollo, chirurgia, ...), progettazione industriale (progettazione meccanica, verifica di prototipi, ...). Un primo esempio di sistema multi-modale è descritto in [3]: il sistema SHARC utilizza in modo integrato la visione stereoscopica ed il feedback acustico e tattile. Il sistema è stato applicato alla costruzione di un ambiente virtuale utilizzando mattoncini virtuali in stile Lego. Il Museo delle Pure Forme [6] rende disponibili un'ampia gamma di modelli virtuali di opere d'arte i cui originali sono ubicati in diversi musei sparsi geograficamente: con essi è possibile interagire con il tatto che complementa un'immagine stereoscopica dell'opera. Il sistema è utilizzabile anche da parte dei non vedenti in quanto rende disponibile la modalità tattile; non offre però alcun esplicito supporto alle difficoltà di percezione precedentemente descritte, offrendo sostanzialmente l'equivalente digitale di un artefatto fisico tradizionale. In [4] viene presentato il sistema multi-modale Haptic Desktop che ha l'obiettivo di costituire un'alternativa ai tradizionali dispositivi di input utilizzati per accedere alle risorse del computer: esso integra la visualizzazione grafica con un sistema aptico bidimensionale e con un ritorno acustico. Un esempio di visualizzazione multi-modale di grafici per l'apprendimento di concetti matematici può essere trovato in [6]. L' Accessible Graphic Calculator estende l'uso dei grafici ad utenti con problemi visivi tramite una trasduzione sonora delle informazioni: le frequenze del suono sono fatte variare in modo proporzionale ai valori delle y. L'andamento del grafico prodotto è reso percepibile attraverso una stampante braille.

Sono stati proposti vari approcci per estendere l'accessibilità ai non vedenti di dati spaziali tridimensionali: in [7] viene presentato un sistema per promuovere la mobilità autonoma dei non vedenti. Il sistema, chiamato MoBIC, è composto da due componenti: di addestramento e di navigazione. La prima aiuta l'utente a pianificare un percorso in un'area urbana: le informazioni di percorso così costruite vengono passate al sistema di navigazione che guida l'utente durante la passeggiata reale.

Un altro sistema, il BATS, è basato su interfacce aptiche ed è fondato sullo studio di mappe bidimensionali [8]. Altri sistemi basati totalmente su interazione aptica si possono trovare in [9] [10] [11]. In [9] viene proposto un sistema per l'esplorazione del modello della mappa di un campus costruita in OpenGL, ma l'esperienza è limitata ad una interazione aptica elementare. In [10] [11] viene sfruttato l'uso combinato di interazione aptica e acustica per la fruizione di dati concernenti ambienti urbani in presenza di traffico. In [10] viene presentato il progetto e l'iniziale implementazione di un'interfaccia per l'addestramento alla mobilità.

Questo progetto dovrebbe permettere all'utente di simulare una passeggiata lungo il modello virtuale di una strada reale in cui informazioni aptiche riguardo la conformazione della strada e messaggi vocali riguardo agli ambienti che vi si affacciano vengono forniti all'utente man mano che esso procede nell'esplorazione. L'applicazione descritta in [11] è caratterizzata da un ambiente altamente strutturato e dinamico che riproduce una reale parte di un centro cittadino che l'utente deve esplorare per apprendere un percorso. L'applicazione è stata sperimentata facendo esplorare prima virtualmente e poi nella realtà l'ambiente riprodotto con risultati soddisfacenti.

Il sistema Omero differisce da tutti questi lavori in quanto non indirizza una particolare istanza del problema (esplorazione di un oggetto, pianificazione di un percorso, conoscenza di ambienti complessi, ...) ma propone una serie di strumenti volti a migliorare la fruizione del contenuto informativo quale ne sia la natura.

La sperimentazione finora ha visto l'impiego nella fruizione di informazioni spaziali 3D, nel quale ambito il sistema ha dato prova di poter gestire sia ambienti virtuali indoor che mappe geografiche. In particolare una planimetria può essere offerta ad un non vedente tramite un'appropriata ricostruzione 3D, in cui ogni componente ha una sua particolare controparte aptico/acustica.

## ***2.5 Impostazione del sistema multimodale***

Il sistema è stato ideato e progettato per offrire all'utente un'esperienza arricchita rispetto alla semplice esplorazione cinestetica diretta. Coerentemente con una visione della realtà virtuale non come semplice approssimazione più o meno completa della realtà equivalente ma come una rappresentazione complementare che deve aiutare la comprensione e la conoscenza, il sistema cerca di sfruttare l'adattabilità dei modelli digitali e la versatilità dei vari tipi d'interazione integrata aptico/acustica per ridurre al minimo le problematiche che interessano l'esperienza tattile nella realtà fisica. A tale scopo diverse funzionalità e strumenti sono stati progettati e implementati.

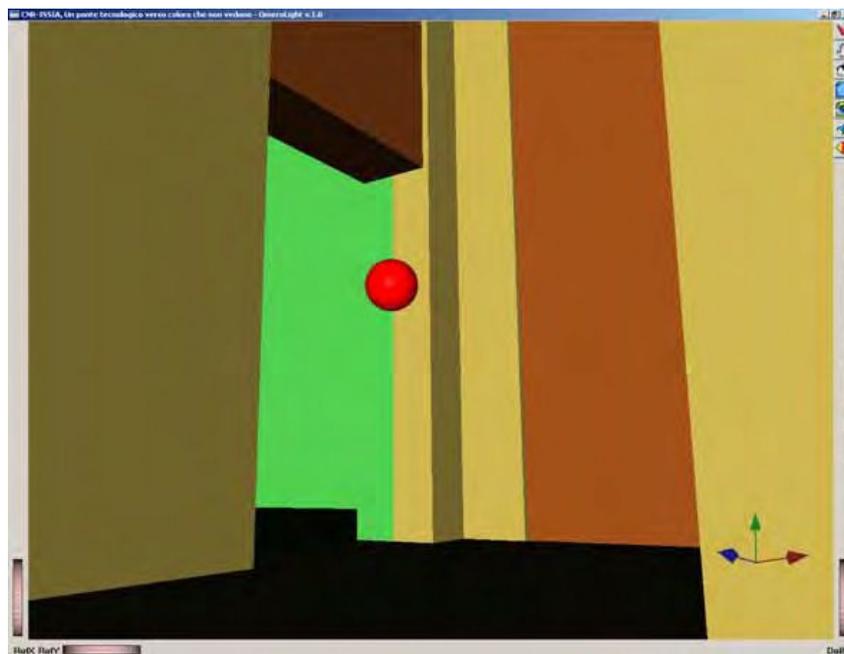
*Effetti aptici.* Oltre alla simulazione del contatto con la geometria tridimensionale della realtà modellata, il sistema utilizza estensivamente il ventaglio delle sensazioni tattili che possono essere generate: forze attrattive (per guidare verso l'oggetto o lungo percorsi di esplorazione predefiniti), vibrazioni o viscosità (utili per evidenziare la presenza di alcune parti del modello) ecc. Questo permette di utilizzare il canale aptico, non verbale ed immediato, anche per trasmettere all'utente la presenza e l'identità di oggetti e le relative informazioni.

*Effetti acustici.* Possono essere usati per aumentare il realismo e la significatività del contatto con la scena oppure sintetizzati per comunicare informazioni attraverso suoni o messaggi testuali.

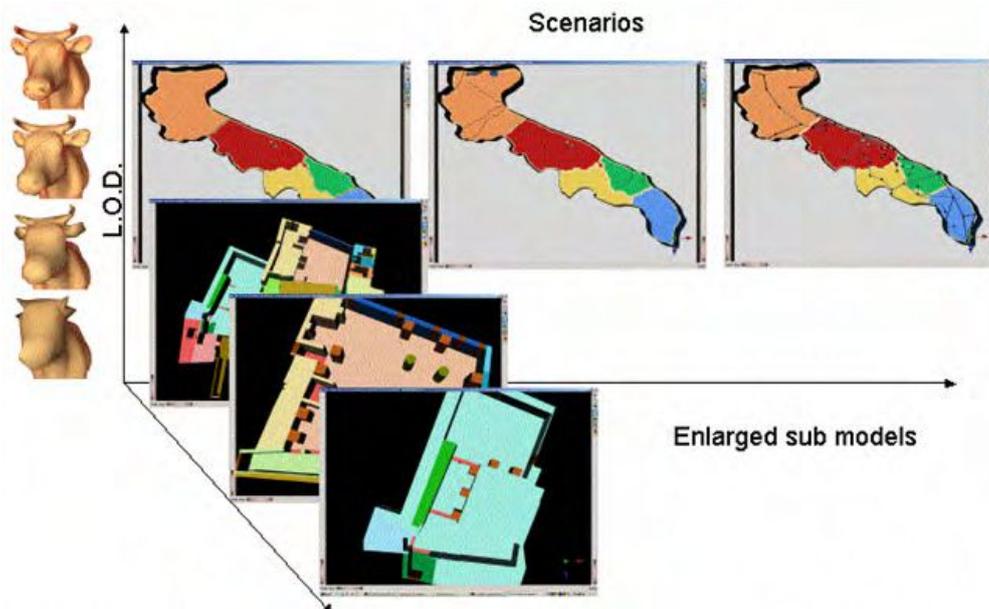
*Oggetti attivi.* L'esplorazione può essere migliorata aggiungendo ad alcuni oggetti un certo grado d'interattività, rispetto alla restante parte della geometria della scena che rimane passiva. Questi oggetti attivi possono essere classificati come: *aptici, acustici o aptico/acustici*, in dipendenza dagli effetti ad essi associati (effetti aptici, effetti acustici o effetti più complessi risultanti da una combinazione di sensazioni aptiche ed acustiche). Ad ognuno di essi può essere ulteriormente associato un messaggio vocale per trasmettere informazioni più complesse di tipo testuale (es. dati storici, artistici, ...).

Gli oggetti attivi possono essere inoltre dinamici o statici: per gli oggetti dinamici è possibile specificare un comportamento che ne cambi posizione e/o assetto all'interno della scena. Effetti, messaggi vocali e comportamento dinamico possono essere attivati automaticamente al semplice contatto tra l'avatar che rappresenta l'utente e l'oggetto attivo, oppure su richiesta dell'utente. Un esempio di oggetto attivo aptico/acustico dinamico è costituito dalle porte nei modelli raffigurati in Figura 1: la porta viene associata ad una vibrazione che permette di distinguerla dalle pareti e, su richiesta dell'utente, può sollevarsi per lasciarsi attraversare dall'avatar e poi chiudersi alle sue spalle, con movimenti che sono sottolineati da un suono appropriato.

*Accesso progressivo all'informazione.* Una quantità troppo grande di informazioni fornite tutte insieme in un unico modello complesso può ingenerare confusione: questo può accadere anche ai vedenti alle prese con immagini troppo complesse ma risulta particolarmente problematico per chi deve esplorare la scena attraverso un senso seriale come il tatto. Il sistema Omero utilizza modelli appositamente organizzati per ripartire l'informazione tra molteplici livelli semantici chiamati *scenari*. La possibilità di definire un oggetto *toccabile* o *non toccabile* rende possibile variare dinamicamente la percepibilità dei singoli scenari: in questo modo i dettagli possono essere aggiunti in modo progressivo (ed eventualmente rimossi se opportuno) mostrando in ogni momento solo le informazioni a cui l'utente è interessato.



**Figura 1 - L'immagine raffigura l'avatar dell'utente (la sfera rossa) mentre attraversa la soglia di una porta dopo averla aperta**



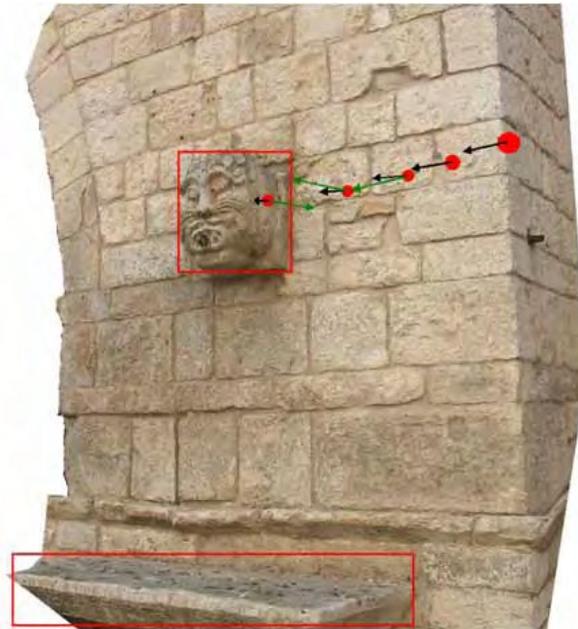
**Figura 2 - Schema di organizzazione per la presentazione progressiva della scena virtuale**

Questa possibilità, comune per la visualizzazione di modelli per vedenti, viene resa possibile così anche ai non vedenti in modo più versatile ed efficiente rispetto alla costruzione ed esplorazione di molteplici artefatti fisici corrispondenti ai diversi scenari. Un analogo discorso può essere fatto sul livello di approssimazione del modello (Level of Details – LOD) rispetto alla geometria della realtà: non solo quindi aggiungere/rimuovere dettagli ma rendere più fine o grossolana la modellazione dell’oggetto di interesse se questo è funzionale a facilitarne la comprensione da parte dell’utente. Anche la scala è un elemento su cui è possibile operare in digitale: attraverso ingrandimenti del modello (che variano la proporzione relativa tra oggetto e polpastrello) si possono rendere percepibili dettagli molto fini che nella realtà sarebbero indistinguibili. È possibile quindi delineare uno spazio tridimensionale in cui ogni asse definisce una modalità di semplificazione applicata al modello come descritto in Figura 2.

*Percorsi guidati.* Il successo nella costruzione di un modello mentale della realtà completo e significativo può essere fortemente influenzato dai percorsi di esplorazione seguiti dal non vedente [13]. È possibile esercitare delle forze attrattive in modo da guidare la mano dell’utente verso una particolare parte del modello o vincolarla a seguire un determinato percorso, che garantisca la visita di tutti gli aspetti più rilevanti del modello, come schematizzato in Figura 3: si può quindi simulare una visita guidata della scena. La possibilità di far dipendere le aree di maggiore interesse da indicazioni fornite da un vedente attraverso il mouse apre un nuovo canale di comunicazione tra vedenti e non vedenti che si incontrano nel mondo virtuale che ciascuno percepisce attraverso le modalità ad esso più consone.

## 2.6 Applicazioni sperimentate

Le funzionalità descritte sono state implementate e verificate attraverso applicazioni specificamente progettate per permettere di capire quanto potessero essere valide ed efficienti per gli utenti finali.



**Figura 3 - Una schematizzazione di un percorso guidato. Le zone all'interno del rettangolo sono i target d'interesse della visita**

Una prima sperimentazione è servita per verificare quanto il dispositivo aptico adottato, un PHANTOM Desktop [14] fosse adatto agli scopi prefissati. Ad un gruppo di dieci utenti (ipovedenti e non vedenti, con diverse storie e livelli di familiarità con gli strumenti informatici) è stato proposto un modello virtuale costruito componendo dei semplici solidi geometrici. Tutti loro, seppure con diversi gradi di difficoltà, hanno potuto riconoscere i solidi toccati.

Ad un più ampio numero di utenti partecipanti ad una mostra nazionale [15], è stato sottoposto il modello di un particolare di un edificio storico rappresentante una gogna (vedi Figura 3) per verificare l'utilità dei percorsi guidati e della interazione aptico/acustica.

La capacità del sistema di assistere la navigazione attraverso realtà tridimensionali complesse è stata intensivamente verificata in altre tre sperimentazioni. La prima è stata indirizzata all'esplorazione della planimetria di un appartamento, un modello relativamente semplice, composto da ambienti di forma rettangolare tutti allineati agli assi, con porte interattive e che ha dato risultati positivi in quanto ogni utente non vedente è riuscito ad acquisire la confidenza necessaria a muoversi autonomamente tra le stanze in un tempo largamente inferiore a quello richiesto da un'esplorazione reale.

L'evoluzione di quest'applicazione è stata applicata all'esplorazione della planimetria di una larga parte del piano terra del Castello Svevo di Bari. Questo modello ha rappresentato una prova notevolmente più complessa per i non vedenti: esso presenta una grande varietà di forme e disposizioni per gli ambienti alcuni anche molto piccoli.

Un'ultima applicazione infine ha interessato l'esplorazione multi-livello della mappa della regione Puglia, con la rappresentazione esemplificativa di aspetti tanto legati alla morfologia ed agli elementi naturali che agli interventi antropici sul territorio. Queste ultime due [16], maggiormente complesse ed emblematiche delle potenzialità dello strumento, saranno dettagliate nel seguito.

### 3 Architettura

#### 3.1 Scelte realizzative

Il sistema Omero è stato progettato come un framework generico per l'esplorazione di modelli 3D cui sono associati particolari contenuti informativi.

Tutti gli strumenti implementati devono poter essere utilizzabili in maniera integrata durante l'esplorazione. Per questo motivo è stato definito uno schema entità relazione (Figura 4) che lega le *features* descritte in un insieme organico. Questa organizzazione va a definire una standardizzazione della scena virtuale al fine di rendere univocamente localizzabili gli oggetti che diverranno attivi e che comporranno scenari e sottomodelli.

Una scena può essere suddivisa in più sotto-modelli, ognuno dei quali può descrivere una parte della realtà modellata, se questa è molto vasta. In questi sotto-modelli possono anche trovare posto versioni ingrandite di particolari del modello globale. Ognuno di questi sotto-modelli può essere organizzato in scenari, ognuno dei quali definisce una vista appropriata di una parte del contenuto informativo del sotto-modello. Formalmente uno scenario è una collezione di oggetti attivi legati fra loro da una certa semantica. Ad ogni scenario può essere associato un determinato effetto, che può essere aptico e/o acustico, che viene detto "globale" ed è sempre attivo durante l'esplorazione dello scenario.

Oltre all'effetto globale può essere definita una visita guidata, in termini di oggetti dello scenario che definiscono una sequenza di punti di interesse (target) lungo un percorso. Gli oggetti possono essere passivi o attivi (cioè associati ad un effetto, che chiameremo locale, aptico e/o acustico ed eventualmente legati ad un determinato comportamento dinamico). Ogni oggetto attivo può essere coinvolto nella visita guidata associata allo scenario. Per ogni oggetto (attivo o passivo) è possibile definire degli attributi di materiale (rigidità, frizione ecc.). I modelli 3D in ingresso possono essere creati specificamente per il sistema oppure possono essere scaricati dal web: il formato in ingresso è il VRML, il formato più utilizzato per la rappresentazione e lo scambio di contenuti 3D (anche se nuovi standard si stanno affermando quali X3D [17] e Collada [18]).

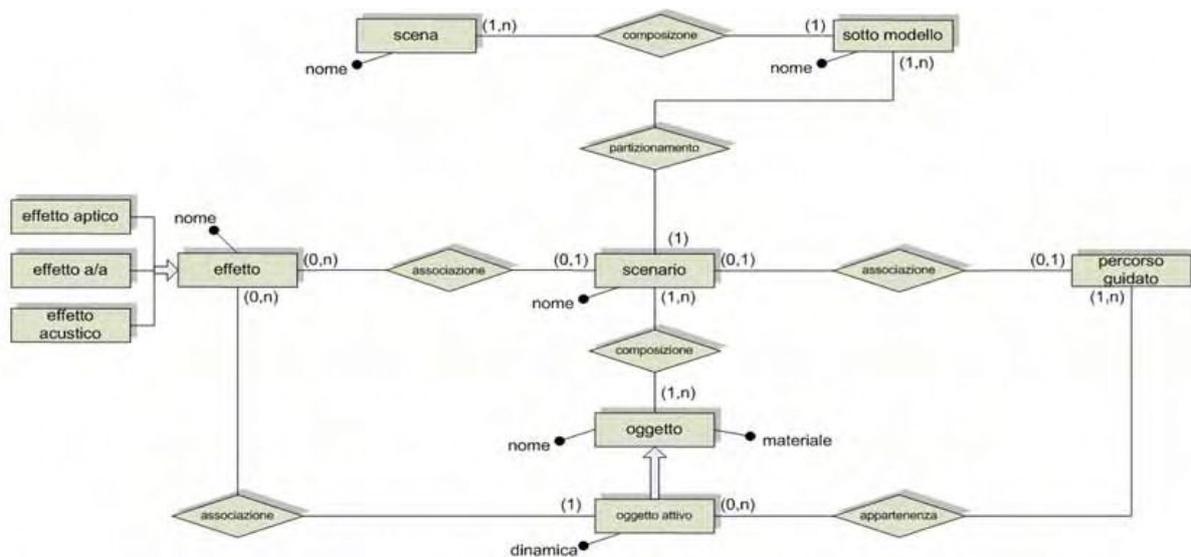


Figura 4 - Il modello entità relazioni che descrive come è organizzata la scena virtuale

### 3.2 Indipendenza dal modello in ingresso

Per soddisfare l'organizzazione descritta in precedenza, il modello in ingresso deve essere costruito opportunamente. Il formato VRML in ingresso permette una descrizione della scena di tipo gerarchico mediante scene graph [19]. In questo modo la scena virtuale viene descritta tramite un albero in cui i nodi intermedi definiscono i vari raggruppamenti di oggetti e altre informazioni riguardanti le caratteristiche di visualizzazione della scena. In ultimo i nodi foglie descrivono le geometrie degli oggetti costituenti la scena. Un particolare nodo di raggruppamento che assume particolare importanza nel funzionamento di Omero è il nodo detto di Switch, il quale permette di discriminare quale dei suoi sotto rami deve essere visitato per la visualizzazione. In questo modo raggruppando opportunamente gli oggetti nella scena si possono creare i sotto modelli e gli scenari relativi. Un'opportuna associazione con i dispositivi di input disponibili all'utente permette di passare da un contesto ad un altro dinamicamente. La scena virtuale può anche essere un modello preesistente, in questo caso l'organizzazione non incontrerà i requisiti definiti nello schema in 2.1 e quindi ci sarà il bisogno di riorganizzare opportunamente il tutto.

Si faccia l'esempio della gogna (Figura 5): inizialmente l'oggetto può essere descritto come una unica forma associata ad un unico nodo nel scene graph, se si desidera definire un percorso guidato sarà necessario estrarre la descrizione relativa al volto leonino ed al sedile al fine di renderli oggetti separatamente identificabili all'interno della scena.

Per aumentare la flessibilità nella descrizione dei modelli in modo trasparente rispetto all'applicativo che deve leggerli e restituirli in modo multi-modale, si è deciso di dividere la parte del modello che descrive il "cosa" (struttura geometrica ed organizzazione gerarchica di sotto-modelli e scenari), descritto in un file VRML, dal "come" (ovvero come i singoli componenti vengono associati agli specifici metodi di restituzione verso l'utente), descritte in un file XML: si possono quindi associare comportamenti diversi a scene diverse ma anche comportamenti diversi per una stessa scena in esecuzioni diverse, rendendo più adattabile l'approccio. Durante la progettazione di una scena (compresa la sua organizzazione in sotto-modelli, scenari, oggetti attivi, ...) è importante stabilire i contenuti informativi che si vogliono trasmettere all'utente. È quindi di particolare importanza l'esperto del domino applicativo per definire quali aspetti sottolineare e quali informazioni devono essere associate ai diversi oggetti.

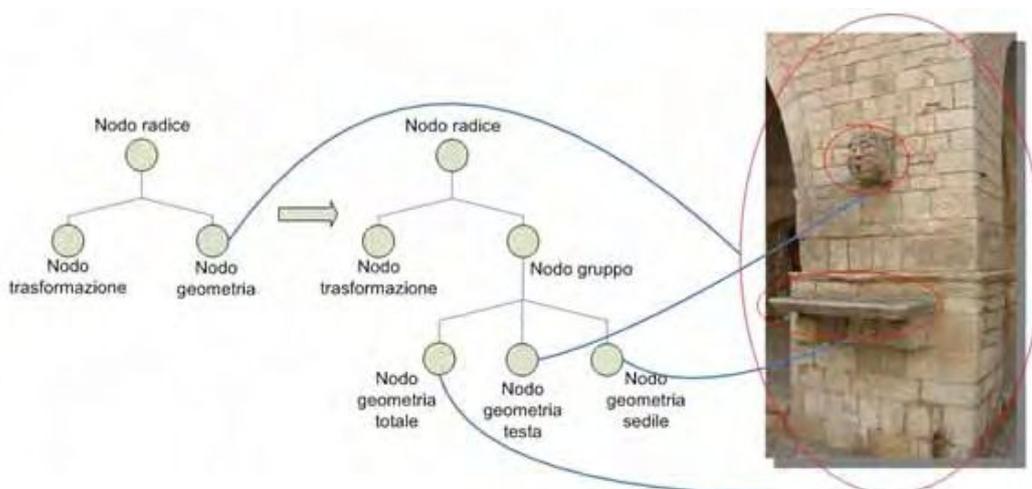
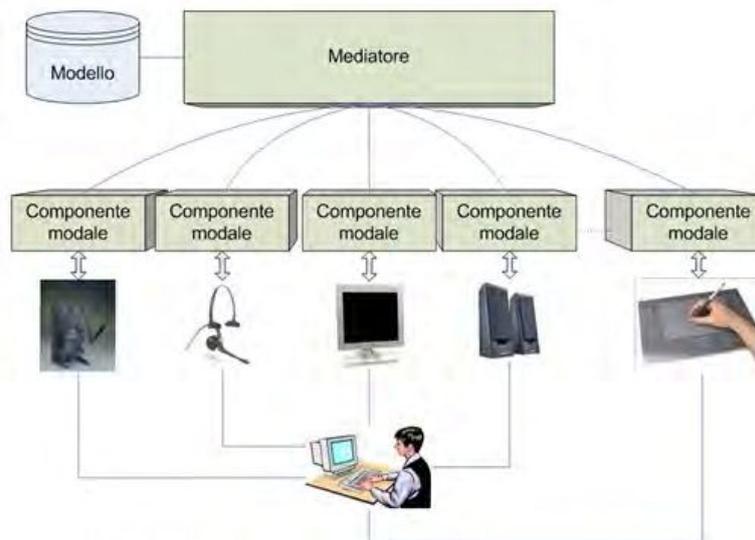


Figura 5 - Esempio di riorganizzazione della scena virtuale



**Figura 6 - Architettura del sistema**

### 3.3 Moduli e relazioni tra moduli

Il sistema vuole essere aperto all'aggiunta e integrazione di nuove modalità d'interazione, indipendente rispetto al modello in ingresso e dal dispositivo d'interazione modale.

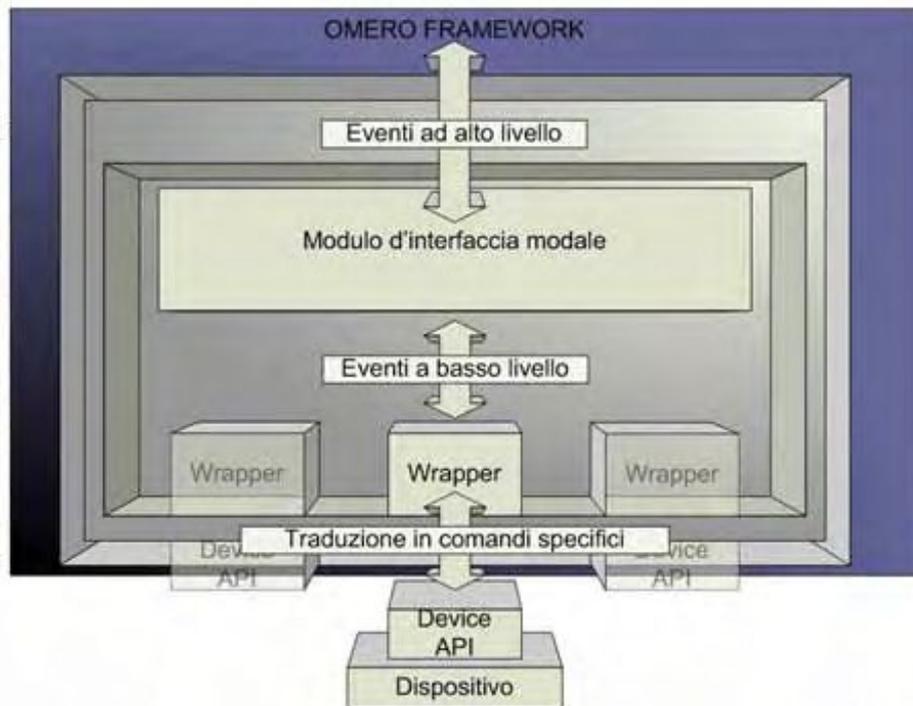
Si può pensare al sistema come un a un middle-ware, tra la scena virtuale caricata e i vari canali informativi.

L'architettura riportata in Figura 6 sottolinea come il modello sia caricato dal file esterno all'interno di un modulo principale chiamato *Mediatore*, cui sono collegati i singoli *Componenti Modali* che gestiscono le particolari interazioni con l'utente. L'architettura può facilmente integrare un nuovo componente modale (un nuovo dispositivo d'interfaccia).

Il modulo Mediatore attiva i componenti modali e gestisce la comunicazione fra essi. Il funzionamento generale è determinato da eventi che i vari componenti modali generano durante l'interazione con l'utente.

Il Mediatore riceve tutti questi eventi (comandi o risposte a comandi) e decide quale evento generare in risposta. Ogni singolo componente modale si astrae rispetto al particolare dispositivo d'interazione che deve gestire: il *Modulo d'Interfaccia Modale* definisce l'API d'interfaccia verso il modulo Mediatore mentre il *Wrapper*, incapsula la particolare API proprietaria relativa al dispositivo fisico. Così il sistema si rende indipendente dal particolare dispositivo adottato per una determinata interazione.

L'architettura interna di un Componente modale è riportata in Figura 7.



**Figura 7 - Architettura del componente modale**

### 3.4 Implementazione corrente

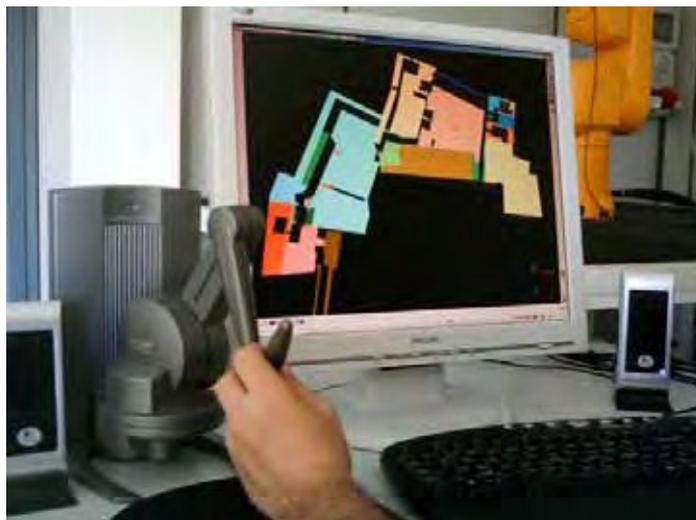
Nella corrente versione di Omero sono stati utilizzati il PHANToM Desktop e PHANToM Omni come dispositivi aptici gestiti tramite due Wrapper uno per l'API GHOST e l'altro per l'API OpenHaptics, la libreria Coin3D per la visualizzazione grafica e la gestione della scena virtuale [19] e la funzionalità TextToSpeech del sistema LOQUENDO [20].

## 4 Sperimentazione con il Castello Svevo di Bari

### 4.1 Caratteristiche del contesto

Un'esigenza sentita da parte dei non vedenti è la possibilità di orientarsi autonomamente in ambienti indoor e outdoor. Per tale motivo la prima sperimentazione è stata effettuata per verificare le capacità del sistema Omero di permettere la conoscenza di un ambiente complesso quale il Castello Svevo di Bari attraverso l'esplorazione virtuale di un suo modello tridimensionale appropriato.

Nel modello è stata inserita una parte della planimetria del piano terra del Castello, selezionando gli ambienti che di solito sono aperti al pubblico: il ponte di ingresso, la biglietteria, la galleria che dall'ingresso porta al cortile esterno, la parte del cortile esterno attraverso cui si accede al complesso centrale del castello, la zona di ingresso al cortile interno, il cortile interno, la gipsoteca, la cappella, i bagni e un passaggio che collega il cortile interno con quello esterno. Di tutti gli altri ambienti che affacciano sul cortile interno sono state riprodotte solo le porte di comunicazione con il cortile stesso, accompagnate da un messaggio che ne chiarisse l'identità. I dati digitali sono stati estratti da un modello CAD realizzato a partire da rilievi svolti in epoca recente.



**Figura 8 - L'interfaccia multi-modale utilizzata da un utente non vedente per interagire con la planimetria 3D di un edificio storico, il Castello Svevo di Bari**

La planimetria originale, in scala 1:200, è stata modificata eliminando le zone non di interesse per i visitatori e per ottenere, per quelle di interesse, una rappresentazione 3D che fosse esplorabile attraverso l'uso di effetti aptici e/o acustici. In Figura 8 si può vedere Omero operare sul modello realizzato.

#### ***4.2 Funzionalità realizzate e sperimentate***

Nel modello realizzato, oggetti come alberi, pozzi, vasi, ecc. sono stati definiti come oggetti acustici attivi ed associati a semplici forme solide. Le porte apribili ed attraversabili sono state definite come oggetti aptico/acustici attivi e dinamici mentre quelle di accesso ad ambienti non visitabili in virtuale sono state definite come oggetti statici ed associate ad opportuni messaggi per indicarne la non praticabilità. Le aree di passaggio, varchi senza porte tra ambienti, sono state definite come oggetti attivi statici aptico/acustici ed associate a delle leggere cunette che le rendono riconoscibili anche apticamente. Mentre la vibrazione associata avverte l'utente della loro presenza, un messaggio vocale identifica gli ambienti alle due estremità del varco. I messaggi vocali sono attivati automaticamente al passaggio oppure su richiesta esplicita fatta premendo un opportuno pulsante sullo stilo dell'interfaccia aptica.

In questo modello è stata verificata l'utilità di ingrandire delle parti della planimetria: oltre al modello complessivo rappresentato in Figura 8, l'utente può esplorare modelli ingranditi di alcuni ambienti di particolare interesse, nei quali è più agevole cogliere i dettagli, le proporzioni e le forme. Durante la sperimentazione sono state inoltre confrontate due versioni del modello complessivo del Castello: il primo che includeva tutti gli oggetti attivi al proprio interno ed il secondo limitato invece ai soli ambienti, porte e varchi. La seconda versione è stata progettata per enfatizzare la topologia del Castello, prescindendo dai dettagli che l'utente ha potuto rilevare e introdurre nel proprio schema mentale durante le visite delle versioni ingrandite delle singole parti. Sono stati anche valutate e confrontate diverse modalità di attivazione dei messaggi vocali, diverse velocità di pronuncia delle frasi e diverse lunghezze dei singoli messaggi.

### 4.3 Sperimentazione e commenti

Due sessioni di test sono state condotte su due differenti gruppi composti ciascuno da quattro non vedenti che non avevano mai visitato il Castello prima: le sessioni sono state video-registrate per permetterne l'analisi a posteriori. Durante il primo test è stato usato un modello del castello formato dalla sola rappresentazione complessiva degli spazi con tutti gli oggetti attivi al proprio interno, senza usare ingrandimenti; i messaggi associati agli oggetti attivi erano attivati dal contatto tra avatar e oggetto mentre i messaggi che identificavano l'ambiente nel quale si trovava l'utente erano attivati a richiesta. Agli utenti è stato chiesto di esplorare liberamente il modello verbalizzando le proprie impressioni. Dopo la visita virtuale è stato loro chiesto di riassumere l'organizzazione degli spazi e di raggiungere determinati punti della planimetria, al fine di verificare se lo schema mentale costruito fosse completo e corretto. Dopo l'esplorazione virtuale ai non vedenti è stato chiesto di compiere una visita reale, durante la quale è stato chiesto di esprimere le aspettative circa la disposizione dei diversi ambienti lungo il percorso deciso in modo libero ed autonomo.

Al gruppo di non vedenti coinvolti nel secondo test è stato invece proposto un modello globale del castello spoglio di tutti gli oggetti attivi a parte le porte ed i varchi. Sono state rese fruibili le versioni ingrandite degli ambienti principali, nei quali sono stati riposizionati tutti i relativi oggetti attivi. In questa seconda esperienza sono stati utilizzati messaggi vocali più stringati e veloci. Gli utenti hanno svolto una visita virtuale preliminare (costituita dall'esplorazione del modello generale seguita da quella dei diversi ingrandimenti) e poi la visita reale di nuovo organizzata e condotta autonomamente e liberamente. In entrambe le fasi di test ciascun non vedente ha utilizzato l'ausilio preferito per il movimento (accompagnatore, cane guida, bastone, ...): anche nel caso dell'accompagnatore questi si è limitato a segnalare piccoli ostacoli non modellati nel virtuale (rami, piccoli gradini, ...) e non ha interferito nella pianificazione e nella conduzione della visita.

La prima sessione di prove ha messo in evidenza come un modello complesso come quello della planimetria del castello pone difficoltà dovute principalmente alle dimensioni e alla disposizione degli ambienti dalla geometria piuttosto irregolare. Inoltre dovendo rappresentare una superficie molto vasta nello spazio di lavoro del dispositivo aptico gli ambienti erano abbastanza piccoli. Una difficoltà ulteriore è stata costituita dalla presenza di oggetti distribuiti nell'intero modello del castello, ad ognuno dei quali è stato associato un opportuno messaggio audio che si attivava al contatto con l'avatar. La combinazione di spazi piccoli e irregolari e la frequente attivazione di messaggi audio ha reso abbastanza ardua e difficoltosa l'esplorazione del modello da parte dei non vedenti, alcuni dei quali hanno sottolineato l'importanza degli oggetti presenti lungo il perimetro degli ambienti perché possono costituire dei validi punti di riferimento mentre in generale hanno prestato poca attenzione agli oggetti che si trovavano nel mezzo degli ambienti (peraltro difficili da incontrare nella realtà sul percorso preferenziale lungo il perimetro degli spazi). Tutti gli utenti sono riusciti, anche se con gradi diversi di precisione, a riepilogare la disposizione degli ambienti che si affacciano sul cortile interno e sull'area di ingresso che comprende il ponte, l'ingresso e la biglietteria. La parte più difficile da comprendere è stata quella relativa alla zona di passaggio tra il cortile esterno e quello interno. In generale, il pulsante presente sullo stilo non è stato utilizzato spesso al fine di conoscere la propria posizione, forse a causa dei messaggi audio attivati frequentemente in seguito al contatto con i passaggi e gli oggetti.

La visita reale effettuata subito dopo ha sostanzialmente confermato l'utilità della visita virtuale preliminare; infatti, tutti gli utenti avevano già un'idea di quello che avrebbero visitato e di come orientarsi di volta in volta.

Le maggiori difficoltà sono state dovute al fatto che non è stato semplice riportare la dimensione degli ambienti virtuali a quella degli ambienti reali. Un altro problema incontrato dai non vedenti è stato quello dell'orientamento nell'ambiente reale, dovuto anche ai sistemi di riferimento diversi utilizzati nelle due visite. Infatti, mentre nella visita virtuale il sistema di riferimento con il quale orientarsi rispetto al modello è unico ed ha un'orientazione costante, nella visita reale quando l'utente cambia la propria direzione deve essere in grado di ruotare in accordo con lo schema mentale costruito con la visita virtuale la qual cosa non è sempre agevole. Alcuni utenti hanno manifestato il desiderio di ripetere la visita virtuale dopo quella reale in modo da affinare lo schema mentale che si erano costruiti.

Durante la seconda sessione di test, ai non vedenti inizialmente è stato proposto il modello della planimetria dell'intero castello con porte e passaggi, ma senza nessun oggetto. In questo modo, gli utenti hanno potuto esplorare il modello concentrandosi sulla disposizione spaziale delle stanze e sull'interazione aptico/acustica con le porte e con i passaggi, senza farsi distrarre dalla presenza degli oggetti e dai relativi messaggi. Dopo aver completato l'esplorazione di questo modello, hanno avuto la possibilità di scegliere le zone del castello da rivisitare utilizzando il modello ingrandito, comprendente tutti gli oggetti. I tempi per l'esplorazione del modello della planimetria del castello durante la seconda sessione di test sono risultati più bassi rispetto a quelli della sessione precedente; questo può essere dovuto ad un'interazione più efficiente degli utenti con quest'ultimo modello rispetto a quella che si aveva con il modello più dettagliato. Anche la capacità di riepilogare la disposizione degli ambienti dopo la visita virtuale è migliorata rispetto alla sessione precedente. Gli utenti hanno usato più di frequente il pulsante sullo stilo per ascoltare i messaggi. I messaggi audio erano stati resi più stringati. Il modello ingrandito dei bagni che nel modello globale sono piuttosto piccoli, è risultato essere quello più visitato. In particolare, un utente dopo aver esplorato quasi tutti gli ambienti ingranditi, alla fine di ogni singola esplorazione è sempre ritornato nel modello globale per integrare in esso i dettagli appena acquisiti. Questo gli ha permesso di comprendere pienamente la disposizione degli oggetti nel cortile interno e di orientarsi tra essi durante la visita reale. Anche durante la seconda sessione di test, gli utenti sono riusciti a portare a termine i compiti assegnati in modo soddisfacente anche se con diversi gradi di autonomia. Le visite reali sono state caratterizzate da un maggior grado di autonomia rispetto a quelle effettuate nella prima sessione; gli utenti hanno inoltre confermato di ritenere utile sia la visita virtuale preliminare sia quella successiva.

Gli utenti hanno giudicato in modo molto positivo il sistema: infatti, avere la possibilità di costruirsi una conoscenza completa e organica di un ambiente, poter pianificare una visita reale avendo conoscenza a priori di quello che si incontrerà nel mondo reale costituisce un vantaggio tangibile che può avere una ricaduta immediata sulla completezza delle esperienze reali.

## **5 Sperimentazione con la regione Puglia**

### ***5.1 Caratteristiche del contesto***

Una seconda sperimentazione con i non vedenti è stata condotta proponendo l'esplorazione di una mappa virtuale della regione Puglia. Questo soggetto si presta bene per una fruizione multi-livello ed ha permesso di sperimentare l'utilità degli scenari, che rappresentano uno degli assi dello schema riportato in Figura 2. Sullo specifico territorio regionale pugliese è possibile costruire più viste: mappe fisiche, politiche, di viabilità ecc.

Con le normali mappe tattili le alternative sono la realizzazione di più mappe tattili (una per ciascuno scenario) oppure la fusione di tutte le informazioni in un'unica mappa tattile, che risulterebbe però molto complessa e di difficile lettura aptica. L'uso delle didascalie in braille inoltre costringe l'utente a spostare l'attenzione dalla forma toccata alla posizione della scritta, causa questa di distrazione. Partendo da dati GIS relativi a province, fiumi e laghi è stata realizzata una mappa con molteplici scenari legati alla morfologia del territorio, che è stata poi completata da informazioni di tipo antropico (ubicazione delle città più importanti e delle principali tratte stradali di collegamento).

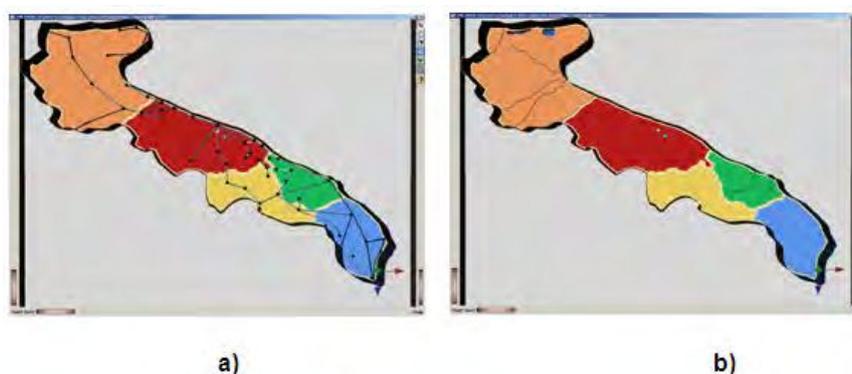
## 5.2 Funzionalità realizzate e sperimentate

Il punto centrale della sperimentazione è stato l'uso degli scenari: l'utente poteva dinamicamente scegliere quale scenario esplorare. Il primo livello contiene i confini regionali e provinciali, realizzati come strutture tridimensionali in rilievo che possono essere scavalcate per passare da una provincia ad un'altra. Tanto i confini che i territori provinciali sono stati definiti come oggetti attivi acustici. Attraverso un effetto aptico di vibrazione l'utente viene informato, scorrendo lungo il confine, di un cambiamento nell'identità dello stesso (per esempio quando, scorrendo lungo il confine regionale, cambiava la regione confinante).

In un secondo scenario è stata rappresentata la rete idrografica della regione (Figura 9.b). Fiumi e laghi sono stati descritti attraverso canyon ed avvallamenti rispettivamente: l'avatar dell'utente può quindi muoversi all'interno di essi per acquisirne il percorso o la forma e le dimensioni. Anche fiumi e laghi sono stati definiti come oggetti attivi acustici i cui messaggi vocali, attivati su richiesta, ne indicano il nome.

Un ulteriore scenario contiene la posizione delle maggiori città, rappresentate con dei prismi esagonali e definiti come oggetti attivi aptici associati ad un effetto di attrazione che cattura l'avatar quando entra nell'area d'influenza. Anche in questo caso un messaggio vocale, sempre attivato a richiesta, descrive il nome della città. In questo scenario è stata sperimentata una forma di assistenza attraverso la guida di una forza attrattiva: per facilitare il ritrovamento delle città, specialmente quelle isolate all'interno del territorio, l'utente poteva chiedere al sistema di essere condotto alla città più vicina.

L'ultimo scenario riproduce le principali vie di comunicazione stradale tra le città (Figura 9.a).



**Figura 9 - Due scenari del modello della Puglia:**  
a) lo scenario relativo alla rete stradale  
b) la rete idrografica

Per la rappresentazione delle strade sono state sperimentate due diverse tipologie: nel primo caso la strada è associata ad un segmento, tra due centri urbani, che esercita una forza attrattiva sull'avatar; nel secondo caso la rappresentazione è analoga a quella dei fiumi. Entrambe le rappresentazioni prevedono un messaggio vocale riguardante il nome della strada e delle città ai suoi estremi. L'utente può spostarsi tra gli scenari utilizzando dei tasti funzione abilitati sulla tastiera.

### **5.3 Sperimentazione e commenti**

Il modello è stato presentato a dieci non vedenti durante una sperimentazione controllata e ad altrettanti utenti durante un meeting dell'Unione Italiana Ciechi, che ha permesso di far esplorare la regione ad utenti italiani di varia provenienza che poco o nulla conoscevano del territorio pugliese.

L'esplorazione è partita dal modello contenente i confini regionali e provinciali. In seguito si è passati all'esplorazione della rete idrografica, delle città principali ed infine della rete stradale. È risultata particolarmente utile la funzione di aiuto nello scenario delle città: essa è stata utilizzata in modo intensivo specialmente durante le prime fasi dell'esplorazione. Durante la sperimentazione in laboratorio è stata proposta la prima tipologia di strada, mentre durante il meeting è stata proposta la seconda.

Durante la sperimentazione controllata, tutti gli utenti del sistema hanno acquisito una buona conoscenza della mappa politica, in quanto alla fine dell'esplorazione possedevano un'idea corretta della forma della regione e delle sue province, della loro collocazione spaziale e dei confini interprovinciali e interregionali. Un solo utente ha incontrato più difficoltà rispetto agli altri ma è riuscito comunque a capire la posizione dei laghi e dei fiumi sul territorio e a farsi un'idea del loro percorso. Tutti gli altri utenti hanno avuto l'occasione sia di trovare conferme alle informazioni che già possedevano sulla rete idrografica pugliese sia di apprenderne di nuove. Qualche difficoltà in più è stata incontrata nell'esplorazione dello scenario relativo alle città, anche se l'esito della prova può ritenersi più che soddisfacente: la maggiore difficoltà incontrata è stata quella di dover memorizzare la posizione delle varie località (molto spesso non note) sul territorio e di metterle in relazione alle altre al fine di riuscire a ritrovarle con maggiore facilità successivamente.

Prestazioni meno soddisfacenti si sono ottenute dalla maggioranza degli utenti nello scenario delle strade: quasi tutti hanno avuto grandi difficoltà a percepire l'attrazione al segmento con cui sono state rappresentate le strade e di conseguenza non sono riusciti a crearsi una rappresentazione organica della rete stradale proposta. Questo risultato ci ha indotto a fornire una rappresentazione alternativa della rete stradale che è stata proposta con buoni risultati in sperimentazione successiva, durante il meeting, in quanto il modello incavato ha permesso di seguire molto più agevolmente il tracciato.

## **6 Conclusioni e sviluppi futuri**

È stato descritto un sistema multi-modale per la fruizione di modelli tridimensionali digitali. L'aggiunta dell'interazione tattile alle più usuali restituzioni grafiche e sonore, rende possibile un'interazione più completa per i vedenti ed apre uno scenario di maggiore accessibilità per gli utenti con disabilità visive. Per questi ultimi, infatti, l'udito ed il tatto costituiscono i canali primari di contatto con la realtà circostante. L'acquisizione di dati a forte connotazione spaziale attraverso descrizioni verbali o sonore non è né naturale né semplice.

Gli artefatti fisici normalmente utilizzati per rappresentare tali informazioni verso il senso del tatto (mappe tattili, modelli in scala, ...) ancorché utili non presentano la stessa flessibilità e dinamicità che può invece essere realizzata nei modelli digitali. L'obiettivo è quindi quello di utilizzare tutti i gradi di libertà offerti dalle applicazioni di realtà virtuale per predisporre una esperienza della realtà che faciliti e renda più efficace il percorso conoscitivo che ciascun utente deve realizzare rispetto all'oggetto di interesse. Questo obiettivo richiede un'attenta messa a fuoco del contenuto informativo e formativo dell'esperienza ed un'appropriata valutazione e progettazione dei canali di comunicazione da utilizzare per trasmettere i contenuti. In questa progettazione è possibile introdurre la ridondanza ed il polimorfismo nella rappresentazione delle informazioni in modo da poter trasmettere gli stessi messaggi verso più sensi dell'utente o per rinforzarne la percezione o per adattarsi a particolari limitazioni/necessità dell'utente stesso.

Il sistema Omero si propone come una piattaforma multi-modale per permettere l'accesso a contenuti informativi, veicolati tramite opportuni modelli 2D/3D, anche ad utenti con disabilità visive, come risultato dell'interazione aptico/acustica oltre che visiva. L'interazione che il sistema propone con il mondo virtuale si avvale di varie funzionalità di supporto e potenziamento dell'interazione aptico/acustica di base: l'esplorazione interattiva, l'accesso progressivo ai contenuti, la possibilità di manipolare la scala della realtà virtuale, la guida dell'utente lungo percorsi di visita della scena, ... .

L'applicazione integrata ed efficiente di questi strumenti sulla scena virtuale richiede una sinergia tra la figura del modellatore della scena e quella dell'esperto del dominio applicativo cui appartiene la scena rappresentata. Sarà compito dell'esperto del dominio applicativo indicare quali sono i contenuti più importanti, gli oggetti più significativi ed i percorsi conoscitivi che devono essere racchiusi nel modello. Sarà compito poi di chi realizza il modello raccogliere queste indicazioni e dar loro la rappresentazione polimorfica all'interno del modello più adatta ad incontrare le esigenze della classe di utenti più ampia possibile.

Varie sperimentazioni condotte con utenti non vedenti, ipovedenti e vedenti hanno dimostrato come l'approccio del sistema sia percepito come naturale ed istintivo. A differenza di altre esperienze di sostituzione sensoriale nelle quali all'utente viene richiesto di sviluppare nuove abilità cognitive o capacità di interpretazione mentale, le interazioni proposte cercano di riferirsi ed utilizzare patrimoni concettuali già presenti e significativi.

L'accesso progressivo all'informazione ha prodotto tempi di apprendimento più rapidi rispetto all'esplorazione di un modello complesso presentato al suo massimo livello di dettaglio. L'uso integrato aptico/acustico ha favorito una trasmissione più compatta dell'informazione, rispetto al binomio <forma>-<descrizione in Braille>, aggiungendo parallelismo al processo di apprendimento.

Sviluppi futuri prevedono la maggiore standardizzazione delle rappresentazioni utilizzate anche per facilitare l'importazione e messa a punto dei modelli, la progettazione di strumenti per il rendering visivo meglio adatto alle esigenze degli utenti ipovedenti (peraltro molto più numerosi e spesso più trascurati dei non vedenti), la realizzazione di interfacce più versatili (riconoscimento automatico della voce, interfacce grafiche di migliore percezione ed uso, ...), strumenti di manipolazione e costruzione dei modelli, strumenti di configurazione e personalizzazione del sistema per adattarlo alle specifiche esigenze del singolo utente.

## 7 Bibliografia

- [1] K. SALISBURY, F. CONTI, AND F. BARBAGLI, “Haptic Rendering: Introductory Concepts”, *IEEE Computer Graphics and Applications*, Vol. 2, 2004, pp. 24-32.
- [2] R. D. JACOBSON, “Representing spatial information through multimodal interfaces”, *Proc. 6th International Conference on Information Visualisation*, 2002.
- [3] T. CHEN, P. YOUNG, D. ANDERSON, J. YU, AND S. NAGATA, Development of a stereoscopic haptic acoustic real-time computer (share), *Proc. SPIE 3295*, 1997
- [4] M. BERGAMASCO, A. FRISOLI, F. BARBAGLI, “Haptics technologies and cultural heritage applications”, *Proceedings of Computer Animation*, 2002. Volume , Issue , 2002, Page(s): 25-32
- [5] J. SOLIS; S. MARCHESCHI; O. PORTILLO, M. RASPOLLI, C.A. AIZZANO, M. BERGAMASCO, “The Haptic Desktop: a novel 2D multimodal device” Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on Volume , Issue , 20-22 Sept. 2004, Page(s): 521 – 526
- [6] J. A. GRADNER, “Access to blind students and professional to mainstream math and science”, *Proceedings of 8th International Conference*, ICCHP 2002, Linz, Austria, July 15-20, 2002.
- [7] H. PETRIE, V. JOHNSON, TH. STROTHOTTE, A. RAAB, S. FRITZ, R. MICHAEL, “MoBIC: designing a travel aid for blind and elderly people”, *The Journal of Navigation*, Vol. 49, No. 1, 1996, pp. 45-52.
- [8] P. PARENTE, G. BISHOP, “BATS: The Blind Audio Tactile Mapping System”, *Proc. of the ACM Southeast Regional Conference*, ACM, March 2003.
- [9] Y. Murai, H. Tatsumi, N. Nagai, and M. Miyakawa, “A Haptic Interface for an Indoor-Walk-Guide Simulator”, ICCHP 2006, Springer, Linz, Austria, July 12-14, 2006.
- [10] F.L. VAN SCOY, V. BAKER, C. GINGOLD, E. MARTINO, D. BURTON, “Mobility Training using Haptic interface: Initial Plans”, *Proc. of Fourth Annual PHANToM User Group*, Boston (US), October 1999.
- [11] C. MAGNUSSON, K. RASSMUS-GRON, “A Dynamic Haptic-Audio Traffic Environmet”, *Proc. of Eurohaptics 2004*, Munich, Germany, June 5-7 2004.
- [12] G. JANSSON, “Perceiving complex virtual scenes without visual guidance”. In *Touch in virtual environments*, M. L. Prentice Hall.
- [13] [www.sensable.com](http://www.sensable.com).
- [14] F. DE FELICE, F. RENNA, G. ATTOLICO AND A. DISTANTE, “A haptic system for the fruition of cultural heritage by blind people”, Building Virtual Rome – Immaginare Roma Antica, World Expo on Virtual Archaeology, Market of Trajan, Rome, Italy, 15 September – 15 November 2005.
- [15] F. DE FELICE, F. RENNA, G. ATTOLICO, A. DISTANTE, “A haptic/acoustic application to allow blind the access of spatial information”, in *Proceedings of WorldHaptics 2007*, Tsukuba, Japan, 22-24 March 2007.
- [16] [www.web3D.org](http://www.web3D.org)
- [17] [www.Collada.org](http://www.Collada.org)
- [18] [www-evasion.imag.fr/~Francois.Faure/doc/inventorMentor/sgi\\_html](http://www-evasion.imag.fr/~Francois.Faure/doc/inventorMentor/sgi_html)
- [19] [www.coin3d.org](http://www.coin3d.org)
- [20] <http://www.loquendo.com/en/index.htm>



## IA E MUSEI: L'ALLESTIMENTO AUTOMATICO

*Francesco Amigoni, Viola Schiaffonati*

Laboratorio di Intelligenza Artificiale e Robotica  
Dipartimento di Elettronica e Informazione  
Politecnico di Milano  
Via Ponzio, 34/5 - 20133 Milano, Italy

amigoni@elet.polimi.it, schiaffo@elet.polimi.it  
<http://www.airlab.elet.polimi.it/>

In questo contributo presentiamo una peculiare applicazione dell'intelligenza artificiale (IA) all'ambito museale: il sistema Minerva, sviluppato dal Laboratorio di Intelligenza Artificiale e Robotica del Politecnico di Milano. Minerva è un sistema di IA in grado di allestire automaticamente musei virtuali, sia per fornire supporto agli allestitori sia per arricchire le possibilità di visita degli utenti. In particolare, presentiamo in dettaglio due versioni di Minerva, una relativa a musei archeologici e una relativa a un nascente "museo del territorio", per dimostrarne le potenzialità e l'adattabilità a contesti diversi.

### 1 Introduzione

Negli ultimi anni l'ambito museale è stato oggetto di radicali cambiamenti riguardanti sia il ripensamento degli strumenti per la sua organizzazione e visita sia la nozione stessa di museo e le sue funzioni all'interno della società. Uno dei settori in cui questi cambiamenti risultano particolarmente evidenti è quello dell'applicazione delle nuove tecnologie dell'informazione e della comunicazione che possono portare a un mutamento delle modalità di allestimento e di visita di un museo [7]. I sistemi informatici offrono agli utenti la possibilità di interagire e muoversi in ambienti virtuali, ricevere informazioni complementari durante la visita di una mostra o di un museo e, infine, gestire in maniera flessibile e personalizzabile le imponenti quantità di dati relative agli oggetti culturali. Da ciò deriva il dissolversi della nozione tradizionale di museo come semplice contenitore e l'aprirsi di nuove possibilità di intendere il museo quale laboratorio di conoscenza.

Il *Progetto Minerva*, nato nel 1995 per iniziativa di Marco Somalvico presso il Laboratorio di Intelligenza Artificiale e Robotica del Politecnico di Milano, ha l'ambizioso obiettivo di estendere il dominio delle attività che possono essere gestite da un sistema informatico nel settore dei beni culturali. Più precisamente, il Progetto Minerva si occupa dello sviluppo di un sistema – chiamato semplicemente *Minerva* – in grado di fornire supporto all'attività creativa di chi allestisce musei e all'attività conoscitiva di chi visita musei, utilizzando tecniche di intelligenza artificiale (IA). In pratica, Minerva è in grado di allestire in modo automatico musei virtuali, a partire dalle collezioni di opere d'arte, di reperti archeologici o, più in generale, di oggetti che si vogliono visualizzare e dagli ambienti in cui tali oggetti devono essere esposti.

In questo contesto, l'allestimento viene considerato (forzando leggermente il significato comunemente attribuito ai termini) come unione di due processi distinti: la fase di *allestimento* vera e propria, ossia l'organizzazione delle opere o degli oggetti in gruppi culturalmente omogenei e ordinati, e la fase di *allocazione*, ossia la collocazione di questi gruppi in ambienti virtuali preservando il raggruppamento imposto durante il processo di

allestimento. Si noti come la fase di allestimento (nel senso definito sopra) è principalmente legata a criteri artistici, per esempio la creazione di collegamenti culturali fra le opere, mentre la fase di allocazione (nel senso definito sopra) è principalmente legata a criteri di tipo geometrico, per esempio la disposizione di opere mantenendo un'adeguata distanza fra le opere. Il sistema Minerva supporta l'utente nella fase di allestimento e svolge automaticamente quella di allocazione.

Un sistema come Minerva può essere impiegato per assolvere diverse funzioni. In particolare, nello sviluppo di Minerva sono state seguite due direzioni: la prima vede il sistema come supporto per l'allestimento di musei, utilizzato da chi si occupa di progettare e realizzare allestimenti, la seconda vede il sistema come strumento per creare musei virtuali personalizzati, utilizzato da una più ampia categoria di utenti a scopi informativi, didattici e ludici. Lungo la prima direzione, è stata realizzata una versione di Minerva per il supporto all'allestimento di opere provenienti dalla collezione del Museo Archeologico di Milano [5] (versioni precedenti dello stesso sistema sono descritte in [2][3][4][6]). Lungo la seconda direzione, sono state realizzate delle versioni di Minerva per la creazione di musei virtuali relativi all'Isola Comacina [1] e a un progetto INTERREG sull'arte dello stucco nel Parco dei Magistri Comacini. Due di queste versioni, Minerva Archeologico e Minerva L'arte dello stucco, saranno presentate in dettaglio nelle Sezioni 2 e 3.

È opportuno evidenziare che, anche nei casi in cui Minerva è utilizzato in funzione di supporto agli allestitori, non si tratta di voler sostituire professionisti umani con macchine opportunamente programmate, quanto piuttosto di voler fornire ai professionisti (e in generale a chi ne abbia bisogno) supporto per le loro attività con l'intento di realizzare una mutuamente proficua e sinergica interazione uomo-macchina. L'idea è di sfruttare ciò che ciascun soggetto è in grado di fare meglio; ad esempio, al sistema informatico possono essere delegati i compiti allestitivi di più basso livello (calcolo degli spazi e degli ingombri, verifica dell'adeguamento a norme, ...), mentre l'allestitore può occuparsi di attività maggiormente creative, come il costruire legami fra le opere, che siano in grado di raccontare una storia al visitatore [2].

Precisiamo inoltre che, nello sviluppo di Minerva, l'attenzione è stata principalmente posta al "nucleo" del sistema (descritto nella Sezione 1.1 di seguito), tralasciando l'interfaccia grafica. Quest'ultimo è uno degli aspetti da migliorare per promuovere Minerva da prototipo a sistema pronto per l'impiego in contesti reali.

### **1.1 L'architettura di Minerva**

Nonostante le differenti versioni, le modalità di utilizzo e l'architettura del sistema Minerva sono sostanzialmente uniche. L'utente può selezionare le opere d'arte, scegliere l'ambiente in cui desidera allestirle e impostare alcuni criteri che guidano il processo di allestimento automatico. Il sistema elabora tali informazioni ed è in grado di rendere fruibili i risultati ottenuti mediante un modello virtuale tridimensionale o bidimensionale del museo, navigabile dall'utente con tecniche di realtà virtuale o tramite *point-and-click*. È importante sottolineare che il sistema Minerva presenta un alto livello di automazione, reso possibile dall'utilizzo di tecniche di IA. Gli utenti non scelgono direttamente i singoli oggetti da inserire nei musei virtuali né li dispongono negli ambienti virtuali; al contrario, gli utenti selezionano alcuni temi e criteri che sono utilizzati poi dal sistema per recuperare (da un database) gli oggetti specifici e per disporli nel museo. Da questo punto di vista, Minerva occupa, come discusso nella successiva Sezione 1.2, una posizione unica e distinta rispetto ad altri sistemi di IA applicati alla realtà museale.

Dal punto di vista informatico, le differenti versioni del sistema Minerva sono state implementate come sistemi multiagente. Un *sistema multiagente* è composto da entità computazionali autonome, gli agenti, che interagiscono fra loro. In particolare, nel caso di Minerva, gli agenti collaborano al fine di allestire un museo virtuale. In generale, gli agenti di Minerva sfruttano regole che derivano dall'esperienza di allestitori umani e che sono state inserite nel sistema durante la sua realizzazione. Gli agenti sono dei programmi JAVA che sono stati implementati sfruttando l'ambiente JADE come middleware e l'ambiente JESS come motore inferenziale. Non riportiamo qui la descrizione completa dell'architettura e dei meccanismi interni di funzionamento di Minerva (il lettore interessato può fare riferimento ad alcuni lavori precedenti, come [1] e [5]), ma illustriamo sommariamente il funzionamento di due degli agenti che compongono l'architettura.

Data una collezione di opere e un ambiente nel quale disporle (entrambi scelti dall'utente), due agenti operano per trovare rispettivamente un allestimento e un'allocazione. Il primo agente, chiamato *agente allestitore*, determina, sulla base dei criteri selezionati dall'utente e in collaborazione con quest'ultimo, l'organizzazione concettuale delle opere. Più precisamente, l'agente allestitore estrae le opere di interesse dal database delle opere e, in base alle indicazioni fornite dall'utente, le suddivide in gruppi di opere omogenee e ordina questi gruppi. Il secondo agente, chiamato *agente allocatore*, individua la miglior disposizione dei gruppi di opere all'interno dell'ambiente selezionato, cercando di mantenere nella stessa area (per esempio, nella stessa stanza) le opere di uno stesso gruppo e di preservare l'ordine dei gruppi. Non è detto che i gruppi di opere prodotti dall'agente allestitore siano allocabili nell'ambiente dall'agente allocatore: in questo caso, i due agenti devono cooperare per trovare una soluzione (anche in collaborazione con l'utente). Per esempio, un gruppo di opere potrebbe essere ulteriormente suddiviso o due gruppi di opere potrebbero essere fusi.

## 1.2 Altri sistemi di IA per i musei

Il sistema Minerva si caratterizza in maniera del tutto originale nel panorama attuale dei sistemi informatici per la fruizione di contenuti culturali. Infatti, nonostante negli ultimi anni sia stata sottolineata l'importanza dell'utilizzo delle nuove tecnologie nel settore dei beni culturali, si tratta principalmente di tecnologie già sviluppate, e non progettate *ad hoc*, che vengono applicate per gestire gli aspetti più superficiali della fruizione. Nel caso di Minerva, invece, gli strumenti utilizzati sono stati appositamente progettati. Tali strumenti forniscono risultati assai significativi sia dal punto di vista tecnico sia dal punto di vista concettuale.

Nel primo caso, infatti, sono state sviluppate tecniche avanzate di IA per garantire la parziale automazione di processi allestitivi – compito non semplice, dato che si tratta di estendere i tradizionali ambiti di applicazione dell'IA.

Nel secondo caso, invece, la creazione e l'utilizzo di tali strumenti di supporto hanno aperto possibilità di personalizzazione e di costruzione di percorsi personalizzati all'interno dei musei virtuali, consentendo maggiore partecipazione e coinvolgimento degli utenti.

Due sono in particolare gli elementi di maggiore innovazione del sistema Minerva rispetto ad altri strumenti informatici per musei virtuali: la possibilità di creare allestimenti virtuali “tagliati” sugli specifici interessi e competenze dell'utente e l'alto grado di automazione del processo di allestimento. Questi due aspetti distinguono il sistema Minerva da altri che sono sostanzialmente sistemi di gestione di basi di dati che trattano collezioni e che consentono la creazione di musei virtuali scegliendo direttamente le opere e disponendole a mano.

Negli scorsi anni la virtualità è stata un tema centrale dell'applicazione di strumenti informatici ai musei. Oggigiorno tutti i più importanti musei offrono tour virtuali accessibili via web, per diffondere e pubblicizzare le loro collezioni. In questi sistemi, tuttavia, sono possibili solo minime personalizzazioni per l'utente che, generalmente, può visitare solo musei virtuali già allestiti e predeterminati (si vedano ad esempio [8][10][11][12]).

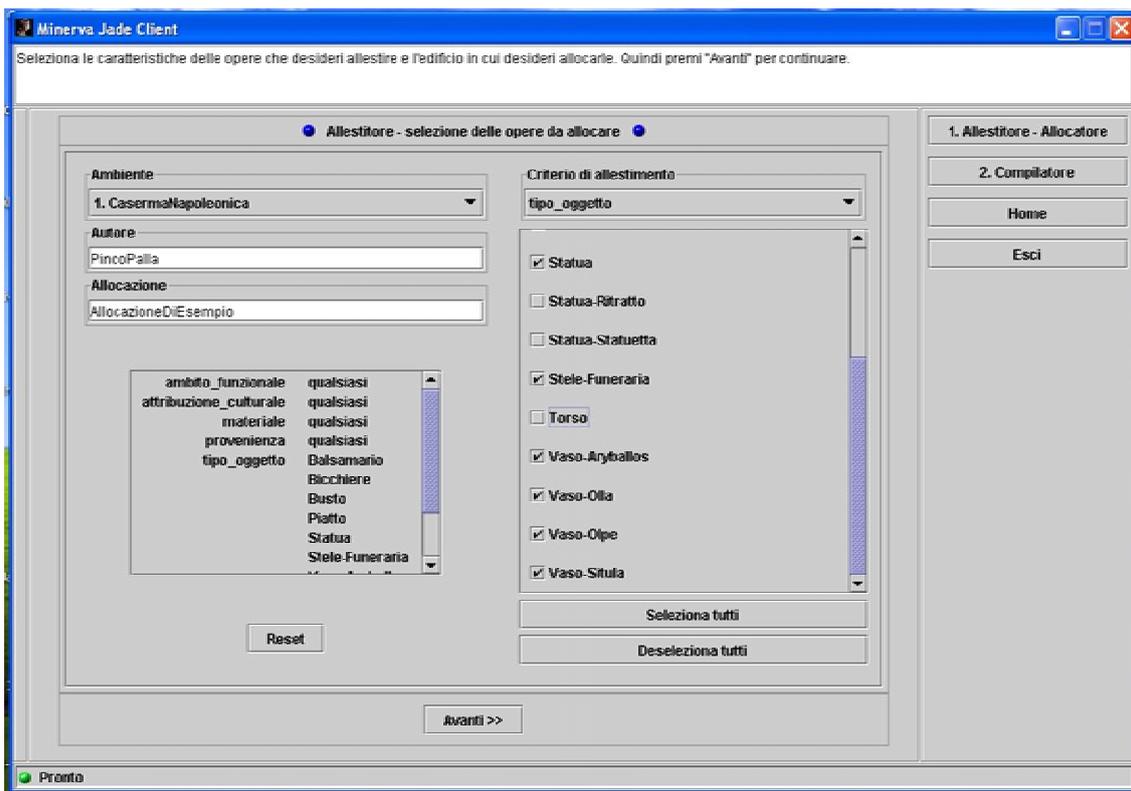
La virtualità può essere però adottata anche per obiettivi più specifici. Per esempio, quando le opere non possono essere esposte per problemi di spazio o perché particolarmente fragili. In questo scenario, i sistemi che sembrano essere più vicini al sistema Minerva sono quelli che offrono all'utente la possibilità di scegliere direttamente le opere d'arte da inserire nel museo virtuale (si vedano ad esempio [9][10][13][14]).

In conclusione, il sistema Minerva esibisce un livello di automazione decisamente maggiore di qualsiasi altro strumento in circolazione. L'utente non sceglie una ad una le opere da allestire e neppure le dispone manualmente nell'ambiente virtuale. Al contrario, l'utente sceglie dei criteri attraverso i quali selezionare quali opere saranno esposte nello spazio virtuale e secondo quale ordine. Il risultato che il sistema restituisce non è così preconosciuto dall'utente del sistema, ma è frutto dell'interazione "creativa" fra utente e sistema.

## **2 Minerva Archeologico**

Per descriverne il funzionamento, consideriamo in primo luogo la versione archeologica di Minerva, chiamata *Minerva Archeologico*, nata da una collaborazione con il Museo Archeologico di Milano [5] (versioni precedenti dello stesso sistema sono descritte in [2][3][4][6]). L'utente, che in questo caso è tipicamente un allestitore, può selezionare un insieme di opere d'arte fra quelle presenti nel database (in questo caso si tratta di reperti archeologici) e un ambiente in cui collocarle.

Nel caso mostrato in Figura 1, l'utente ha deciso di allestire alcune opere d'arte nelle stanze della Caserma Napoleonica, un edificio storico nel centro di Milano. Nella parte centrale della finestra mostrata in figura, si nota il meccanismo con il quale l'utente può scegliere quali opere d'arte allestire. L'utente può scegliere gli oggetti da allestire secondo i seguenti criteri: ambito funzionale, attribuzione culturale, materiale, provenienza e tipo. Si noti che l'utente non sceglie i singoli oggetti, ma solo insiemi di oggetti in accordo con i criteri selezionati. Nel caso specifico di Figura 1, l'utente ha scelto di vedere le statue, le steli funerarie e diversi tipi di vasi presenti nel database. Naturalmente, l'utente può tornare a questa prima fase dell'allestimento se non è soddisfatto del risultato prodotto nelle fasi seguenti.



**Figura 1 - Scelta delle opere**

La seconda fase dell'allestimento consiste nel suddividere in gruppi gli oggetti selezionati e nel determinare l'ordine in cui questi gruppi saranno esposti nell'ambiente selezionato. Ogni gruppo può essere, a sua volta, suddiviso in ulteriori sottogruppi.

Per esempio, in Figura 2 (vedi pagina seguente) l'utente ha suddiviso le opere in quattro gruppi ordinati secondo il tipo di oggetto: le steli funerarie, i balsamari, i bicchieri e i piatti. L'utente può ulteriormente intervenire rifinando i gruppi: il gruppo delle steli funerarie può essere ulteriormente suddiviso in quelle romane e in quelle etrusche. Inoltre l'utente può modificare l'ordine dei gruppi selezionando un gruppo e agendo sui tasti con le frecce per spostarlo verso l'alto o verso il basso.

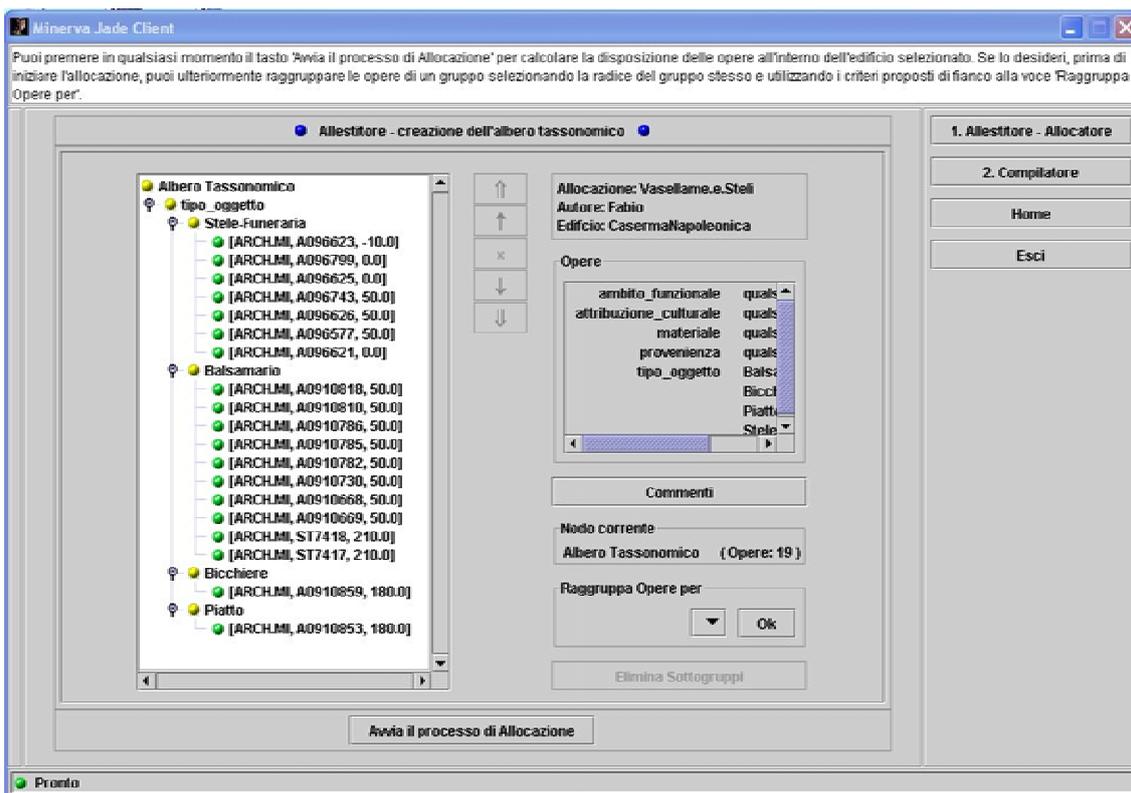


Figura 2 - Allestimento

Una volta che l'utente è soddisfatto con la fase di allestimento, ossia con il raggruppamento e l'ordine delle opere, può dare avvio alla fase di allocazione, in cui le opere vengono automaticamente collocate da Minerva su teche che sono poi posizionate all'interno delle stanze dell'ambiente selezionato. L'utente può agire su alcuni parametri del processo di allocazione che controllano la disposizione delle opere nell'ambiente (Figura 3).

Questi parametri comprendono l'occupazione massima e minima dell'area e del perimetro delle stanze, il numero massimo di opere che possono essere messe su una teca, la frazione di spazio che deve essere lasciata libera su una teca e l'ordine (orario oppure antiorario) secondo il quale le opere sono collocate nelle stanze. Durante il processo di allocazione Minerva cerca di allocare i gruppi ordinati delle opere nelle stanze dell'ambiente considerate in sequenza. In generale, il sistema cerca di allocare ogni gruppo in una stanza. Quando una stanza è troppo piccola per contenere un intero gruppo (stante i valori dei parametri di occupazione dello spazio specificati dall'utente), alcune delle opere del gruppo devono essere allocate nelle stanze successive. Quando la stanza, invece, è troppo grande e un gruppo non la riempie completamente, il sistema chiede all'utente se un altro gruppo può essere allocato nella stessa stanza oppure se la stanza può essere lasciata non del tutto occupata.

Nel caso riportato in Figura 4 il sistema chiede all'utente se delle statue romane debbano essere allocate nella stanza in cui si trovano già dei busti romani che non la riempiono completamente. In questo caso l'utente prende una decisione secondo l'effetto estetico che desidera ottenere con il suo allestimento. Si noti che questa scelta difficilmente può essere fatta dal sistema in automatico, dato che dipende da molti fattori come il tipo e il numero delle opere d'arte nei gruppi.

## ALLOCATORE

Clear

Nome allocazione:  Ambiente:

Opzioni avanzate

**IN GRANDE**

Teche:

Algoritmo:

Sequenzializzatore:

Aggiungi gruppi

Dividi gruppi

Max Occupazione Lineare:

Min Occupazione Lineare:

Max Occupazione Superficiale:

Min Occupazione Superficiale:

RCC:

**IN PICCOLO**

Opere:

Algoritmo:

Aggiungi teche:

Dividi teche:

Max Occupazione Lineare:

Min Occupazione Lineare:

Max Occupazione Superficiale:

Min Occupazione Superficiale:

Min RAT:

Max RAT:

Figura 3 - Allocazione

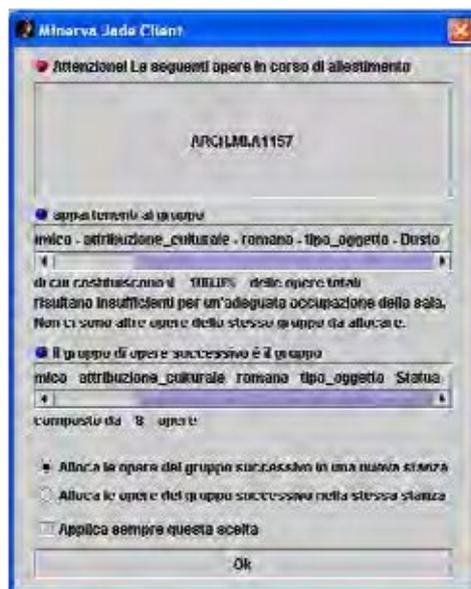


Figura 4 - Gestione di situazioni critiche nell'allocazione

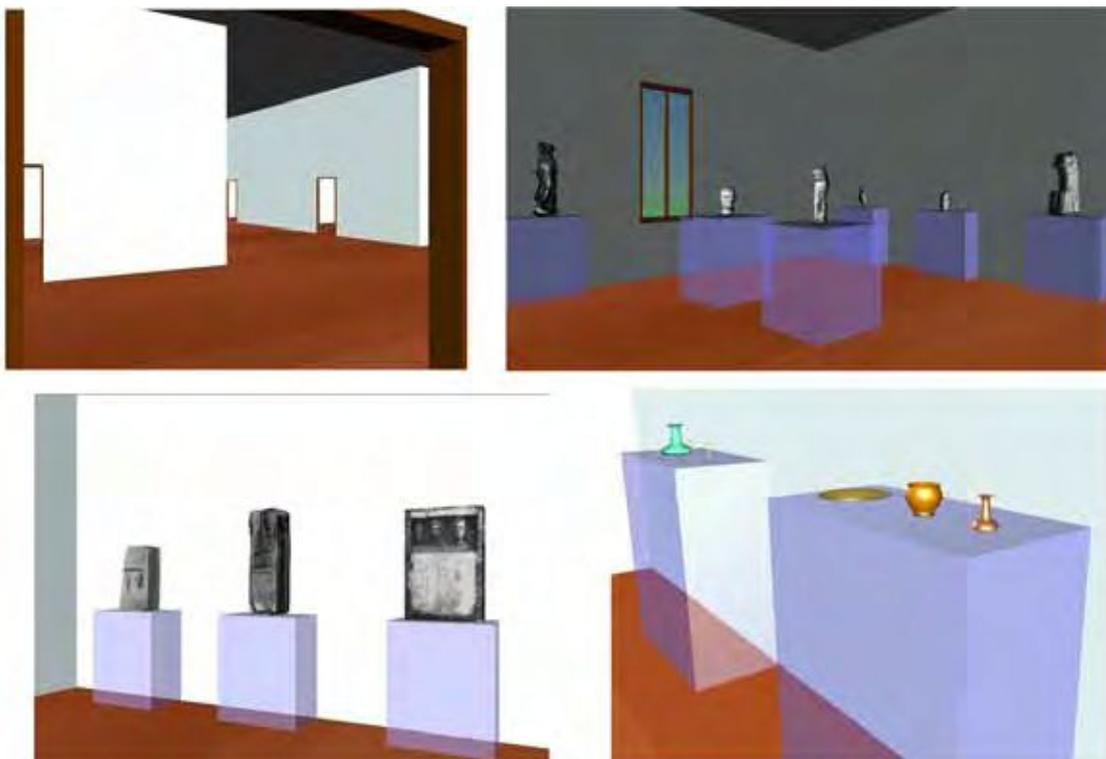


Figura 5 - Visita di un museo virtuale automaticamente allestito

PROGETTO MINERVA - SCHEDA DELL'OPERA

DETTAGLI	
CODICE OPERA	A096625
TITOLO	(sconosciuto)
MATERIALE	pietra calcarea
TIPO DI OGGETTO	Stele-Funeraria
PROVENIENZA	Europa - Italia
FORMA	(sconosciuto)
SOGGETTO	due fabbri
DATAZIONE	0.0
AUTENTICITA'	AUTENTICO
ATTIPRESERVAZIONE_CULTURALE	romana
AMBITO FUNZIONALE	funerario
TECNICA	(sconosciuto)
ALTEZZA cm	103.0
LARGHEZZA cm	88.0
PROFONDITA' cm	30.0

DESCRIZIONE	
<p>Stele con corpo quadrato coronato da un timpano dispiantato, decorato solo sul lato frontale; i fianchi sono ripianati a martellina media. Sopra una base con iscrizione, parte di corniciatura, compare una nicchia rettangolare, coronata da un fregio sopra un modanatura a gola; cui segue nel lato superiore una sottile fascia liscia leggermente inclinata verso l'interno; all'interno della nicchia vi sono due strati sovrapposti. Sopra la nicchia è un timpano fiancheggiato da due elementi architettonici di rilievo soltanto l'imposta; il campo frontale è coronato alla base da un fregio semplice, che prosegue lungo gli spioventi con una modanatura interna a</p>	

Figura 6 - Scheda di un'opera

Una volta che le fasi di allestimento e di allocazione sono completate, l'utente può visitare il museo così allestito in un ambiente VRML di realtà virtuale (Figura 5). Ogni opera d'arte è riprodotta virtualmente in accordo con il suo modello tridimensionale ottenuto dall'oggetto reale o da sue rappresentazioni (come per esempio fotografie). In questo museo virtuale, l'utente può navigare liberamente oppure seguire una lista di *viewpoints*. L'utente, inoltre, semplicemente cliccando sulle opere d'arte, può accedere a schede che riportano informazioni dettagliate sulle opere stesse (come ad esempio un modello VRML migliore e una descrizione testuale, Figura 6).

### **3 Minerva L'arte dello stucco**

Un'altra versione del sistema Minerva, denominata *Minerva L'arte dello stucco*, consente di fruire in maniera personalizzata dei risultati del progetto INTERREG IIIA "L'arte dello stucco nel Parco dei Magistri Comacini (Intelvesi, Campionesi e Ticinesi) delle Valli e dei Laghi: valorizzazione, conservazione e promozione". Più precisamente, il sistema Minerva L'arte dello stucco che presentiamo in questa sezione nasce dalla constatazione che un sistema di IA è il più adatto a gestire la complessità della comunicazione di un progetto interdisciplinare, che spazia dalla lettura iconografica all'interpretazione artistica, dalla documentazione storica alle analisi scientifiche, dalla metodologia della conservazione alla valorizzazione e al ruolo dei beni culturali nello sviluppo locale. Il sistema Minerva ha consentito, infatti, di realizzare un museo virtuale esplorabile in funzione degli interessi del visitatore, consentendo di comunicare con la massima ridondanza ("voglio rivedere quello che conosco già", ovvero "muovermi secondo la logica che è familiare alle mie competenze") o con il massimo di informazione ("voglio lasciarmi trasportare dentro gli aspetti che non conosco di un mondo molto vario").

Il sistema Minerva L'arte dello stucco permette all'utente di fruire al meglio delle informazioni e delle conoscenze ottenute dal progetto INTERREG attraverso la creazione di percorsi personalizzati all'interno di tali contenuti. In questo contesto, l'interazione dell'utente con il museo virtuale è supportata da una serie di tecnologie informatiche oggi ampiamente diffuse. Per esempio, i contenuti sono presentati attraverso pagine HTML arricchite dalla presenza di immagini e da collegamenti ipertestuali. L'utente può accedere al museo virtuale utilizzando un comune web browser (per esempio Internet Explorer o Mozilla Firefox), disponibile su tutti i sistemi informatici più diffusi. Attraverso il web browser l'utente può visitare il museo virtuale interagendo con le pagine del museo stesso e approfondendo i propri interessi in modo da costruire un percorso assolutamente personale fra gli oggetti risultanti dalla ricerca (siano essi descrizioni, fotografie, rilievi, collegamenti ecc.).

Oltre alla personalizzazione e alla flessibilità che caratterizzano il museo virtuale, già evidenziati con il sistema Minerva Archeologico presentato in precedenza, un risultato significativo legato specificamente al sistema Minerva L'arte dello stucco riguarda la possibilità di esporre nei musei personalizzati non solo le opere fisiche (chiese, cappelle, sculture, oggetti), ma anche i collegamenti contestualmente rilevanti fra questi e il territorio circostante. In questo senso si può senza dubbio affermare che Minerva L'arte dello stucco offre agli utenti la possibilità di costruire veri e propri racconti a partire da pochi oggetti, facendone emergere le connessioni con altri. Queste connessioni riflettono le diverse prospettive dalle quali possono essere analizzati i risultati del progetto.

Nel seguito descriviamo una tipica interazione con il sistema Minerva L'arte dello stucco da parte di un utente, con l'obiettivo di illustrare l'insieme dei possibili percorsi a sua disposizione.



**Figura 7 - Pagina iniziale con i casi studio**

Anche Minerva L'arte dello stucco è, come le altre versioni di Minerva, un prototipo che mostra alcune limitazioni, peraltro facilmente superabili. Per esempio, l'aspetto grafico dell'interfaccia è migliorabile per aumentarne l'usabilità; inoltre, il numero di oggetti (chiese, stucchi, ...) che, al momento, è di circa 200 può essere ulteriormente accresciuto.

Nella pagina iniziale (Figura 7) l'utente può scegliere da quale caso studio partire per la creazione del suo museo virtuale semplicemente selezionando l'immagine che rappresenta il caso studio di suo interesse (si tratta dei casi studio considerati nel progetto INTERREG: Santuario della Madonna dei Ghirli a Campione d'Italia, Madonna del Restello a Castiglione Intelvi, S. Lorenzo a Laino, S. Maria a Scaria Intelvi). Sulla medesima pagina, l'utente può leggere una breve descrizione dei diversi percorsi di ricerca che fungono da linee guida per l'esplorazione del museo. I percorsi di ricerca previsti sono: "Percorsi di lettura storico-artistica", "Il contesto storico, sociale ed economico", "Materiali e tecniche", "Rilievo e restauro" e "Galleria fotografica e informazioni utili", che rappresentano le differenti anime del progetto INTERREG.

Dopo aver selezionato il caso studio di proprio interesse, nella nuova pagina (Figura 8) l'utente ha accesso a maggiori informazioni relative alla chiesa e ai contenuti dei percorsi di ricerca ad essa relativi. Si noti che i contenuti informativi sono differenziati in base alla chiesa scelta.

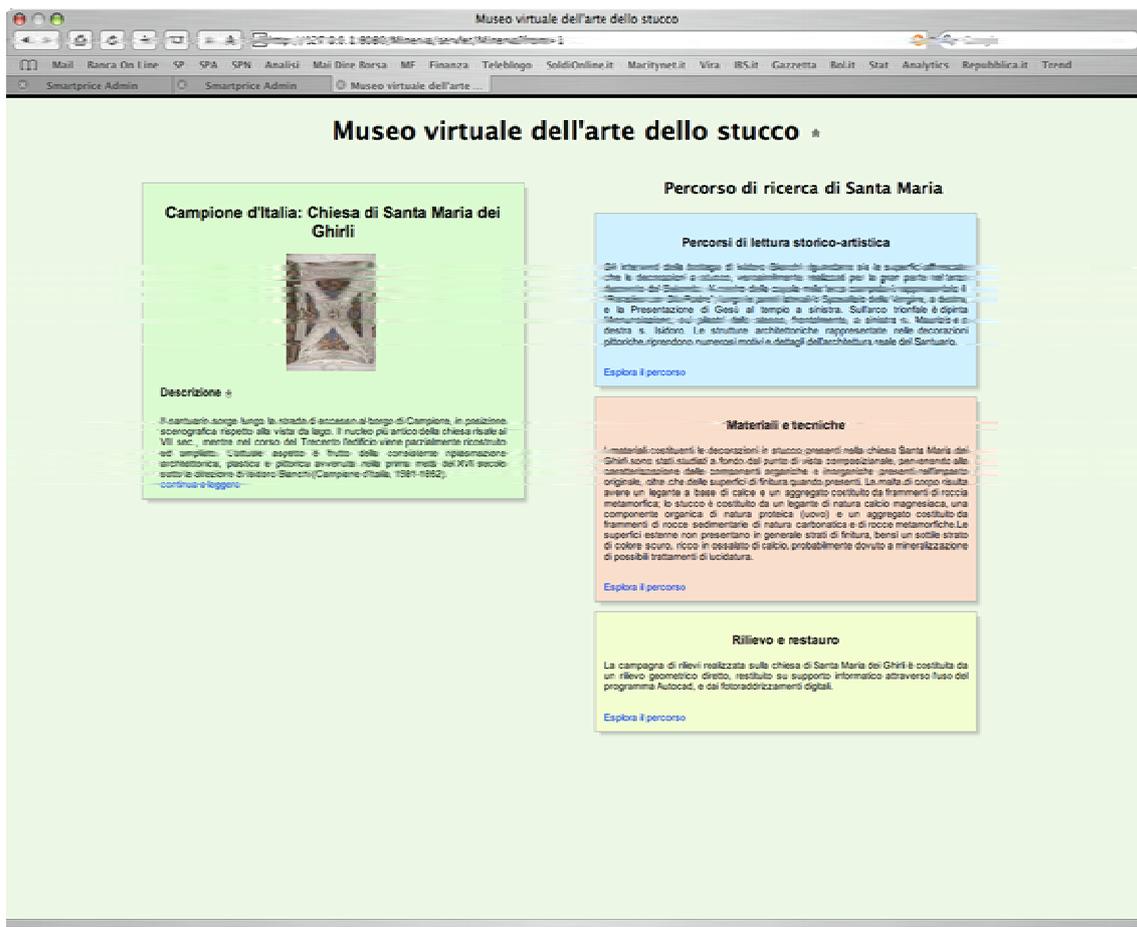


Figura 8 - Pagina relativa al caso studio scelto

L'utente è libero di selezionare quale percorso di ricerca approfondire. Nella pagina scelta è presentata una descrizione della chiesa riferita nello specifico al percorso di ricerca selezionato (nel caso di Figura 9, il percorso scelto è "Materiali e tecniche" per il caso studio della Madonna del Restello). Inoltre, in accordo al caso studio e al percorso di ricerca prescelto, compaiono una serie di approfondimenti ad esso relativi.

Gli approfondimenti sono scelti ed ideati sulla base dei risultati che ciascun partner ha conseguito all'interno del progetto INTERREG: ogni approfondimento vuole indirizzare l'utente nella scelta dei criteri e dei siti sulla base dei quali intende creare, con il supporto del sistema, il suo museo virtuale. Per esempio, l'utente potrebbe decidere di vedere i soggetti "angeli" presenti in tutte e quattro le chiese.

Va sottolineato che il sistema estende la ricerca degli oggetti da presentare anche al di là dei casi studio trattati in questo progetto, per comprendere altri oggetti inseriti nel database che hanno connessioni con gli oggetti dei casi studio: in questo modo il museo virtuale personalizzato prende forma al di là dei confini iniziali.

Museo virtuale dell'arte dello stucco

http://127.0.0.1:8080/Minerva/servizi/Minerva?theme=8&id\_chiesa=2

Museo virtuale dell'arte

## Museo virtuale dell'arte dello stucco

**Castiglione d'Intelvi: Oratorio della Madonna del Restello**



**Materiali e tecniche**

Se gli studi storico-artistici sul Restello indicano che il periodo di costruzione e decorazione della chiesa è contenuto in un ventennio ed evidenziano una campagna isolata e decorativa unitaria, le analisi hanno invece rivelato un'eterogeneità compositiva delle malte di corpo, dello stucco vero e proprio e delle finiture nei vari punti di campionamento. Inoltre è stata rinvenuta cartapesta all'interno delle statue e si sono rivelate tracce di piombo sulle superfici di finitura di alcuni stucchi da collegare probabilmente all'uso di attrezzi a stampa.

**I materiali utilizzati negli stucchi della Madonna del Restello**

**Legione trasversale lucida dello stucco dofile del secondo angelo**

Le malte di corpo mostrano composizioni eterogenee, con leganti a base di calcare calcico, di calcare calcico debolmente magnesioso e di gesso, spesso caratterizzati dalla presenza di grumi, di aggregati, a bassa classazione ed alto addensamento, sono prevalentemente costituiti da frammenti di rocce... [continua a leggere](#)

**Come è fatto uno stucco**

**Capitello in stucco ideato a sinistra dell'ingresso**

Con il termine "stucco" si indica un particolare tipo di decorazione in rilievo, realizzata in materiale plastico bianco, eventualmente tinto con pigmenti, che induce ad affaria. Lo stucco è generalmente costituito da un'armatura e da più strati di malta sovrapposti, a granulometria degradata... [continua a leggere](#)

**Descrizione dei materiali impiegati per la realizzazione degli stucchi**

**Pietruzzi Natta e Come scegliere stoffe**

In generale si può dire che uno stucco è una miscela di almeno due componenti: legante e aggregato. Gli stucchi hanno una varietà compositiva notevole e gli impasti possono essere divisi in tre gruppi principali: a base di gesso, con eventuali additivi (leganti organici) e... [continua a leggere](#)

**La ricetta dello stucco in letteratura**

**Pubblicazione di un saggio sulla fattura**

La trattatistica architettonica descrive con dovizia di particolari le malte e, spesso, anche gli stucchi da utilizzare nelle decorazioni. Il punto di riferimento è sempre Vitruvio. Le ricette parlano sempre dell'indicazione che i manufatti in stucco debbano essere preparati mediante una... [continua a leggere](#)

**Le analisi di laboratorio**

**Presenza di tracce di piombo nelle analisi di finitura**

I campioni prelevati mediante l'impiego di bisturi e piccoli scalpelli seguendo le indicazioni fornite dalla raccomandazione UNI Normale 380, sono stati complessivamente 29. Su di essi sono stati condotti analisi mineralogico-petrografiche mediante indagini diffrattometriche a raggi X (XRD) su... [continua a leggere](#)

**Scopri gli oggetti del museo virtuale**

sulla base degli approfondimenti di questo caso-studio  
Minerva ti consente di cercare

- Stucco a base di calcare
- Stucco a base di gesso
- Stucco a base di calcare e gesso
- Struttura di sostegno
- Polveri di marmo
- Elementi di tracciatura
- Finiture

[Torna indietro](#)

**a partire da questi luoghi**

- Santa Maria dei Ghis
- Madonna del Restello
- San Lorenzo
- Santa Maria di Scaria
- Tutti i casi studio

[Trova oggetti](#)

Figura 9 - Pagina relativa al percorso di ricerca scelto



**Figura 10 - Museo virtuale creato dall'utente**

Il museo virtuale che è stato creato viene presentato in una nuova pagina (Figura 10), che offre una serie di oggetti che soddisfano i criteri di ricerca selezionati: una parte degli oggetti è quella presente nei quattro casi studio di partenza, mentre un'altra parte comprende gli oggetti ad essa correlati (per varie motivazioni) e che sono esterni ai casi studio. Qui si rivela la funzione sussidiaria del sistema informatico, che suggerisce all'utente nuovi accostamenti e percorsi, che erano impliciti nella combinazione di criteri selezionata, ma dei quali l'utente stesso non era forse consapevole.

A questo punto l'utente può accedere alle pagine/schede di ogni singolo oggetto (Figura 11): ricordiamo che gli "oggetti" trattati da Minerva sono sia oggetti in senso fisico che "pacchetti" di informazioni e di relazioni. La pagina di ciascuno di essi può infatti contenere commenti e precisazioni scaturite dalle letture che le diverse aree disciplinari (quella storico-artistica, l'indagine scientifica ecc.) hanno riservato al medesimo oggetto: in questo modo proprio gli oggetti del museo virtuale permettono il concretizzarsi dell'obiettivo di approccio multidisciplinare al tema della ricerca, lo stucco, in quanto offrono in ogni scheda una lettura "trasversale" del medesimo soggetto. A questo punto l'utente può tornare ai percorsi di ricerca principali del caso studio di cui l'oggetto è testimonianza significativa, oppure scoprire altri oggetti correlati a quello di partenza, e da qui continuare a seguire il proprio percorso e a costruire il proprio racconto all'interno del museo virtuale.

**Museo virtuale dell'arte dello stucco** #

**Lunetta in controfacciata (navata sinistra)**



Nel campione analizzato (campione 9) sono stati riscontrati ben 9 strati di finitura di cui i primi 6 con calce magnesiaca e quelli superiori legati alla doratura.

**Tematiche correlate**

Lo strato di finitura in oro si presenta in precarie condizioni di conservazione: si riscontra un distacco degli ultimi strati di finitura, con numerose mancanze, e un consistente deposito di ossidazione.

**Oggetti Correlati**

La malta riscontrata in questo campione ha un legante calcio magnesiacco e un aggregato carbonatico con frammenti metamorfici. Lo stucco ha un legante calcio magnesiacco con idromagnesite e un aggregato carbonatico e silicatico. Sopra sono sovrapposti 9 strati di finitura: i primi sei strati sono a base di calce magnesiacca, poi vi è uno strato con giallo di cromo e ossidi di ferro, uno organico, uno di oro e infine si trova un deposito di particelle varie.

**Oggetti Correlati:**

**Percorsi di lettura storico-artistica** [Vai al tema](#)

La decorazione è stata realizzata dalla scuola del Colombo

**Oggetti Correlati**

**Figura 11 - Scheda di un oggetto**

## 4 Conclusioni

In questo contributo abbiamo presentato, attraverso l'illustrazione di due versioni, il sistema Minerva sviluppato dal Laboratorio di Intelligenza Artificiale e Robotica del Politecnico di Milano. Una delle maggiori differenze fra le diverse versioni di Minerva riguarda l'estensione del concetto di oggetto da esporre nel museo virtuale. Se nelle prime versioni di Minerva questo era inteso in senso tradizionale, quale opera d'arte o reperto archeologico, a partire dalla versione relativa all'Isola Comacina [1], gli oggetti da esporre non sono più solo i modelli virtuali delle opere con le loro informazioni specifiche (foto, descrizioni, ...), ma anche informazioni più generali sulle connessioni tra gli oggetti e il territorio circostante. In questo modo il sistema fornisce agli utenti la possibilità di creare racconti differenti a partire da pochi oggetti e di scoprire le relazioni fra questi e altri oggetti (come edifici o siti archeologici) situati nella stessa area. L'importanza di un approccio di questo tipo, oltre che nell'alto grado di personalizzazione che fornisce all'utente, risiede nella possibilità di rendere interessanti, anche per un utente comune, anche i pezzi meno importanti (per esempio frammenti) che possono comporre la collezione di un museo. Alla base, infatti, vi è l'idea che i musei virtuali creati utilizzando il sistema proposto possano essere dei ponti tra gli oggetti esposti in un museo fisico e la conoscenza che questi portano con sé (come quella, per esempio, relativa alle tecniche di scavo con cui sono stati rinvenuti). In questa prospettiva ogni oggetto, anche quello meno pregevole dal punto di vista artistico, diventa in grado di raccontare una storia connettendosi a informazioni più ampie. Nel realizzare il sistema Minerva L'arte dello stucco si è seguita questa stessa direzione in modo che gli oggetti siano in grado di raccontare storie e di costruire legami con altri oggetti e con il territorio circostante.

Tale linea di ricerca si presenta molto promettente per quanto riguarda eventuali sviluppi futuri. La virtualità e l'utilizzo delle nuove tecnologie nel settore dei beni culturali, infatti, sono ben lungi dall'aver trovato piena applicazione e dall'aver esaurito la loro spinta innovativa. Molte nuove applicazioni possono essere realizzate per la costruzione di "musei del territorio" a partire dagli spot che si vogliono valorizzare. Più in particolare, nel futuro intendiamo migliorare alcuni aspetti tecnici di Minerva, per esempio per permetterne la

completa fruizione via web, e applicare Minerva all'allestimento di altri musei per testare ulteriormente le sue potenzialità e la sua applicabilità ad ambiti diversi.

## 5 Bibliografia

- [1] AMIGONI, F., DELLA TORRE, S., SCHIAFFONATI, V. (2006), "Yet Another version of Minerva: The Isola Comacina Virtual Museum", *Atti di "ECAI2006 (Seventeen European Conference on Artificial Intelligence) Workshop on Intelligent Technologies for Cultural Heritage Exploitation"*, Riva del Garda, 28 agosto 2006, p. 1-5.
- [2] AMIGONI, F., SCHIAFFONATI, V. (2003), "The Minerva Multiagent System for Supporting Creativity in Museums Organization", *Atti di "IJCAI2003 (Eighteenth International Joint Conference on Artificial Intelligence) Workshop on Creative Systems: Approaches to Creativity in AI and Cognitive Science"*, Acapulco, Messico, 9-10 agosto 2003, p. 65-74.
- [3] AMIGONI, F., SCHIAFFONATI, V. (2004), "The Minerva Multiagent System for Museums Organization", *Intelligenza Artificiale*, 1(1), febbraio 2004, p. 39-43.
- [4] AMIGONI, F., SCHIAFFONATI, V. (2004), "L'intelligenza artificiale per l'allestimento automatico dei musei", *Kermes - La rivista del restauro*, XVII, luglio-settembre 2004, p. 68.
- [5] AMIGONI, F., SCHIAFFONATI, V. (2005). "A New Version of Minerva for Organizing Archeological Museums", *Atti di "AI\*IA (Associazione Italiana per l'Intelligenza Artificiale), Workshop on Cultural Heritage"*, Milano, 20 settembre 2005.
- [6] AMIGONI, F., SCHIAFFONATI, V., SOMALVICO, M. (2001), "Minerva: An Artificial Intelligent System for Composition of Museums", *Atti di "International Cultural Heritage Informatics Meeting (ICHIM01)"*, Volume 2 (Short Papers/Posters and Demos), Milano, 3-7 settembre 2001, p. 389-398.
- [7] ANTINUCCI, F. (2004), *Comunicare nel museo*, Laterza, 2004.
- [8] British Museum, Londra, Regno Unito. <http://www.thebritishmuseum.ac.uk>. Ultimo accesso giugno 2007.
- [9] Cardiac Virtual Museum. <http://arrhythmia.hofstra.edu/vrml/museumn/museumn.html>. Ultimo accesso giugno 2007.
- [10] Centre Pompidou, Parigi, Francia. <http://www.centrepompidou.fr>. Ultimo accesso giugno 2007.
- [11] Museo National Del Prado, Madrid, Spagna. <http://museoprado.mcu.es>. Ultimo accesso giugno 2007.
- [12] The National Gallery, Londra, Regno Unito. <http://www.nationalgallery.org.uk>. Ultimo accesso giugno 2007.
- [13] Virtual Museum of Canada. <http://www.virtualmuseum.ca>. Ultimo accesso giugno 2007.
- [14] WALCZAK, K., CELLARY, W., WHITE, M. (2006). "Virtual Museum Exhibitions". *IEEE Computer*, 39(3), p. 93-95.



αρωγή

## UN LABORATORIO DISTRIBUITO SU BASE GEOGRAFICA PER LA RICOMPOSIZIONE VIRTUALE ASSISTITA DI FRAMMENTI

*Nicola Mosca, Giovanna Carlomagno, Floriana Renna, Giovanni Attolico, Arcangelo Distante*

Istituto di Studi sui Sistemi Intelligenti per l'Automazione (ISSIA)  
Consiglio Nazionale delle Ricerche (CNR)  
Via Amendola, 122/D – 70126 – Bari  
*attolico@ba.issia.cnr.it*

### 1 Introduzione

La ricomposizione della vela del S. Matteo, affrescata da Cimabue nella Basilica Superiore di S. Francesco in Assisi e frammentata in più di 140.000 pezzi a seguito del terremoto del 1997, costituisce una sfida di notevole complessità.

L'elevato numero di frammenti (nettamente superiore a quello solitamente gestito con la modalità tradizionale), la loro grande difformità in dimensioni e forma, la tecnica utilizzata da Cimabue che rende la pellicola pittorica potenzialmente vulnerabile alla prolungata, anche se attenta, manipolazione che la ricomposizione tradizionale implica hanno suggerito ai responsabili del restauro della basilica di esplorare l'uso di strumenti informatici nella ricomposizione.

A complicare ulteriormente il problema, differenziandolo anche rispetto alla ricomposizione di un puzzle che istintivamente richiama alla mente, c'è il fatto che i frammenti raccolti non coprono la superficie intera della vela, contengono probabilmente frammenti di un affresco adiacente sulla volta della basilica e presentano dei contorni danneggiati la cui corrispondenza esatta non è garantita.

L'Istituto Centrale del Restauro ha collocato i frammenti in più di 1.000 contenitori, incastonandoli in un supporto, il *foam*, in grado di offrire loro stabilità e produrre un contrasto di colore abbastanza significativo tra fondo e frammenti. Di ciascun contenitore è stata poi acquisita un'immagine, utilizzando opportuni riferimenti per valutare le variazioni di colore, le disomogeneità di luminosità e le eventuali distorsioni geometriche introdotte nel processo d'acquisizione. Questo materiale digitale ha costituito la base da cui si è partiti per sviluppare uno strumento che consentisse la ricomposizione virtuale dei frammenti.

#### 1.1 Analisi del problema

L'analisi dei pochissimi tentativi fatti in passato per sviluppare strumenti per la ricomposizione virtuale di frammenti ha evidenziato due approcci fondamentali.

Nel primo ciascun frammento viene esaminato e classificato dagli operatori che gli assegnano un certo numero di caratteristiche visive scegliendole all'interno di un vocabolario predefinito (numero e tipologia di colori, attribuzione a specifiche parti dell'affresco, caratteristiche di tessitura, presenza/assenza di disegni interni, ...): queste caratteristiche vengono codificate come annotazioni testuali del frammento stesso. Tale classificazione permette, durante la ricomposizione, di selezionare rapidamente con l'aiuto degli strumenti informatici i frammenti d'interesse, in base alle caratteristiche visive desiderate [1].

Questo approccio ha il vantaggio di una grande semplicità concettuale e di realizzazione e può far uso di strumenti informatici già commercialmente disponibili. Ha anche l'apparente vantaggio di basarsi esattamente sulle caratteristiche che gli operatori normalmente utilizzano per il proprio lavoro. Esso però presenta alcune serie limitazioni. Richiede che gli operatori esaminino e classifichino uno per uno tutti i frammenti, con gli inevitabili problemi derivanti sia dalla disomogeneità della classificazione tra i diversi operatori (che dovrebbero valutare in modo comune caratteristiche visive spesso fortemente soggettive) sia dall'incostanza delle valutazioni dello stesso operatore nel tempo (stanchezza, cali di attenzione, ...). Inoltre la classificazione dovrebbe essere eseguita una sola volta, valutando in modo definitivo (e quindi avendo ben chiare dall'inizio) tutte le caratteristiche utili: l'aggiunta di nuove caratteristiche in corso d'opera imporrebbe infatti il riesame dell'intera collezione di frammenti con intuibili costi economici e di tempo.

Il secondo approccio è basato su strumenti per la ricomposizione automatica dei frammenti. Attraverso opportune rappresentazioni delle immagini dei frammenti e di un'immagine dell'affresco prima della frammentazione vengono individuate, per ciascun frammento, le posizioni nelle quali esso può essere collocato con maggiore probabilità. A questo punto l'operatore può valutare le alternative e confermare quelle che appaiono convincenti [2], [3].

In questo approccio è naturalmente indispensabile la disponibilità di un'immagine dell'affresco prima della frammentazione che deve essere di qualità sufficiente a rendere significativo il confronto con i frammenti. Inoltre l'apporto dell'operatore si limita all'esame ed alla convalida delle proposte del sistema. Non viene offerto alcun supporto alla ricerca di ulteriori alternative nei casi in cui le proposte fatte non contengano le posizioni corrette: per poter ricollocare frammenti che non siano stati riconosciuti dal sistema bisogna procedere in modo completamente libero e sostanzialmente in modalità tradizionale senza alcun supporto del sistema informatico.

Dall'analisi di diversi problemi di ricomposizione, anche differenti dallo specifico caso della vela del S. Matteo, è emerso che queste due logiche di lavoro (ricerca di corrispondenza tra frammento ed immagine dell'affresco ove possibile e ricerca di frammenti che possano essere accostati usando tutte le informazioni disponibili) possono e devono essere impiegate in modo congiunto e sinergico. Inoltre è apparso evidente che l'operatore porta nel processo di ricomposizione un contributo di cultura e di esperienza che consente valutazioni visive ulteriori rispetto a quanto può essere analizzato dagli algoritmi di analisi automatica delle immagini. Il sistema di supporto deve quindi integrarsi nella ricomposizione lasciando ampi spazi decisionali e flessibilità di scelta all'operatore stesso, responsabile ultimo del processo, il quale deve potersi avvalere degli aiuti informatici per aumentare l'efficienza del proprio lavoro senza esserne inutilmente condizionato in termini procedurali. È necessario che si realizzi una sinergia tra operatore e strumenti informatici che consenta ad entrambi di esprimere pienamente il proprio potenziale nel procedimento di ricomposizione [4].

## 1.2 La soluzione proposta

L'approccio seguito è consistito quindi nello sviluppare un sistema per la *ricomposizione virtuale assistita* di frammenti. *Virtuale* perché lavora sulle immagini dei frammenti piuttosto che sugli oggetti fisici che rimangono custoditi nel cantiere dei frammenti e la cui manipolazione è quindi limitata a quanto necessario per replicare nella realtà il riassetto messo a punto sulle immagini all'interno del sistema. *Assistita* (e non automatica) perché durante l'intero processo l'operatore non è sostituito dal sistema ma assistito nel suo lavoro dalle funzionalità offerte che si propongono di migliorarne l'efficienza e l'efficacia [5]. L'operatore rimane al centro dell'intero processo, avendo pieno *controllo* e *responsabilità* del modo in cui i diversi strumenti offerti vengono applicati.

Questa impostazione ha richiesto come primo passo la realizzazione di una stazione di lavoro (Figura 1) che replicasse in digitale la modalità tradizionale di ricomposizione. L'applicazione, realizzata in ambiente Windows, è stata concepita in modo tale che l'operatore, non necessariamente esperto di strumenti informatici, non sia costretto a modificare il proprio approccio all'attività di ricomposizione e possa replicare in video quanto è abituato a compiere in cantiere.



**Figura 1 - La stazione client utilizzata dai restauratori. Sul monitor di sinistra è visibile lo spazio di lavoro sul quale (se disponibile) viene visualizzata a risoluzione massima la parte dell'immagine di riferimento dell'affresco su cui si sta lavorando. Questo schermo costituisce la controparte virtuale del tavolo sul quale in cantiere viene collocata la foto a dimensioni reali dell'affresco per ricercare la collocazione dei frammenti. Sul monitor centrale un'altra finestra mostra l'immagine dell'affresco, scalata per poter essere visibile completamente, che viene utilizzata per cambiare l'area di lavoro. Altre finestre (monitor a destra) vengono dedicate ad ospitare contenitori virtuali, equivalenti ai cassetti utilizzati in cantiere per organizzare i frammenti in gruppi logicamente omogenei**

La configurazione multimonitor proposta consente di presentare e gestire in modo funzionale lo spazio di lavoro che contiene il tavolo virtuale di ricomposizione, la miniatura dell'affresco e i contenitori virtuali dei frammenti. I contenitori virtuali costituiscono l'astrazione dei cassetti fisici usati dai restauratori per catalogare i frammenti in cantiere ma, contrariamente a quanto avviene nella realtà, un frammento può essere archiviato contemporaneamente in due o più contenitori. All'operatore viene offerta la possibilità di lavorare su di un tavolo virtuale sul quale, se disponibile, può essere mostrata l'immagine a piena risoluzione dell'area di interesse dell'affresco prima della frammentazione. L'area di lavoro può essere cambiata spostandosi con il mouse su una miniatura dell'immagine dell'affresco, riprodotta in un'altra finestra. Su questo tavolo virtuale l'operatore può trascinare le immagini dei frammenti e muoverle, attraverso rotazioni e traslazioni, per collocarle nella posizione corretta. A tale scopo, attualmente viene utilizzato un mouse con 6 gradi di libertà, che può permettere al sistema di evolvere per la gestione di frammenti solidi nello spazio tridimensionale.

La selezione dei frammenti può essere svolta in diversi modi. Innanzitutto attraverso la selezione diretta, come si farebbe in cantiere, esaminando i frammenti e scegliendoli in modo appropriato. Il sistema però offre strumenti più efficienti: l'operatore può scegliere, posizionandosi in modo libero, uno o più dettagli dell'immagine dell'affresco (se disponibile) da usare come esempi da fornire ad un motore di ricerca su immagini che seleziona i frammenti più simili per caratteristiche visive e quindi maggiormente candidati ad essere collocati, il tutto in tempi assolutamente non paragonabili con quelli necessari ad esaminare la collezione di oggetti fisici in cantiere. Può utilizzare come esempi altri frammenti (perché già collocati in una certa area o perché l'immagine di riferimento non è disponibile) per far crescere delle aree già ricomposte. Può anche combinare le due modalità, raffinando delle richieste iniziate a partire da particolari dell'affresco attraverso frammenti ricollocati nell'area di lavoro.

Questa versatilità del motore di ricerca sulle immagini fa sì che l'operatore possa utilizzare il sistema non solo per la ricomposizione in senso stretto ma anche per suddividere i frammenti (eventualmente in modo gerarchico attraverso partizioni successive), altra attività comune nella pratica di cantiere. Avere a disposizione gruppi omogenei per contenuto pittorico riduce il numero di oggetti da considerare durante le ricerche. Una simile classificazione dei frammenti, logicamente simile a quella proposta dagli approcci che prevedono la classificazione manuale di ciascun oggetto, offre in realtà rispetto a quest'ultima notevoli vantaggi. In qualunque momento, infatti, se viene identificata una nuova caratteristica visiva di interesse rispetto alla quale classificare i frammenti, si può provare a tradurla in una forma misurabile mediante opportuni algoritmi di analisi. Fatto ciò, il sistema può, con la velocità e la consistenza tipica dei sistemi informatici, esaminare e classificare l'intera banca dati anche rispetto alla nuova caratteristica di interesse. Questo significa risparmiare il tempo prezioso degli operatori, ottenere un risultato non affetto da fluttuazioni inter-operatore o soggettive e permettere lo svolgimento della classificazione in modo incrementale anche ad operazioni di ricomposizione già avviate.

Il confronto fatto dagli operatori tra frammenti ed immagine di riferimento è stato migliorato ed agevolato da possibilità offerte dal digitale (utili ad enfatizzare il contenuto informativo delle immagini e a renderlo meglio percepibile) ed impossibili nel mondo reale: i colori dell'immagine di riferimento possono essere corretti, sulla base dei frammenti già collocati, in modo da avvicinarsi a quelli dei frammenti senza costringere l'operatore a compiere tale operazione mentalmente; il contrasto e la luminosità del fondo possono essere variati dinamicamente; il frammento da collocare può essere reso semitrasparente per permettere la

simultanea valutazione della sua pellicola pittorica e del fondo; i frammenti già posizionati possono essere resi temporaneamente invisibili per evitare di occludere il fondo nel collocare i nuovi, ...

Il sistema inoltre prevede il superamento della localizzazione geografica del lavoro di ricomposizione (Figura 2). La sua architettura, attraverso un server (un elaboratore con più processori simmetrici con un'elevata quantità di memoria di massa) ed una collezione di stazioni di lavoro client (una per ciascun operatore), è stata progettata per rendere il laboratorio virtuale di ricomposizione distribuibile su base geografica. Non è dunque più necessario che tutti gli operatori convergano in un'unica sede, causa di significative inefficienze nella pratica di cantiere (nel caso specifico ubicato necessariamente presso la Basilica di Assisi). Più operatori, che possono ciascuno risiedere ovunque sia disponibile una connessione ragionevolmente efficiente (modem telefonico anche a bassa velocità) ad Internet e quindi con il server, possono cooperare al progetto di ricomposizione del medesimo affresco.

L'architettura client-server consente di suddividere il carico di lavoro in modo efficiente: le interrogazioni del database vengono gestite dal server centrale rendendo possibile ai restauratori di proseguire l'attività di ricomposizione sulla stazione client loro assegnata.



**Figura 2 - Il sistema sviluppato prevede il superamento della localizzazione geografica del cantiere di ricomposizione. Il laboratorio virtuale è composto da un server (un elaboratore con più processori simmetrici con un'elevata quantità di memoria di massa) ed una collezione di stazioni di lavoro client (una per ciascun operatore). Più operatori, da qualunque parte del mondo purché sia disponibile una connessione anche a bassa velocità ad Internet, possono cooperare in modo organizzato alla ricomposizione del medesimo affresco**

L'organizzazione dei dati tra client e server rispecchia questa impostazione: infatti, mentre le immagini dei frammenti sono ospitate su ciascuna stazione client le differenti caratteristiche estratte dalle immagini dei frammenti sono memorizzate sul server di ricerca.

## **2 Costruzione del database del sistema**

La realizzazione del sistema ha richiesto l'estrazione e la pre-elaborazione delle immagini dei singoli frammenti e la costruzione del database che ospita i loro dati caratteristici. Disporre di una tale collezione è fondamentale sia perché il sistema consente ad ogni operatore di manipolare i singoli frammenti nello spazio di lavoro sia per lo sviluppo dei vari moduli che portano alla costruzione del database all'interno del quale l'operatore può effettuare ricerche a partire da opportuni esempi. Dalle immagini digitali dei contenitori dei frammenti fornite dall'Istituto Centrale per il Restauro, attraverso il Modulo di Estrazione dei Frammenti, sono stati generati i file che contengono le immagini dei singoli frammenti ( $F_1, F_2, \dots, F_k$ ) che costituiscono la "collezione dei frammenti". Per ciascuna di queste immagini, attraverso il Modulo di Estrazione delle Maschere, è stata generata la relativa maschera il cui insieme costituisce la "collezione delle maschere" ( $M_1, M_2, \dots, M_k$ ). Le immagini dei singoli frammenti e delle relative maschere sono allocate su ogni macchina client per massimizzare la velocità di accesso ai dati. Ogni frammento con la propria maschera è stato poi elaborato dal Modulo di Analisi del Colore (per estrarre le caratteristiche cromatiche  $I_1, I_2, \dots, I_k$ ) che ha consentito di generare anche la "palette globale" dei colori e dal Modulo di Analisi della Tessitura che ha generato le descrizioni  $T_1, T_2, \dots, T_k$ . I dati prodotti dal Modulo di Analisi del Colore e dal Modulo di Analisi della Tessitura popolano il database allocato sul server.

### ***2.1 Estrazione delle immagini dei frammenti e creazione delle maschere***

I frammenti sono conservati in appositi contenitori, in cui sono vincolati da uno strato di foam che impedisce che possano spostarsi accidentalmente. In ciascun contenitore ne sono posizionati (in dipendenza dalla grandezza) fino ad oltre 300.

Per estrarre le immagini dei singoli frammenti dalle immagini dei contenitori sono stati sviluppati algoritmi di elaborazione basati sull'analisi delle caratteristiche cromatiche delle immagini. Lo strato di foam sul quale sono collocati i frammenti è di colore molto scuro e pertanto un'analisi del valore delle terne RGB di ogni pixel dell'immagine di partenza consente di distinguere i frammenti dal fondo. La tecnica utilizzata scandisce l'immagine del contenitore nella regione di interesse ossia nella sola zona dell'immagine contenente i frammenti. A partire da un pixel potenzialmente appartenente ad un frammento si procede con tecniche iterative all'inserimento dei pixel adiacenti in una struttura dati che raccoglie tutti i pixel del frammento in questione.

Durante l'estrazione, le immagini dei singoli frammenti ricevono un identificativo univoco che descrive il cassetto di appartenenza e, attraverso un codice numerico ed un'immagine opportunamente prodotta, la sua posizione all'interno del cassetto: ciò permette di ritrovare rapidamente in cantiere il corrispondente oggetto fisico.

L'informazione su quali pixel in ciascuna immagine facciano parte del frammento vero e proprio (da visualizzare nel sistema e su cui applicare le analisi di colore e tessitura) e su quali appartengano al fondo è data da un'immagine binaria (maschera): essa rende anche possibile eseguire una stima della superficie di ciascun frammento, permettendo di ordinare i frammenti restituiti dalla fase di ricerca anche in base alla loro dimensione.

La costruzione di ciascuna maschera 0 ha richiesto i seguenti passi:

- applicazione di un algoritmo di clustering dei colori
- binarizzazione
- filtraggi.

L'algoritmo di clustering utilizzato, fast global k-means [6], suddivide i colori presenti nell'immagine in sottoinsiemi disgiunti con caratteristiche cromatiche omogenee (cluster). Le specifiche tecniche usate, di tipo incrementale, aggiungono progressivamente i cluster alla soluzione attraverso una procedura di ricerca deterministica e globale. L'algoritmo per risolvere il problema con M cluster (M da definire a seconda dell'applicazione) risolve in modo sequenziale tutti i problemi intermedi con rispettivamente 1,2,...,M-1 cluster. Il metodo è efficace e non dipende da condizioni iniziali o parametri che devono essere impostati empiricamente [8]. Una serie di test preliminari ha consentito di fissare a 10 il numero di cluster in cui suddividere l'intera immagine di ogni frammento (costituita dal frammento vero e proprio e dallo sfondo) per avere un numero sufficiente di rappresentanti sia per i colori del frammento vero e proprio sia per lo sfondo.

Il successivo processo di binarizzazione permette di rappresentare l'immagine clusterizzata con due colori (bianco per il frammento, nero per lo sfondo) attraverso l'applicazione di una soglia sulla componente di luminosità dei prototipi ottenuti dal processo di clusterizzazione, al fine di separare i cluster relativi allo sfondo da quelli associati ai frammenti. La stessa soglia è stata utilizzata per la quasi totalità delle immagini consentendo di ottenere delle immagini binarie soddisfacenti.

Le immagini ottenute dal processo di binarizzazione presentano ancora un gran numero di artefatti: infatti regioni di dimensioni diverse (da piccoli punti ad aree di dimensioni apprezzabili) vengono classificate in modo errato a causa delle loro caratteristiche cromatiche. Per rimuovere le regioni più piccole ed assegnarle alla classe corretta senza modificare in modo significativo i particolari relativi alla forma dei frammenti nelle immagini binarie, è stato applicato in modo ricorsivo un filtro mediano. Le regioni più grandi, ma sensibilmente più piccole dei frammenti, assegnate dal processo di binarizzazione ad una classe errata sono state individuate ed eliminate per mezzo di un ulteriore algoritmo di rilevazione dei blob basato su uno schema di region growing.

## ***2.2 Modulo di analisi del colore e misure di similarità***

Tra le attività legate alla ricomposizione, un ruolo chiave è svolto dalla valutazione della similarità fra immagini la quale è fortemente legata alle caratteristiche del colore. Dato il ruolo centrale rivestito dall'interazione fra il sistema e i restauratori, le misure di similarità tra immagini devono riprodurre il più possibile le proprietà percettive del sistema visivo umano.

Tra i diversi spazi del colore si è cercato quello più conforme alla percezione umana con l'obiettivo di massimizzare l'accordo tra sistema e restauratori.

Il CIELab è uno spazio metrico del colore standard nel quale ogni colore viene definito tramite i valori delle coordinate L, a, b che sono rispettivamente le componenti di luminanza, rosso-verde e blu-giallo. Lo spazio CIELab può essere considerato come uno spazio del colore approssimativamente uniforme dal punto di vista della percezione dato che a distanze uguali nello spazio del colore corrispondono differenze pressoché uguali nella percezione del colore stesso. La differenza nella percezione fra i colori, nello spazio del colore CIELab, può essere quindi definita come la distanza euclidea fra due colori. Questa metrica si basa su dati provenienti dagli esperimenti di adattamento e discriminazione del colore eseguiti su larghe regioni di colore uniforme in condizioni di adattamento fissate. E' stato però ampiamente dimostrato che la discriminazione e la percezione del colore dipendono profondamente dalla struttura spaziale dell'immagine e quindi la metrica CIELab può mostrare dei limiti quando viene applicata per valutare la similarità di immagini reali nelle quali i colori sono distribuiti in modo molto articolato [9].

La metrica S-CIELab, un'estensione spaziale della rappresentazione CIELab introdotta da Zhang e Wandell, riproduce attraverso la norma euclidea le distanze del colore percepite da un osservatore umano anche in presenza di distribuzioni spaziali complesse. Essa si basa sull'applicazione di un opportuno filtraggio spaziale ai dati dell'immagine a colori al fine di simulare lo sfocamento spaziale che normalmente viene eseguito dal sistema visivo umano. La metrica deve quindi tenere conto dell'influenza della distribuzione spaziale sul colore percepito e di conseguenza sulla discriminazione cromatica. L'algoritmo S-CIELab parte dalla conversione dei dati nello spazio dei colori opponenti. Ognuno dei tre piani nello spazio dei colori opponenti viene successivamente filtrato in accordo con la sensibilità spaziale che l'occhio umano ha per ogni componente: quelle cui l'occhio umano non è molto sensibile vengono ridotte mentre quelle che non sono visibili vengono completamente cancellate. La rappresentazione filtrata viene poi trasformata nello spazio del colore XYZ; è possibile quindi ottenere le componenti Lab. Si ottiene quindi la rappresentazione S-CIELab dell'immagine [10] che può consentire sia di presentare i colori misurati sulle immagini mediante una tavolozza compatibile e plausibile con i colori realmente percepiti dall'occhio umano sia di migliorare le valutazioni di similarità compiute dagli operatori.

Entrambi i risultati sono significativi poiché è l'operatore che deve guidare il processo di ricerca al fine di selezionare all'interno del database i frammenti di interesse reale per il suo lavoro. I risultati dei test effettuati hanno dimostrato che l'S-CIELab genera un istogramma più compatto del CIELab standard e che la palette S-CIELab presenta colori che sono più vicini a quelli percepiti. L'analisi dei risultati evidenzia che i colori più frequenti nello spazio S-CIELab giacciono molto vicini al colore percepito; inoltre i pixel delle immagini sono associati, nello spazio S-CIELab, a pochi colori aventi una distanza da quello atteso molto vicina alla minima differenza percepibile. Ulteriori esperimenti hanno mostrato che la ricerca eseguita usando gli istogrammi nello spazio S-CIELab è più selettiva. La percentuale di frammenti restituiti dal sistema e successivamente validati dal restauratore in alcuni casi è maggiore se si lavora nello spazio S-CIELab. Inoltre, l'uso della metrica S-CIELab ha consentito anche di posizionare correttamente dei frammenti che non erano mai stati riconosciuti in precedenza come appartenenti alle aree di interesse [11].

È importante comunque notare che i risultati ottenuti con entrambe le metriche possono essere influenzati da alcuni fattori non facilmente controllabili:

- alcuni dei pezzi originati dal processo di frammentazione non sono presenti nella collezione dei frammenti
- il restauratore è responsabile dell'intero processo di ricomposizione, e quindi è necessario tenere conto della soggettività dell'operatore nella valutazione dei frammenti restituiti dal sistema in seguito alla ricerca per similarità
- le palette di riferimento CIELab ed S-CIELab dell'intero affresco usate nei test di similarità coprono solo una percentuale definita dall'utilizzatore di tutti i colori presenti nei frammenti.

Lo spazio Lab dei colori è stato discretizzato: per ogni asse è stato valutato un range di validità e su di esso un insieme di valori discreti uniformemente distribuiti: ogni colore Lab rilevato nelle immagini viene quindi approssimato dai valori discreti più vicini. Ogni range è stato calcolato trasformando gli otto vertici dello spazio RGB nello spazio Lab e quindi aumentando leggermente ogni range della componente Lab ottenuto. Ciò rende possibile rappresentare ogni colore Lab come una terna di indici.

Per stimare la similarità si è partiti dalla tecnica di intersezione degli istogrammi tradizionale [12] che, per ogni colore comune a due immagini, calcola il minimo tra le frequenze con cui esso è presente nei due istogrammi. La stima di similarità è data dalla somma dei valori minimi così calcolati. Tale tecnica, estremamente intuitiva e semplice da implementare, ha lo svantaggio di risultare lenta quando applicata al confronto di istogrammi con un grande numero di colori. Per velocizzare il calcolo della similarità è stato quindi seguito un approccio che si avvale di una "palette globale" dei colori presenti nei frammenti dell'intera collezione (ordinati in modo decrescente rispetto alla loro frequenza) dalla quale sono stati rimossi alcuni colori ritenuti poco significativi perché poco frequenti. Sono stati considerati significativi i colori le cui frequenze sommate sono in grado di garantire una percentuale di copertura del dipinto non inferiore al 97%. Ciò seleziona circa 3.800 colori sul totale dei quasi 40.000 presenti nella palette globale completa (il primo colore scartato con queste impostazioni è presente in solo 100.000 pixel, pari ad una immagine di circa 330 x 330 pixel). La palette globale ridotta può essere utilizzata per descrivere tutti i frammenti su una base comune. I nuovi istogrammi sono costituiti dalle frequenze, valutate rispetto al numero complessivo di pixel del frammento, con cui ciascun colore della palette globale ridotta è presente nel frammento stesso. Questa nuova rappresentazione permette di impostare un nuovo algoritmo per valutare la similarità tra i frammenti sulla base delle loro caratteristiche cromatiche in modo semplice e rapido.

### *2.3 Analisi della tessitura e misure di similarità*

La principale limitazione delle misure di similarità che si basano sull'analisi statistica del colore presente nelle immagini consiste nel non poter sfruttare l'informazione relativa alla distribuzione spaziale dei colori (il calcolo per il confronto delle immagini si basa sugli istogrammi). Per poter sfruttare anche l'informazione spaziale relativa alla distribuzione dei colori nelle immagini, è necessario ricorrere all'analisi della tessitura che, nella comunità della visione artificiale, indica lo studio delle proprietà delle immagini legate al modo in cui i colori (o i livelli di grigio) sono distribuiti al loro interno [13].

L'approccio proposto si basa sul calcolo delle matrici di co-occorrenza [14]. I colori relativi alle immagini dell'intera collezione di frammenti sono immagazzinati nella palette globale ridotta e vengono espressi nello spazio del colore Lab; i pixel delle immagini di partenza, caratterizzati dai colori non appartenenti alla palette comune, vengono associati al colore della palette comune dal quale presentano la minima distanza euclidea. I colori appartenenti alla palette globale vengono ridotti, mediante l'applicazione di un algoritmo di clustering (fast global k-means). La scelta del numero degli  $N$  colori chiave influenza lo spazio occupato da ciascuna matrice di co-occorrenza ( $N \times N$ ), per cui tale numero deve essere piuttosto basso: una serie di test effettuati ha portato a considerare come ottimale un numero di colori chiave pari a 64. Il contenuto di ogni cella della matrice di co-occorrenza, individuata dai colori rispettivamente sulla riga  $i$  e sulla colonna  $j$ , corrisponde al numero di volte che quella coppia di colori viene osservata in due pixel adiacenti dell'immagine. Per esaltare le variazioni dei colori presenti nell'immagine che non possono essere ricavate dall'istogramma, vengono cumulate solo le occorrenze tra pixel caratterizzati da colori distinti. Infine, le matrici vengono normalizzate rispetto alle occorrenze cumulate nelle singole celle della matrice.

Alcune scelte operative consentono di ridurre i tempi di calcolo della matrice di co-occorrenza. Per ragioni di simmetria, viene presa in considerazione solo la metà dei pixel appartenenti all'intorno  $3 \times 3$  di ciascun pixel oggetto di elaborazione. Inoltre, per evitare di dover gestire casi particolari, i pixel del bordo dell'immagine non sono oggetto di elaborazione.

La misura di similarità usata, basata su una variante della tecnica di intersezione degli istogrammi, lavora su due matrici di co-occorrenza relative a due immagini distinte  $A$  e  $B$ ; in particolare, la similarità fra due immagini viene calcolata come la somma dei valori minimi relativi alla frequenza di co-occorrenza di ciascuna coppia di colori in ogni immagine. Il massimo punteggio di similarità che si può ottenere fra due immagini, in seguito alla normalizzazione precedentemente effettuata, è pari ad 1. Il tempo di calcolo della similarità cresce linearmente all'aumentare del numero delle celle delle matrici di co-occorrenza da confrontare ma risulta essere sempre trascurabile rispetto al tempo necessario a caricare da disco una matrice di co-occorrenza.

Nel corso dello sviluppo del sistema è stato anche investigato il metodo Fbim (Feature based interaction map) che consente di descrivere le interazioni, in termini di differenza di intensità dei livelli di grigio, tra coppie di pixel presenti in un'immagine [15]. Il metodo Fbim si articola attraverso due passi fondamentali: la costruzione della Mappa delle Interazioni e l'estrazione delle feature da tale mappa (direzionalità, simmetria, regolarità, orientazione del pattern, ...).

L'algoritmo sviluppato genera l'istogramma esteso delle differenze dei livelli di grigio (EGLDH, Extended Gray-Level Difference Histogram) che per ogni pixel tiene conto delle differenze dei livelli di grigio tra esso e una serie di posizioni dell'immagine che vengono individuate da una distanza massima  $d_{max}$  e da una variazione angolare identificata da un angolo  $\theta$  e un passo  $d$ . A partire dall'istogramma esteso si costruisce la Mappa Polare delle Interazioni che consente di rilevare la relazione tra i livelli di grigio di coppie di pixel in un'immagine in dipendenza della loro posizione relativa (distanza e orientazione). Attraverso una trasformazione cartesiana è possibile ottenere la Mappa Cartesiana che ha il vantaggio di conservare la struttura dell'immagine analizzata.

Le caratteristiche da estrarre dalla Mappa delle Interazioni dovrebbero essere invarianti alla rotazione e possibilmente anche allo scaling al fine di descrivere la tessitura a prescindere

dall'orientazione dei singoli frammenti, disposti in modo casuale nei cassettei. In questo modo, frammenti adiacenti nel dipinto originale e presumibilmente contenenti la stessa informazione di tessitura possono essere descritti tramite le stesse caratteristiche ed il sistema potrebbe fornire una risposta migliore alle interrogazioni. Un'analisi del metodo ha evidenziato che la Mappa Polare al variare della rotazione subisce uno shift ciclico lungo le righe; ciò ha portato a scegliere come feature per la descrizione della tessitura due proiezioni significative della Mappa Polare che sono in grado di fornire informazioni invarianti rispetto alla rotazione.

I metodi di analisi della tessitura che lavorano solo sui livelli di grigio non sono però sufficienti a rilevare tutte le interazioni spaziali fra i pixel di un'immagine. Le informazioni sulla tessitura infatti vengono perse nel momento in cui i colori dell'immagine, espressi in livelli di grigio, sono caratterizzati dalla stessa intensità. Ogni frammento è stato descritto da un insieme più o meno numeroso di pattern circolari e l'analisi della tessitura è stata eseguita sui tre canali L, a, b di ciascun pattern considerato, cui è stata associata una coppia di mappe di interazione. Parte della sperimentazione ha riguardato la suddivisione in classi omogenee delle mappe di interazione associate ai diversi pattern circolari, per ciascuno dei canali analizzati. L'analisi numerica dei risultati ha consentito di individuare alcuni indici ritenuti significativi e consistenti per descrivere le proprietà caratteristiche di ciascuna classe di tessitura.

### 3 Correzione del colore

Le immagini digitali fornite dall'Istituto Centrale per il Restauro, ed utilizzate per lo sviluppo del sistema di assistenza alla ricomposizione virtuale dell'affresco, presentano caratteristiche cromatiche problematiche nonostante la cura posta dagli operatori durante l'acquisizione affinché i dispositivi di ripresa riducessero al minimo le variazioni geometriche, di luminosità, di colore e di scala. Inoltre l'unica immagine dell'intera vela prima della frammentazione di cui si dispone, utilizzabile come riferimento, è una foto acquisita diversi anni prima del terremoto in condizioni di illuminazione non note ed avente colori molto differenti da quelli dei frammenti. Tutto questo ha reso necessario lo studio di tecniche di correzione del colore per ridurre l'impatto di tali inconvenienti sul corretto funzionamento del sistema.

Il metodo usato si basa sul calcolo di una matrice di trasformazione che, attraverso una roto-traslazione delle terne RGB, consente a coppie di colori corrispondenti di ridurre la propria distanza nello spazio del colore. I coefficienti della matrice di trasformazione vengono determinati risolvendo un sistema lineare sovradimensionato del tipo  $Ax = b$ , dove la matrice **A** contiene le terne RGB dei colori da correggere, la matrice **x** è la trasformazione da applicare e **b** contiene le terne RGB dei colori di riferimento. Il metodo dei Minimi Quadrati [16] cerca la migliore soluzione **x** affinché sia possibile minimizzare la distanza euclidea tra le terne da correggere e le terne dei colori di riferimento; per applicare tale metodo è stato utilizzato il metodo di decomposizione ai valori singolari (SVD) della matrice **A** [17].

Per ridurre le variazioni di colore tra le immagini dei diversi cassettei è stato applicato il metodo descritto sfruttando come riferimento colorimetrico le color chart contenute in ogni cassetto. Per ciascun cassetto, all'interno della color chart sono state estratte le aree relative a diciotto colori. Tra tutte le color chart disponibili ne è stata scelta una che, per ampiezza dell'intervallo di colori e luminosità, costituisse il riferimento migliore.

L'algoritmo è stato applicato a tutte le color chart disponibili, allo scopo di calcolarne la relativa matrice di trasformazione da applicare, in seguito, all'immagine del cassetto corrispondente.

La funzionalità del sistema di ricerca dei frammenti all'interno del database, basata sul paradigma "query-by-example", consente di fornire al sistema come esempi o piccole aree estratte dalla vela di riferimento o frammenti o parti di essi. Il sistema analizza l'esempio sulla base delle sue caratteristiche cromatiche che vengono confrontate con quelle di tutti i frammenti immagazzinati nel database, per poi restituire le immagini dei cinquecento frammenti ad esso più simili rispetto al criterio di similarità implementato. L'indice di similarità, calcolato usando la tecnica di intersezione degli istogrammi modificata, consiste in un valore nell'intervallo [0,1] che misura quanto ciascun frammento restituito dal sistema sia simile all'immagine di esempio fornita; il valore "1" indica la più alta similarità e si ottiene solo quando il frammento restituito coincide con l'esempio dato.

Per eseguire la ricerca usando come esempi piccole aree estratte dalla vela di riferimento ed ottenere risultati consistenti, è stato necessario correggere i colori della vela, ovvero ridurre la differenza rispetto al colore dei frammenti corrispondenti attraverso opportune trasformazioni [18]. A causa della grande estensione, l'affresco ha subito nel tempo (a causa di processi fisici, chimici, ...) trasformazioni cromatiche differenti nelle sue diverse regioni. Per correggere i colori dell'intera immagine non è, quindi, sufficiente un'unica trasformazione ma diventa necessario identificare correzioni locali la cui validità può essere ipotizzata solo in un'area di limitata estensione. La correzione cromatica dell'immagine della vela permette anche di semplificare il lavoro dei restauratori che, per posizionare correttamente i frammenti muovendoli nell'area di lavoro, possono confrontare le immagini dei frammenti e l'immagine di riferimento in modo più agevole.

La metodologia applicata per la correzione dell'immagine della vela di riferimento si basa sull'estrazione di piccole aree (patch) dall'immagine dell'affresco e dalle corrispondenti zone dei frammenti già ricollocati. Da queste aree vengono estratte le terne corrispondenti che permettono di risolvere il sistema e determinare la trasformazione cercata.

Le sperimentazioni sono state condotte operando secondo due modalità differenti: utilizzando patch monocromatiche o policromatiche. Nel primo caso vengono estratte da un frammento e dalla porzione corrispondente della vela delle aree che risultano essere omogenee in termini di colore; nel secondo caso si considerano interi frammenti e si estraggono dalla vela le aree corrispondenti (stessa forma e dimensione).

Le componenti RGB contenute nei vettori **A** e **b**, nel primo caso, sono costituite dal valor medio dei colori presenti in patch omogenee dal punto di vista del colore (monocromatiche) estratte manualmente dall'immagine di riferimento e dalle corrispondenti aree dei frammenti.

Gli esperimenti sulle patch monocromatiche sono stati condotti lavorando su due zone distinte dell'affresco. Sono stati generati quindi due differenti sistemi la cui soluzione ha dato origine rispettivamente alle due distinte matrici di rototraslazione, utilizzate per correggere le corrispondenti aree dell'affresco.

Le sperimentazioni effettuate hanno messo in evidenza che le matrici di correzione, quando applicate all'immagine di riferimento, rendono i colori della vela molto simili a quelli dei frammenti reali. Ciò consente di rendere significativa la ricerca dei frammenti nel database a partire da dettagli estratti dall'immagine di riferimento corretta e, nello stesso tempo, rende

effettivamente più agevole il lavoro degli addetti alla ricomposizione dell'affresco che si trovano a dover confrontare immagini simili dal punto di vista cromatico.

Obiettivo della sperimentazione condotta a partire da patch policromatiche è stato di ottenere un metodo per eseguire la correzione del colore della vela di riferimento che possa operare in modo automatico, ossia un metodo che calcoli la correzione del colore in una regione intorno ad ogni frammento posizionato senza che l'operatore debba manualmente individuare ed estrarre corrispondenze di colore. Per contenere gli effetti del rumore presente nelle due immagini, l'istogramma di entrambe è stato descritto utilizzando solo un insieme ridotto di colori predominanti, identificati sfruttando il medesimo algoritmo di clustering "Fast Global k-means". Al fine di conoscere le associazioni tra i due insiemi di colori sono state verificate due differenti metodologie.

Nel primo approccio, i due set di colori ordinati rispetto alla coordinata R delle terne RGB forniscono le equazioni del sistema  $\mathbf{Ax} = \mathbf{b}$  da risolvere per calcolare la matrice di trasformazione del colore. Questa sperimentazione è stata eseguita utilizzando sei cluster nello spazio del colore.

Il secondo metodo procede alla ricerca delle corrispondenze dei colori effettuando un confronto pixel a pixel tra l'immagine del frammento e quella della patch della vela di riferimento. Affinché un tale confronto sia possibile, è necessario rendere le due immagini sovrapponibili. Tale risultato è stato raggiunto attraverso una fase di pre-elaborazione che ha consentito di ruotare e scalare l'immagine della patch estratta dalla vela di riferimento fino a portarla in corrispondenza al frammento corrispondente. Si è deciso di effettuare queste pre-elaborazioni sull'immagine della vela per evitare di perdere parte delle informazioni contenute nell'immagine del frammento che ha una maggiore risoluzione. In questo caso la scelta dei colori è stata compiuta utilizzando otto cluster. Si è proceduto quindi alla scansione delle immagini così ottenute per raccogliere le associazioni dei colori in esse presenti. L'algoritmo costruisce la Matrice dei Conteggi, caratterizzata da un numero di righe pari al numero dei colori della vela e un numero di colonne pari al numero di colori del frammento (8x8). Per ogni pixel i colori corrispondenti nel frammento e nell'immagine della vela determinano rispettivamente l'indice colonna e l'indice riga della Matrice dei Conteggi: la posizione così identificata viene incrementata. Al termine della scansione delle due immagini le associazioni più frequenti dovrebbero identificare le coppie di colori corrispondenti. Tuttavia tale metodo è influenzato negativamente dalla rumorosità dell'immagine di riferimento il cui gran numero di colori spuri condiziona il processo di clusterizzazione, facendo apparire significativi colori non rappresentativi dell'immagine.

#### **4 Valutazioni**

Il primo dimostratore è stato sviluppato su stazione singola e dotato solo delle funzionalità principali richieste per la convalida del nuovo approccio alla ricomposizione: esso è stato provato su una copia di una parte della vela (specificamente realizzata dall'Istituto Centrale del Restauro) ed ha ricevuto l'incoraggiante accettazione della comunità del restauro. Questo ha consentito di avviare con fiducia il completamento di un sistema dotato di tutte le funzionalità, tecnologiche e scientifiche, necessarie per una sfida come la ricomposizione di affreschi della complessità della vela di S. Matteo. Due stazioni client ed il server dell'architettura sono stati sviluppati ed installati presso l'ISSIA-CNR di Bari mentre una stazione client è stata configurata presso l'ICR di Roma, dove il prezioso lavoro degli operatori ha consentito di completarne la messa a punto.

Il risultato è stato e si mantiene comunque aperto all'inserimento di nuovi strumenti che la comunità del restauro e la ricerca scientifica possano rendere necessari e disponibili. La flessibilità e la potenza di questa architettura sembra sufficiente a permetterne l'uso proficuo nell'attività quotidiana degli operatori impegnati nella sfida costituita dalla ricomposizione di frammenti per migliorare l'efficacia e l'efficienza del loro lavoro.

Gli operatori che hanno sperimentato il dimostratore hanno colto e confermato come gli strumenti informatici per il riassetto virtuale possano e debbano essere intesi come elementi che si aggiungono in modo complementare al loro bagaglio di conoscenze. Tali strumenti si caratterizzano per la loro capacità di gestire in modo integrato e veloce una grande quantità di informazioni, consentendo innanzitutto una riduzione di tempo associata alla possibilità di riferirsi a dati oggettivi. Il sistema permette di effettuare rapidamente il confronto di uno o più frammenti con tutti gli altri secondo un'analisi di parametri quali: colore, tessitura, area, ... . Ciò può fornire all'operatore in brevissimo tempo gruppi coerenti di frammenti sui quali poter lavorare. Lo studio del frammento e la sua collocazione possono essere agevolati da una serie di funzioni accessorie: l'ingrandimento della superficie pittorica (che può evidenziare particolari che possono sfuggire ad un'osservazione diretta), la resa in trasparenza del frammento, l'uso diretto di coordinate geometriche, la correzione del colore.

Esso ha inoltre dimostrato di poter eliminare i *tempi di disturbo* e il *frastuono di fondo*: la possibilità infatti, di avere contemporaneamente presenti su uno schermo continuo l'immagine di riferimento generale e porzioni di essa ingrandite, elimina la difficoltà, che si verifica nella dimensione operativa reale, di effettuare un riscontro immediato fra *cassetto-frammento-piano di lavoro*, in particolare quando quest'ultimo si estende su una superficie molto vasta. Allo stesso tempo la concentrazione dell'operatore è agevolata dal muoversi all'interno non più di uno spazio tridimensionale ma bidimensionale. Infine la ricomposizione virtuale dei frammenti sull'immagine di riferimento può avere anche il vantaggio di fornire immediatamente ai restauratori la restituzione pittorica della superficie dipinta ricomposta, informazione importante soprattutto quando si è in presenza di configurazioni particolari quali superfici curve come quelle della volta della Basilica di S. Francesco.

La sperimentazione del sistema nella sua configurazione attuale si è articolata in due fasi. Dall'aprile al dicembre 2003 l'attività di ricomposizione virtuale dell'affresco è stata affidata in modo esclusivo ai restauratori che presso la sede ICR hanno utilizzato una stazione client, collegata al server ubicato presso l'ISSIA-CNR di Bari, hanno acquisito la necessaria conoscenza e familiarità sulle funzionalità del sistema ed hanno verificato le diverse funzionalità da esso offerte operando sulle immagini dei frammenti della vela del S. Matteo. Da tale lavoro sono emerse indicazioni utili a migliorare e arricchire le caratteristiche del sistema. Un esempio di tali funzionalità aggiuntive è costituito dalle liste di esclusione: l'operatore può non solo limitare la ricerca dei frammenti ad una parte del database (restringendola solo ai frammenti contenuti in un contenitore specificato) ma può anche escludere dalla ricerca tutti i frammenti contenuti in apposite liste, definibili su base individuale utilizzando diversi criteri logici (mancanza di elementi pittorici significativi, appartenenza dei frammenti ad aree dell'affresco diverse da quella di interesse in un particolare momento della ricomposizione, ... ). L'interrogazione del database avveniva connettendosi al server tramite Internet con un modem a 33 kbit/s. Il risultato dell'attività di ricomposizione svolta in questo periodo è stato il posizionamento di 418 frammenti virtuali, tutti relativi all'area dell'affresco (cornice decorativa esclusa). La superficie così ricomposta è pari a circa 0,72 m<sup>2</sup>, pari cioè a circa il 2,1 % della superficie dell'intero affresco.

Tale stima è stata ottenuta considerando il rapporto fra il numero di pixel contenuti nell'area occupata dalla pellicola pittorica nella totalità dei frammenti ricollocati e il numero di pixel contenuto nell'area di un frammento virtuale corrispondente ad un cm<sup>2</sup>. Prendendo in considerazione le immagini dei frammenti elaborate, la superficie ricomponibile dell'affresco è stata stimata essere pari a 33,5 m<sup>2</sup>. Al termine di tale fase l'ICR, sulla base delle proprie difficoltà a mantenere allocate le necessarie risorse e di considerazioni legate alle scadenze delle attività di cantiere, ha valutato di sospendere la sperimentazione con il sistema.

In seguito a tale decisione l'ICR ha continuato attività di ricomposizione dei frammenti secondo la metodica tradizionale, benché consapevole che la rimozione dei frammenti dal foam e dai cassetti avrebbe potuto pregiudicare l'immediato reperimento dei frammenti fisici relativi ad ulteriori risultati ottenuti utilizzando il sistema. I frammenti ricollocati in modalità digitale vengono infatti, attraverso un identificativo digitale univoco, associati al cassetto di giacenza ed alla posizione all'interno dello stesso. La rimozione dei frammenti dai cassetti può vanificare tale associazione. Tale eventuale problema può comunque essere risolto ricreando la corrispondenza tra immagini digitali ed oggetti fisici, collocando questi ultimi in nuovi cassetti (ovviamente non necessariamente corrispondenti a quelli precedenti), acquisendo nuove immagini di questi cassetti e ritrovando la nuova posizione del frammento fisico corrispondente a ciascuna immagine contenuta nel sistema.

Dal 15 settembre 2004 al 13 marzo 2006, si è svolta una seconda fase sperimentale durante la quale il CNR è riuscito ad allocare risorse proprie per la sperimentazione del sistema e l'attività di ricomposizione virtuale della vela del S. Matteo. Tali risorse, pur scelte in modo da coniugare il più possibile la familiarità con gli strumenti informatici con quella verso le immagini di tipo artistico e le tecniche pittoriche, non avevano, a differenza del personale ICR della prima fase, alcuna precedente esperienza di cantiere legata alla ricomposizione dei frammenti. In parallelo è proseguita anche l'attività di ricerca riguardo lo sviluppo di nuovi algoritmi per aumentare e migliorare le funzionalità e le prestazioni del sistema benché non sia stato possibile integrare tali nuovi risultati scientifici nel sistema adottato per la ricomposizione virtuale dell'affresco.

L'attività di ricomposizione svolta in questo periodo ha portato al posizionamento di 2233 frammenti, il che ha prodotto, sommato ai frammenti collocati dall'ICR nella prima fase, ad un risultato complessivo di 2651 frammenti, con i quali si è conclusa nel marzo 2006 l'attività in oggetto. La quasi totalità di tali frammenti ricollocati virtualmente appartiene all'area utile dell'affresco, verso la cui ricomposizione si è preferito dedicare la maggior parte dell'impegno. Si è stimato che la superficie complessivamente ricomposta sia equivalente a circa 2 m<sup>2</sup>, pari cioè a circa il 6 % della superficie totale dell'affresco.

Il fatto che la superficie ricomposta sia quasi triplicata, mentre il numero dei frammenti posizionati sia aumentato di più di un fattore sei fa comprendere come durante la seconda fase della ricomposizione virtuale sia stato possibile ricollocare frammenti caratterizzati da un'estensione significativamente inferiore (rispetto a quelli posizionati in precedenza).

La tabella mostra in modo riassuntivo i dati relativi alla sperimentazione svolta.

	<b>ICR (con il sistema)</b>	<b>ISSIA (con il sistema)</b>
<b>Frammenti ricollocati</b>	418	2651
<b>Superficie totale ricomposta [cm<sup>2</sup>]</b>	7216,64	20787,1
<b>Superficie frammento più grande [cm<sup>2</sup>]</b>	240,65	250,55
<b>Superficie media frammenti [cm<sup>2</sup>]</b>	17,2647	7,84125
<b>Area ricomposta [pixel]</b>	112582715	324290183

La figura 3 mostra i risultati conseguiti dall'Istituto Centrale del Restauro utilizzando sia il sistema di ricomposizione virtuale (418 frammenti) che la modalità tradizionale. L'affresco così ricomposto è stato ricollocato sulla volta della Basilica.

La figura 4 mostra i risultati ottenuti presso l'ISSIA-CNR, utilizzando esclusivamente il sistema, da parte di risorse con preparazione artistica ma senza specifica esperienza di ricomposizione di affreschi.



**Figura 3 - Immagine della volta della Basilica di S. Francesco in Assisi dopo il riposizionamento dell'affresco di S. Matteo. I frammenti sono stati ricomposti dall'Istituto Centrale del Restauro per alcune centinaia con il sistema per la ricomposizione virtuale assistita e per la restante parte con la metodologia tradizionale**



**Figura 4 - Il risultato del processo di ricomposizione svolto all'interno dell'ISSIA-CNR utilizzando esclusivamente il sistema di ricomposizione virtuale assistita, senza fare mai uso dei frammenti fisici in cantiere**

È utile al termine di questo lungo percorso condurre alcune considerazioni. Il problema della ricomposizione dei frammenti è ricorrente. Tanto in Italia che all'estero esistono pitture frammentate il cui valore artistico o storico suggerisce ed auspica un tentativo di ricomposizione. Anche in altri ambiti (un esempio sono i ceppi funerari lapidei rinvenuti da tempo e catalogati anche in forma digitale in territorio pugliese) l'uso di uno strumento evoluto permetterebbe di tentare la ricomposizione di frammenti il cui destino alternativo sembra essere la musealizzazione in modo disaggregato quando non addirittura lo stoccaggio nei magazzini dei musei stessi.

In molti di questi casi il sistema sviluppato potrebbe essere impiegato con un minimo sforzo, essenzialmente legato alla preparazione dei dati digitali di input. In qualche caso sono addirittura già disponibili immagini pittoriche dei frammenti e la sperimentazione richiede essenzialmente l'accesso ai dati e la disponibilità di personale da dedicare all'attività di ricomposizione. Il caso della vela di S. Matteo ha costituito una sfida particolarmente difficile, soprattutto per il numero elevatissimo di frammenti che, in larga maggioranza, avevano estensioni estremamente contenute ed erano quindi poveri di informazioni visuali che ne consentissero la caratterizzazione. Inoltre la pellicola pittorica è spesso fortemente deteriorata (abrasioni, viraggio dei colori, ...).

Il confronto con l'immagine di riferimento è reso difficile dalla sua risoluzione molto inferiore a quella dei frammenti e dai colori molto diversi da quelli dei frammenti (l'immagine è stata acquisita diversi anni prima del terremoto in condizioni di illuminazione e di ripresa non note). L'approccio seguito, esplicitamente pensato per indirizzare l'intera problematica della ricomposizione dei frammenti invece che in modo limitato il singolo caso di studio, ha consentito di realizzare uno strumento generale che costituisce il supporto di una metodologia innovativa di ricomposizione, favorevolmente valutata ed accolta dalla stessa comunità del restauro.

## Bibliografia

- [1] Reconstructing the temple ceiling (brujo)  
<http://www.research.ibm.com/peru/reconstr.htm>
- [2] Progetto Mantegna, <http://www.pd.infn.it/~labmante>
- [3] G. IACOVITTI, Supporti strumentali per le operazioni di riassettaggio: tecniche di elaborazione delle immagini,  
<http://www.icr.beniculturali.it/Strumenti/Documenti/Q7.pdf>
- [4] Dall'Utopia alla Realtà 1 - Notizie dal cantiere di S. Francesco in Assisi, 2000,  
<http://www.icr.beniculturali.it/Strumenti/Documenti/Utopiareal.pdf>
- [5] G. ATTOLICO, A. DISTANTE, "Il riassettaggio virtuale assistito di frammenti", Dall'utopia alla realtà, Note dal cantiere dei frammenti della Basilica Superiore di S. Francesco in Assisi, n. 2, April 2001,  
<http://www.icr.beniculturali.it/Strumenti/Documenti/Utopiareal2.pdf>
- [6] F. RENNA, G. CARLOMAGNO, N. MOSCA, G. ATTOLICO, A. DISTANTE, "Virtual recomposition of frescos: separating fragments from the background", Proc. IEEE, 17<sup>th</sup> International Conference on Pattern Recognition (ICPR2004), Cambridge, United Kingdom, 23-26 August 2004, pp. 819-822.
- [7] A. LIKAS, N. VLASSIS, J.J. VERBEEK, "The Global K-Means Clustering Algorithm", Pattern Recognition 36(2), pages 451-461, 2003
- [8] S. THEODORIDIS, K. KOUTROUMBAS, *Pattern Recognition*, Academic Press, San Diego, 199, ISBN 0-12-686140-4
- [9] X. ZHANG AND B. A. WANDELL, "Color Image Fidelity Metrics Evaluated Using Image Distortion Maps", submitted to Elsevier Preprint, 1998
- [10] S-CIELAB: <http://white.stanford.edu/~brian/scielab/scielab.html>
- [11] F. RENNA, L. VENTURINO, G. ATTOLICO, A. DISTANTE, "Spatial sensitive color metric in the virtual aided recomposition of frescos", Color Science and Imaging Technologies, Proc. SPIE, vol. 4922, pp. 15-22, 2002
- [12] M.J. SWAIN, D.H. BALLARD, "Color indexing", International Journal of Computer Vision, Vol. 7, N. 1, Nov. 1991, pp. 11-32, 2004
- [13] B. JÄHNE, "Digital Image Processing: concepts, algorithms and scientific application", Springer-Verlang, pp. 185-192, 1995
- [14] S. SHIM, T. CHOI, "Image Indexing by Modified Color Co-Occurrence Matrix", Proc. of Int. Conf. on Image Processing, 2003
- [15] D. CHETVERIKOV, "Texture Analysis Using Feature Based Pairwise Interaction Maps", Pattern Recognition, Special Issue on Color and Texture, vol. 32, pp. 487-502, 1999
- [16] G.H. GOLUB, C.F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, 1996
- [17] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING, B.P. FLANNERY, *Numerical Recipes in C*, Cambridge University Press
- [18] G. CARLOMAGNO, F. RENNA, N. ANCONA, N. MOSCA, G. ATTOLICO, A. DISTANTE, "Correcting colours for aided recomposition of fragments", Proc. of 4<sup>th</sup> Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP2004), Kolkata, India, 16-18 December 2004, pp. 593-598.

## L'INTELLIGENZA ARTIFICIALE IN ARCHEOLOGIA

*Juan Antonio Barceló*

Departament de Prehistòria  
Facultat de Lletres  
Universitat Autònoma de Barcelona  
Bellaterra, Spain

[jbarcelo@seneca.uab.es](mailto:jbarcelo@seneca.uab.es)

<http://seneca.uab.es/prehistoria/Barcelo/>

Traduzione di Glauco Mantegari

Questo è un corso sull'Intelligenza Artificiale in Archeologia. Di cosa si tratta?

Sebbene non esista una precisa definizione di *Intelligenza Artificiale (IA)*, possiamo intenderla come il tentativo di creare macchine che ragionino ed agiscano come gli esseri umani, che siano capaci di apprendere e di usare la loro conoscenza per risolvere autonomamente dei problemi. I computer sono particolarmente adatti ad eseguire computazioni meccaniche, attraverso l'uso di regole fisse di programmazione. Ciò permette alle macchine artificiali di portare a termine compiti semplici e monotoni in maniera efficiente ed affidabile, azioni per le quali gli esseri umani sono meno portati. Di fronte a problemi più complessi, le cose si complicano. A differenza dell'uomo, i computer hanno difficoltà a comprendere situazioni specifiche e ad adattarsi a nuove situazioni. L'Intelligenza Artificiale si propone di migliorare il comportamento della macchina nell'affrontare compiti di elevata complessità.

Le principali motivazioni che stanno alla base della ricerca sull'IA sono due:

- *indagare la natura ed i meccanismi dell'Intelligenza Umana*. Il computer è utilizzato per modellare, indagare e verificare teorie sul comportamento umano;
- *sviluppare programmi più efficaci facendo in modo che operino in maniera "intelligente"*. L'obiettivo è di "migliorare" il comportamento e l'utilità di un programma, anche senza che questo operi in maniera analoga alla modalità attraverso la quale le persone portano a termine lo stesso compito.

Le direzioni di ricerca della moderna Intelligenza Artificiale, intraprese per la risoluzione di specifici problemi di dominio, sono molte. Esempi di problemi di IA sono la *computer vision* [1] (la creazione di un sistema in grado di interpretare un'immagine come farebbe un archeologo) ed il *natural language processing* [2] (la costruzione di un sistema che possa comprendere una spiegazione archeologica allo stesso livello di un esperto umano).

Nel corso degli ultimi cinquant'anni, la ricerca in ambito IA si è focalizzata prevalentemente sulla soluzione di problemi specifici. Numerose soluzioni sono state concepite e migliorate per ottenere il risultato desiderato in maniera efficiente e affidabile.

Per questo motivo l'Intelligenza Artificiale è suddivisa in molte branche, che vanno dall'*Intelligent Pattern Recognition* all'*Artificial Life*, passando attraverso la *Evolutionary Computation* e la *Computational Intelligence*. Tuttavia la maggior parte delle applicazioni mostrano solo un remoto legame con l'intelligenza.

Certamente l'IA è una disciplina che ha raccolto crescente popolarità ed interesse, ma sta divenendo un'etichetta alla moda, una realtà che viene coperta dall'uso di un gran numero di *buzz-words*.

L'Intelligenza Artificiale è necessaria, in Archeologia?

Gli obiettivi sono:

- *Indagare la natura ed i meccanismi del ragionamento archeologico*. Il computer è utilizzato per modellare, indagare e verificare teorie sul comportamento umano nel passato basandosi sui resti materiali.
- *Sviluppare programmi più efficaci, facendo in modo che i dati archeologici vengano processati in maniera "intelligente"*. Il computer "migliorerebbe" il modo di ragionare degli archeologi anche senza che ci sia un'analogia con il meccanismo attraverso il quale normalmente noi portiamo a termine lo stesso compito.

## 1 Il concetto di Problema in Archeologia

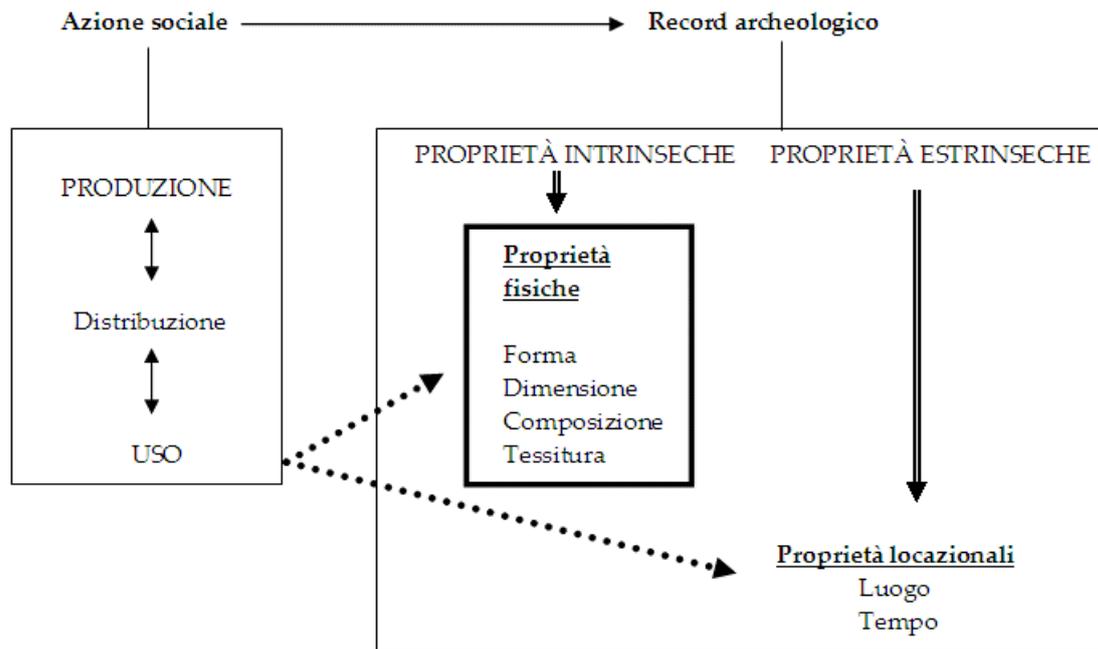
L'idea di *Problema* costituisce il punto di partenza di qualunque ricerca di Intelligenza Artificiale: una persona ha un problema quando ha un obiettivo che non può essere raggiunto in maniera diretta. Ogni volta che non ci si può muovere da una data situazione alla situazione desiderata semplicemente attraverso l'azione, scatta il ricorso al pensiero. Il pensiero ha il compito di individuare una qualche azione che possa mediare tra la soluzione esistente e la soluzione desiderata. In un certo senso possiamo dire che la risoluzione di problemi (*Problem Solving*) sia una qualsiasi sequenza di operazioni cognitive guidata da un obiettivo [3][4]:

$$\text{Problema} = \text{Obiettivo} + \text{Ostacolo}$$

Quando il raggiungimento di un obiettivo è ostacolato, abbiamo un problema; quando però conosciamo il modo per aggirare l'ostacolo o sappiamo come rimuoverlo, il problema è di minore entità. Gli archeologi ragionano molto raramente in termini di problemi da risolvere. La maggior parte delle volte l'Archeologia appare come una disciplina che si occupa di *descrizioni*. Cos'è, dunque, un problema archeologico?

Le domande che generalmente ci poniamo per comprendere il passato sono: di che cosa si tratta? Quando è stato fatto? Chi l'ha fatto? Come è stato fatto? Perché è stato fatto? Questi sono esempi tipici di problemi archeologici. Gli artefatti archeologici hanno forme differenti, dimensioni differenti, composizioni differenti e tessiture differenti. Dovremmo descrivere le differenze in questi elementi e spiegare le origini o le cause della variabilità. Perché le asce di pietra hanno forme e dimensioni differenti? Perché le tombe hanno contenuti differenti? Perché i vasi di ceramica hanno tessiture differenti? In generale la produzione, l'uso e la distribuzione sono i processi sociali che, in qualche modo, hanno prodotto (*causa*) le differenze osservate e la variabilità (*effetto*) (Figura 1).

L'obiettivo dell'Archeologia è di scoprire ciò che non può essere visto (*cause sociali*) nei termini di ciò che è invece visibile (*effetti materiali*). La causa non osservabile di un effetto materiale osservabile è un Problema. Gli oggetti possiedono determinate proprietà fisiche perché sono stati prodotti in un modo tale per cui avevano queste caratteristiche e non altre; ed erano prodotti in quel modo perché, almeno in parte, questi oggetti erano destinati ad alcuni specifici usi e non ad altri: si trattava di strumenti, materiale di scarto, edifici, contenitori, combustibile ecc. Se gli oggetti compaiono in alcuni luoghi e non in altri era perché il loro utilizzo avveniva proprio in quei luoghi e in quei momenti.



**Figura 1 - Un archeologo automatizzato analizza le evidenze archeologiche nel contesto dell'attività sociale identificando i modi in cui le persone hanno prodotto (e/o utilizzato) l'artefatto, i bisogni ai quali ha sopperito e la storia del suo sviluppo**

Ciò equivale a dire che i cambiamenti e le modifiche nella forma, nella dimensione, nella tessitura, nella composizione e nella localizzazione che si riflettono sulla realtà naturale come risultato dell'azione umana (lavoro) sono determinate in qualche modo da queste azioni (produzione, uso, distribuzione) che hanno reso possibile la loro esistenza. Dobbiamo inoltre prendere in considerazione le circostanze ed i contesti (sociali e naturali) nei quali le azioni sono state compiute ed i processi (sociali e naturali) che hanno agito in quel determinato luogo successivamente alla causa originaria, perché possono avere alterato gli effetti originari delle azioni primarie. Spiegare il Passato significa stabilire la sequenza delle azioni sociali necessarie a determinare come la natura sia stata trasformata e come l'artefatto sia stato prodotto in quanto risultato del lavoro umano.

Per risolvere questi problemi dobbiamo rimuovere l'ostacolo che impedisce il raggiungimento dell'obiettivo. *In Archeologia questo ostacolo è costituito dalla non-osservabilità delle cause che generano l'effetto.*

Non possiamo infatti avere una piena percezione dei processi di formazione delle evidenze archeologiche perché:

- la loro natura non è osservabile (non posso osservare il "potere sociale" ma i suoi effetti);
- sono terminate nel presente (Aristotele scrisse la sua *Metafisica* più di duemila anni fa. Non so con certezza se abbia scritto di persona il libro perché si tratta di un'azione del passato. Posso tuttavia inferire che egli sia l'autore perché alcune copie del libro ed alcuni testimoni a lui contemporanei parlano di Aristotele e dei suoi lavori);
- si manifestano molto lentamente. Le montagne e le valli sono fenomeni attuali che si sono originati molto tempo fa per l'effetto congiunto di processi geologici, che possono essere in azione anche in questo preciso luogo ed in questo preciso istante, ma ad una

scala temporale così lenta che non sono in grado di percepire i loro effetti nell'arco della mia vita;

- sono olistici (globali). Non posso osservare come la società umana cambi ed evolva perché ci sono più di sei miliardi di persone nel mondo e non posso avere una percezione delle modalità di azione dei cambiamenti sociali su ciascun individuo.

*Vogliamo essere in grado di vedere ciò che non può essere visto.* Siamo di fronte ad un *problema inverso*: data l'attuale evidenza dell'effetto dovremmo essere in grado di *predire* l'esistenza passata della causa. L'Archeologia è un lavoro di *inverse engineering* perché dovremmo scoprire come le azioni sociali producano i loro effetti materiali; solo così saremo in grado di predire un'azione sociale data la presenza del suo effetto. Esiste una vasta schiera di problemi che sono generalmente descritti affermando che la risposta è nota; non lo è, tuttavia, la domanda. In alternativa, sono noti i risultati e le conseguenze, ma non la causa. Un lavoro di *inverse engineering* può essere definito dall'associazione contingente di:

uno *stimolo in input* (la condizione osservabile)

e

una *risposta in output* (la causa ipotetica)

Il concetto di predizione proviene dall'idea che un problema sarà considerato risolto con successo se saremo stati in grado di produrre, nella maggioranza dei casi, associazioni corrette di tipo comportamentale su alcuni elementi di test. In maniera formale:

*Dato un insieme di esempi di E*

*Fornisci una funzione F*

*che fornisca una predizione dei valori di E*

Una conseguenza materiale che viene osservata *predice* l'azione sociale che l'ha originata se e solo se esiste una distinzione statistica rilevante tra gli eventi che sono conseguenza dell'azione e gli eventi che non sono stati generati da questa azione o da questo processo [5].

Le attività di predizione sono generalmente denominate *attività di supervised learning*, sulla base del fatto che gli esempi sono paragonabili alle informazioni che vengono fornite da un insegnante o da un supervisore ad uno studente [6][7]. L'attività di *supervised learning* consiste nell'acquisire concetti di valore generale a partire da esempi specifici che servono da addestramento. Gli esempi forniti con l'addestramento sono istanze che possono appartenere ad uno specifico concetto – ed in tal caso sono esempi positivi – oppure no – in tal caso sono esempi negativi.

Nell'ambito delle terminologia qui utilizzata ciò che viene appreso è generalmente indicato come *target function*, mentre gli input, e gli output ad essi associati, sono considerati, rispettivamente, come argomenti e valori di una funzione non nota. L'apprendimento viene quindi concettualizzato lungo linee di computazione. I dati forniti sono considerati come *valori di input*, mentre i dati da predire sono considerati come *valori di output*. Il processo di apprendimento è inteso, pertanto, come l'acquisizione della capacità di computare la *target function*. Quando una persona impara come categorizzare un sottoinsieme di eventi in un determinato modo sta facendo di più che riconoscere semplicemente le istanze che ha incontrato. Quella persona sta anche apprendendo una regola che potrebbe essere applicata ad altre istanze. Il concetto di categoria consiste fondamentalmente in questa "regola di raggruppamento"; sono regole come questa quelle che ognuno costruisce quando elabora e definisce concetti.

## 2 Il Problem Solving in Archeologia

Il *problem solving* in Archeologia può essere definito concettualmente come un tipo di apprendimento, dal momento che può essere inteso come l'acquisizione, a partire da dati esistenti, di conoscenza (*decision rules*), che permetta poi di predire osservazioni future. Questo processo implica necessariamente il riconoscimento di una "connessione" tra alcuni elementi osservati dei dati il che, a sua volta ed in maniera estensiva, implica l'esistenza di alcuni elementi non osservati. Di conseguenza, per essere in grado di vedere ciò che non si può vedere, ovvero la causa o il processo di formazione delle evidenze archeologiche, abbiamo bisogno di:

*Dati:*

Una descrizione iniziale di un'entità teorica

Un'istanza di questa entità

Una spiegazione dell'associazione tra il concetto e la sua istanza

Alcuni criteri di tipo operativo

*Determinare:*

Una generalizzazione dell'istanza che sostituisca la descrizione iniziale e che sia collegata con la spiegazione e con i vincoli operativi.

Possiamo tradurre questa definizione generale in un approccio statistico. In questo modo l'apprendimento statistico può essere caratterizzato nei termini seguenti: dato un insieme di esempi di addestramento, l'errore nel risultato dell'apprendimento è dato dall'*errore empirico*, che consiste in una misura basata sui dati di addestramento. La teoria dell'apprendimento statistico si focalizza su due elementi principali:

- la *Asymptotic Analysis*: può essere provato che, con un numero crescente di esempi, l'errore empirico converga verso l'errore reale?
- la *Learning Rate*: se il punto precedente è stato dimostrato, quanto velocemente l'errore empirico converge verso l'errore reale?

Discuteremo meglio di questo approccio in seguito.

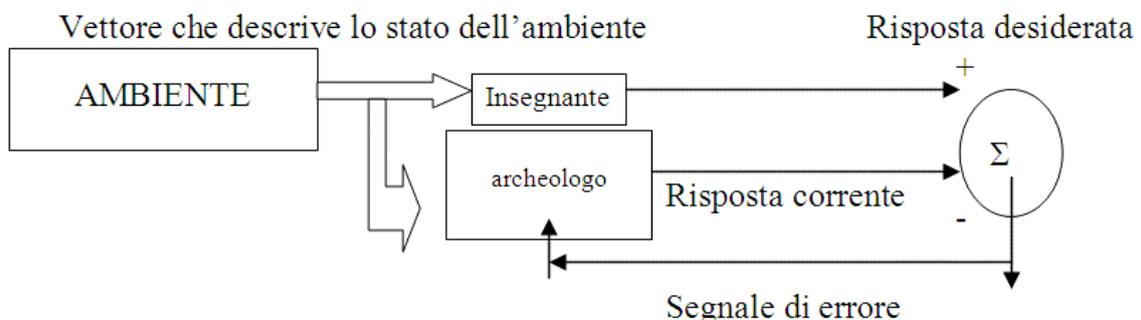


Figura 2 - Diagramma di flusso che illustra il meccanismo di *supervised learning*

Tutto questo significa che, senza una Conoscenza Generale, non ci può essere apprendimento. Se non abbiamo a disposizione istanze note di concetti pre-definiti non possiamo stabilire le associazioni causali, perché la *causalità* non è definita in termini di mere correlazioni statistiche. Esiste, tuttavia, un'altra forma di apprendimento, denominata *Unsupervised Learning*, che consiste nel processo di raggruppamento di esempi di input secondo classi di similarità. La maggior parte dei metodi di *unsupervised learning* consistono, di fatto, in meccanismi di *clustering* che si basano su una qualche misura di similarità. Ogni oggetto è rappresentato come un insieme non ordinato (un vettore) di elementi. Oggetti "simili" sono tutti quelli che possiedono valori simili sui diversi elementi che li costituiscono. Si può pertanto pensare di raggruppare gli esempi in modo da minimizzare le distanze all'interno del gruppo e, al contempo, massimizzare le distanze tra gruppi differenti, tenendo conto dei vincoli sul numero dei gruppi che possono essere formati.

I due meccanismi di apprendimento possono essere tradotti in Archeologia nei termini generalmente noti come Classificazione e *Clustering*, oppure in quelli di apprendimento e partizione. Il *problema della classificazione* è equiparabile a quello del *supervised learning* ad eccezione del fatto che nel *supervised learning* abbiamo un certo numero di *output target* che rappresentano etichette definitorie di classi. I casi esemplificativi definiscono quale tipo di oggetto appartenga a quale categoria. L'obiettivo, in un problema di classificazione, consiste nello sviluppare un algoritmo in grado di assegnare ogni artefatto, rappresentato da un vettore  $x$ , ad una delle classi  $c$  (cronologia, funzione, origine ecc.). Il problema consiste nel trovare la migliore corrispondenza possibile tra i pattern di input (elementi descrittivi) e la risposta desiderata (classi). Un insieme, finito o infinito, di pattern (binari o veri e propri vettori dotati di un valore) deve essere distinto in classi ed un certo problema è definito da un insieme di pattern di addestramento selezionati, che vengono forniti insieme ai corrispondenti nomi delle classi cui appartengono. L'obiettivo è di classificare *tutti* i pattern quanto più correttamente possibile.

Il *clustering* è il processo della suddivisione degli esempi di input in classi di similarità. Gli archeologi, per sviluppare una qualche forma di apprendimento concettuale, hanno utilizzato il *clustering* per anni, al posto di una reale classificazione. È tuttavia importante comprendere quale sia la differenza che corre tra *clustering* e classificazione. Data la differenza tra avere espliciti elementi in comune e implicite relazioni, la classificazione, intesa come apprendimento con una funzione predittiva, non può essere ridotta al *clustering*.

La nostra discussione si è fin qui basata sull'assunzione che il modo di effettuare un'analisi archeologica di tipo causa-effetto sia quello di trasformare alcuni elementi di input in categorie tematiche fisse che possano descrivere l'azione sociale nel corso del tempo. Ciò costituisce sicuramente una grossolana semplificazione della realtà. Le azioni sociali sono notoriamente difficili da descrivere attraverso concetti o etichette di classe fissi, sebbene sia presente una naturale tendenza "accademica" ad agire in questo modo. L'Archeologia moderna ha dimostrato la quasi impossibilità di trovare un insieme adeguato di etichette semplici che possano descrivere tutte le situazioni riscontrabili nella storia umana. È interessante notare che si può riscontrare un elevato tasso di soggettività anche nei casi in cui delle tipologie siano state definite con cura; è un fatto abbastanza comune che archeologi differenti che lavorano con gli stessi dati definiscano la medesima realtà attraverso l'uso di categorie differenti.

### 3 Le Reti Neurali Artificiali in Archeologia

Nella parte restante del corso le Reti Neurali verranno utilizzate come una *learning machine* di tipo distribuito, adattivo e generalmente non-lineare per la risoluzione di problemi archeologici e per la rappresentazione di modelli di classificazione.

Una Rete Neurale Artificiale (*Artificial Neural Network, ANN*) rappresenta un paradigma di processamento dell'informazione ispirato alle modalità con le quali il cervello processa l'informazione. L'elemento chiave di questo paradigma è la struttura del tutto originale del sistema di processamento dell'informazione, che è composto da un elevato numero di elementi altamente interconnessi (neuroni) che lavorano all'unisono per la risoluzione di problemi specifici. Le Reti Neurali Artificiali, come le persone, apprendono attraverso esempi.

Un neurone naturale è costituito da due elementi, chiamati *sinapsi* e *dendrite* (Figura 3, sinistra). I dendriti sono le estensioni del neurone che lo connettono ad altri neuroni per formare una rete neurale, mentre le sinapsi sono delle porte di accesso che si connettono ai dendriti provenienti da altri neuroni. Attraverso queste connessioni vengono trasmessi impulsi elettrici e l'informazione viene trasportata secondo la temporizzazione e la frequenza con le quali questi impulsi vengono emessi. Un neurone riceve in questo modo informazione da altri neuroni, la processa e la rilascia a sua volta ad altri neuroni (Figura 3, destra). Il neurone deve pertanto generare un qualche genere di output che si basi sugli input che si accumulano. Il neurone integra gli impulsi che riceve e, quando questa integrazione eccede un certo limite, emette a sua volta un impulso.

Un neurone artificiale non è un robot, ma solamente un tipo di calcolo utilizzato per integrare algebricamente un certo numero di input, e produce solamente un singolo output numerico, equivalente alla somma pesata degli input. Gli input sono misure, ovvero elementi descrittivi di una qualche evidenza archeologica. L'output consiste in una assegnazione di classe, ossia nel risultato di un'operazione di classificazione o di *clustering*. Come i neuroni nel nostro cervello, i neuroni artificiali sono collegati l'uno con l'altro; anche qui non esiste un cavo fisico che connette i meccanismi virtuali. Il collegamento tra neuroni artificiali si traduce nell'esecuzione simultanea ed in parallelo di calcoli singoli.

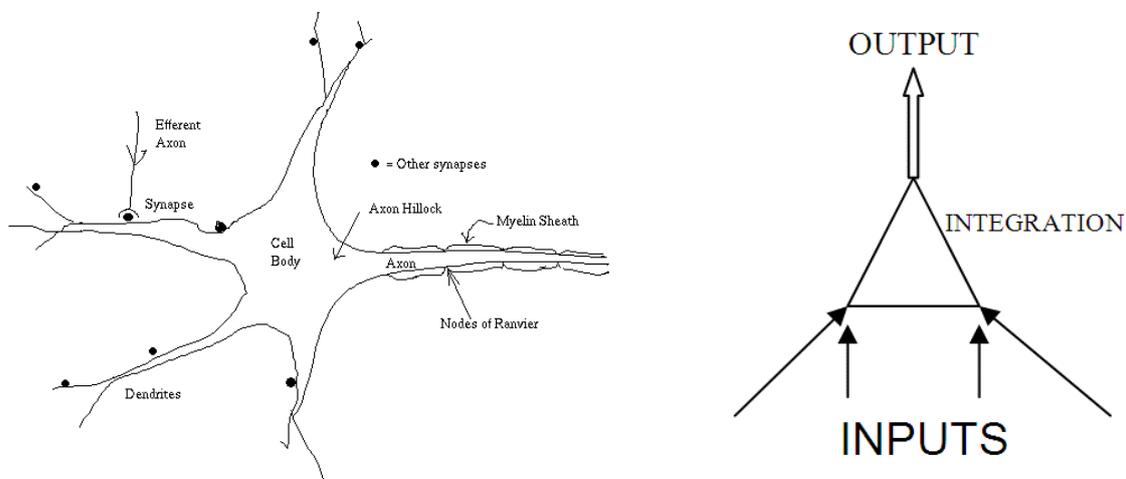
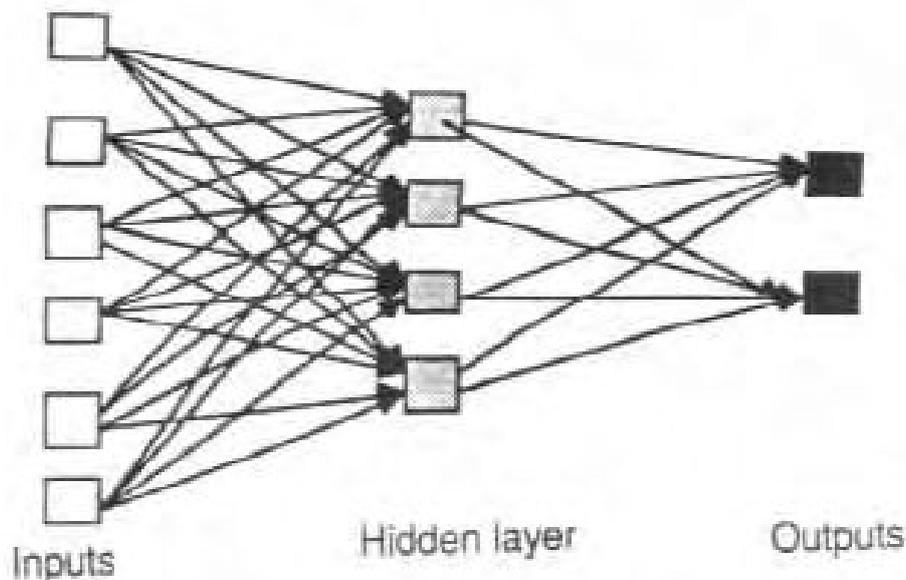


Figura 3 - Rappresentazione schematica di un neurone (a sinistra) e della sua attività (a destra)



**Figura 4 - La topologia di una Rete Neurale a tre strati, con uno strato nascosto**

Questo è ciò che chiamiamo l'*architettura o topologia della rete*. La topologia che studieremo non è, pertanto, quella di un cervello umano, ma una topologia di tipo semplice, progettata per la facile implementazione di calcoli paralleli su un computer digitale. Una delle forme più semplici di questa topologia è attualmente costituita da tre strati:

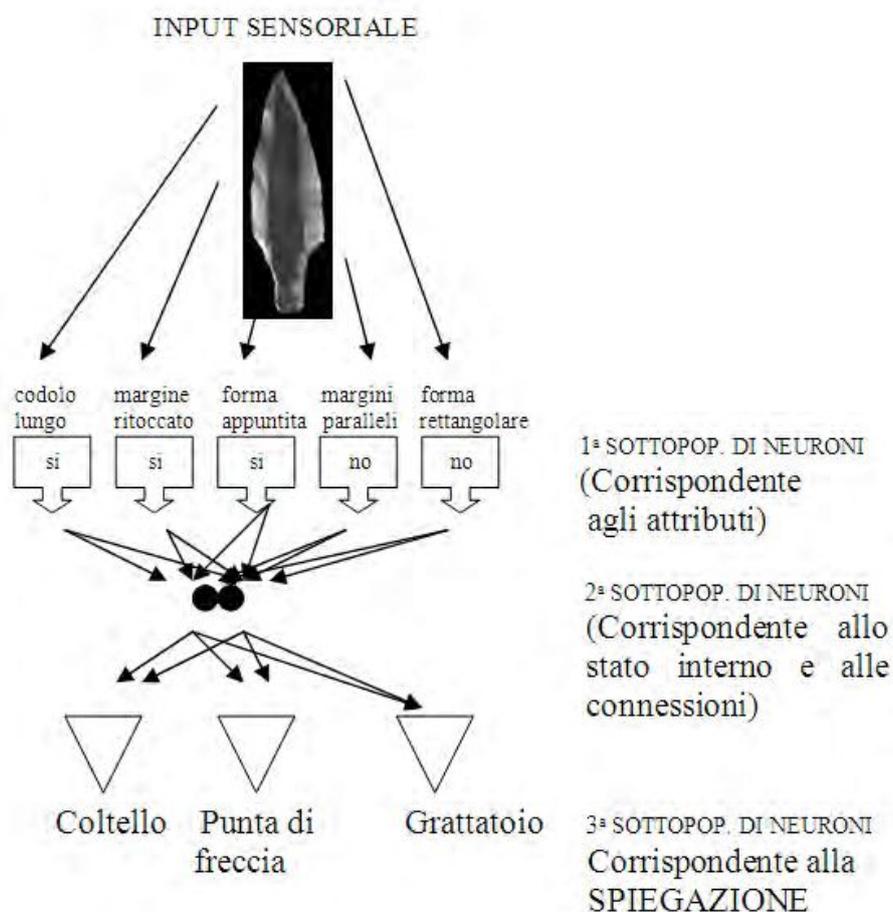
- uno *strato di input* (gli input della nostra rete)
- uno *strato nascosto*
- uno *strato di output* (l'output della nostra rete)

Tutti i neuroni che provengono da uno strato sono connessi con tutti i neuroni dello strato successivo (Figura 4).

Che cosa significa in sostanza tutto questo? Una Rete Neurale è semplicemente una complessa tecnica statistica che permette di associare una spiegazione (output) ad una descrizione (input). I pesi e le connessioni sono solamente i parametri di una equazione complessa che lega, generalmente in modo non-lineare, ogni elemento descrittivo con una categoria esplicativa di carattere generale.

È necessario ricordarsi che i neuroni sono solo calcoli. Lo scopo del sistema, pertanto, è solamente quello di definire un'equazione complessa che restituisca il valore dell'output come risultato di una somma pesata dell'input. In Archeologia i neuroni di input codificano quello che conosciamo, grazie all'osservazione, di ciascuno strumento, in un *vettore*, ovvero in una serie di numeri che indicano il valore di ciascun elemento. Abbiamo un vettore di input per ciascun esemplare nel database. Una seconda sotto-popolazione di neuroni può essere progettata per fornire un vettore che codifichi le possibili soluzioni di un problema: freccia, punta, raschiatoio ecc. (Figura 5).

Ci sono molti modi differenti per definire questa codifica: ogni *neurone di output* può essere una rappresentazione per ogni concetto o per ogni possibile soluzione, oppure ogni singolo concetto può essere rappresentato in maniera distribuita attraverso differenti neuroni.



**Figura 5 - Rappresentazione idealizzata di una rete neurale artificiale in Archeologia**

Traducendo tutto ciò in termini matematici possiamo dire che si ha l'applicazione di un insieme di valori descrittivi. Ogni valore individuale è moltiplicato secondo un peso – che è l'analogo della forza sinaptica – e tutti gli input pesati vengono sommati per determinare il livello di attivazione di un neurone intermedio, il cui scopo è di racchiudere la funzione matematica che lega l'input all'output.

Quando un numero giunge da una connessione, viene prima di tutto moltiplicato per il numero chiamato “peso della connessione” (*connection weight*) la cui funzione è quella di determinare la rilevanza alla connessione stessa (secondo un meccanismo identico a quello della grandezza di un dendrite biologico); successivamente il neurone accumula il risultato complessivo passando il valore attraverso una soglia (*threshold*), la quale emette un nuovo valore numerico quando un certo valore viene raggiunto.

L'output della fase di threshold è a sua volta connesso agli input di molti altri neuroni, che formano, quindi, una rete completa. L'attività di ciascuna unità nascosta è determinata dalle attività delle unità di input e dai pesi sulle connessioni tra l'input e le unità nascoste.

Il comportamento delle unità di output dipende dall'attività delle unità nascoste e dai pesi tra le unità nascoste e quelle di output. Questa semplice tipologia di rete è interessante, perché permette alle unità nascoste di essere libere di costruirsi le proprie rappresentazioni dell'input.

I pesi tra le unità nascoste e quelle di input determinano il momento in cui ciascuna unità nascosta è attiva; pertanto, modificando questi pesi, un'unità nascosta può scegliere che cosa rappresentare. Ogni rete neurale è dotata di conoscenza, contenuta nei valori dei pesi delle connessioni. Il cambiamento della conoscenza immagazzinata nella rete come funzione dell'esperienza implica una regola di apprendimento che determina il cambiamento dei valori dei pesi.

Una rete neurale artificiale può essere considerata come un sistema di processamento dell'informazione che crea una corrispondenza tra un vettore di *input*, che descrive un elemento, ed un vettore di *output*, che provvede all'assegnazione di una classe. Possiamo definire questo genere di approccio al ragionamento archeologico "*category-based*" poiché l'accesso alle proprietà della spiegazione avviene attraverso un processo di categorizzazione. Il vantaggio risiede nel fatto che le reti neurali sono adatte ad apprendere relazioni mal definite a partire da esempi poco chiari, includendo in questi anche relazioni fra tipi di dati differenti.

Una rete neurale è, di fatto, una memoria associativa; è un programma che traduce la realtà in numeri e che è in grado di fornire la risposta corretta quando interrogato utilizzando una descrizione appropriata. Si tratta di una situazione di *problem-solving* molto complessa perché se una memoria associativa si limitasse ad associare solo quegli specifici oggetti individuali che sono stati già visti in precedenza, il sottostante meccanismo di spiegazione archeologica automatica sarebbe di uso piuttosto limitato. Come archeologi, infatti, dobbiamo identificare un insieme di dati visuali nuovi tracciandone una corrispondenza con un determinato tipo di oggetto. La generalizzazione è solo una parte della nostra capacità nell'identificare oggetti; normalmente siamo in grado di identificare azioni sociali compiute nel passato anche in quei casi in cui la manifestazione visibile delle loro conseguenze materiali nel presente non coincide perfettamente con ciò che noi conosciamo a proposito di associazioni causa-effetto precedentemente memorizzate. La capacità di "riconoscere" dal punto di vista archeologico implica l'esistenza di una qualche forma precedente di apprendimento, nella quale le categorie esplicative, potenziali e astratte, siano state create e definite. L'obiettivo del riconoscimento consiste nell'attuare queste identificazioni in maniera corretta, in maniera tale cioè che l'identificazione rispecchi una proprietà significativa del mondo, indipendente dal particolare dato sottoposto ad interpretazione.

Di conseguenza una rete neurale opera con due modalità: quella di addestramento e quella di uso. Nella prima le connessioni possono essere addestrate ad attivare (o a non attivare) i neuroni connessi, in risposta a particolari pattern di input. Nella modalità di uso quando un pattern di input noto viene riconosciuto, l'output ad esso associato diviene l'output corrente.

Questo è esattamente l'attività di *Inverse Engineering* che abbiamo considerato precedentemente: dati valori appropriati di input (descrizioni) e di output (spiegazioni), dobbiamo calcolare i pesi che legano l'input e l'output. Qui risiede la principale differenza con un approccio statistico di tipo tradizionale. In una disciplina descrittiva, infatti, abbiamo a disposizione solo gli input; in questo caso non possiamo apprendere o generalizzare alcunché. Come accade nelle discipline sperimentali, abbiamo bisogno di conoscere la causa e l'effetto per essere poi in grado di inferire la connessione appropriata tra l'una e l'altro.

La metodologia di addestramento è analoga al modo in cui si insegnerebbe ad un bambino a leggere o a contare, mostrandogli alcuni numeri o lettere ed assegnando ad essi alcuni valori. Una rete neurale verrà addestrata esattamente nello stesso modo, fornendole un insieme di casi noti nei quali alcune descrizioni numeriche sono assegnate a categorie numeriche.

L'apprendimento implica l'esecuzione di calibrazioni tra le connessioni esistenti tra i neuroni, ossia il cambiamento dei valori dei loro rispettivi pesi. Quando alla rete vengono presentate istanze note delle categorie che devono essere apprese e si riscontra che l'output non è quello desiderato, che cosa farà la rete? La risposta è ovvia: modificherà alcuni pesi delle connessioni. Dal momento che i pesi delle connessioni nella rete sono inizialmente definiti casualmente, il risultato iniziale sicuramente non soddisferà alcun criterio di errore; un algoritmo appropriato modificherà, pertanto, il peso di alcune connessioni per ottenere un risultato migliore. L'algoritmo modificherà il peso delle connessioni finché non otterrà il risultato desiderato (questa è la fase di apprendimento). Successivamente si forniranno allo strato di input della rete altri esempi e si continuerà con la calibrazione dei pesi fino a che, finalmente, sarà ottenuto l'output atteso per ogni esempio. L'intero insieme di esempi di addestramento deve essere mostrato alla rete molte volte per poter raggiungere un risultato soddisfacente. Come non ci si aspetterebbe che un bambino imparasse a leggere dopo aver visto ogni lettera o parola una sola volta, così la rete ha bisogno di molti esempi. Il "valore di errore" viene computato come differenza tra il valore "corretto" atteso e il valore di output attuale. Questo valore di errore viene quindi "propagato" all'indietro (*backpropagated*) ripercorrendo la rete verso il basso e modificando i pesi in maniera proporzionale al contributo di ciascun elemento al valore totale di errore. Questo meccanismo viene ripetuto per ogni esempio appartenente all'insieme di apprendimento, mentre il rendimento sull'insieme di test migliora. La tecnica di addestramento citata viene definita "*error backpropagation*". Alla fine dell'apprendimento si confida nel fatto che la rete sia in grado di risolvere il problema; diciamo quindi che la rete ha appreso e che la sua "conoscenza" è immagazzinata dai differenti pesi delle connessioni. Di conseguenza, una rete neurale rappresenta un *modello di comportamento* costruito sulla base di un certo numero di esempi di questo comportamento. Il modello viene costruito attraverso l'*apprendimento*. La Rete Neurale, "ignorante" all'inizio, viene modificata utilizzando questi esempi e diviene un *modello* che rende conto del comportamento osservato (le relazioni tra le variabili descrittive contenute negli esempi).

Come conseguenza di questa computazione parallela, la rete neurale costruisce un sistema complesso di equazioni non lineari in grado di distinguere ogni differenza che esiste tra le differenti categorie di output. In effetti si tratta di un classificatore non lineare che può essere confrontato con la regressione non lineare o l'analisi discriminante. La rete neurale è in grado di individuare regole di classificazione anche quando le categorie di output si sovrappongono parzialmente oppure quando le tecniche statistiche standard non siano state in grado di trovare alcuna possibile discriminazione.

L'apprendimento per *backpropagation* si basa sull'assunzione che gli esempi di associazione tra dati e concetti siano in numero sufficiente. Ma da dove provengono questi dati? Le uniche possibilità che abbiamo sono l'osservazione controllata, l'etnoarcheologia o la replica sperimentale. Operando in questo modo saremo in grado di classificare un oggetto come membro di un ampio numero di categorie note sulla base delle sue proprietà visibili, quali la sua dimensione, la sua forma, la sua tessitura, la sua composizione e la sua localizzazione.

In secondo luogo questa identificazione ci permette di accedere ad un vasto insieme di informazioni già archiviate che riguardano questo particolare tipo di oggetto, includendo tra di esse la sua funzione e varie forme di aspettative circa i suoi comportamenti futuri. Questo schema a due fasi ha il vantaggio di associare ogni proprietà di tipo esplicativo con ogni oggetto.

Una rete neurale risolve problemi attraverso il riconoscimento e, con l'aiuto del risultato ottenuto, è in grado di effettuare ulteriori riconoscimenti. La risposta fornita dalla rete neurale esemplifica non solo lo stimolo o le sue sorgenti, ma anche le interazioni accumulate tra tutte le possibili sorgenti dello stesso stimolo o di stimoli simili, in maniera proporzionale alla frequenza con la quale essi sono stati percepiti.

Le reti neurali, con la loro notevole capacità di derivare un significato da dati complicati o imprecisi, possono essere utilizzate per estrarre pattern e individuare tendenze troppo complesse per essere identificate sia da esseri umani sia da altre tecniche informatiche. Una rete neurale addestrata può essere considerata alla stregua di un "esperto" della categoria di informazioni che le è stato chiesto di analizzare. Questo esperto può essere utilizzato per fornire proiezioni di fronte a nuove situazioni di interesse e per rispondere a domande di tipo *what if*. Ulteriori vantaggi sono i seguenti:

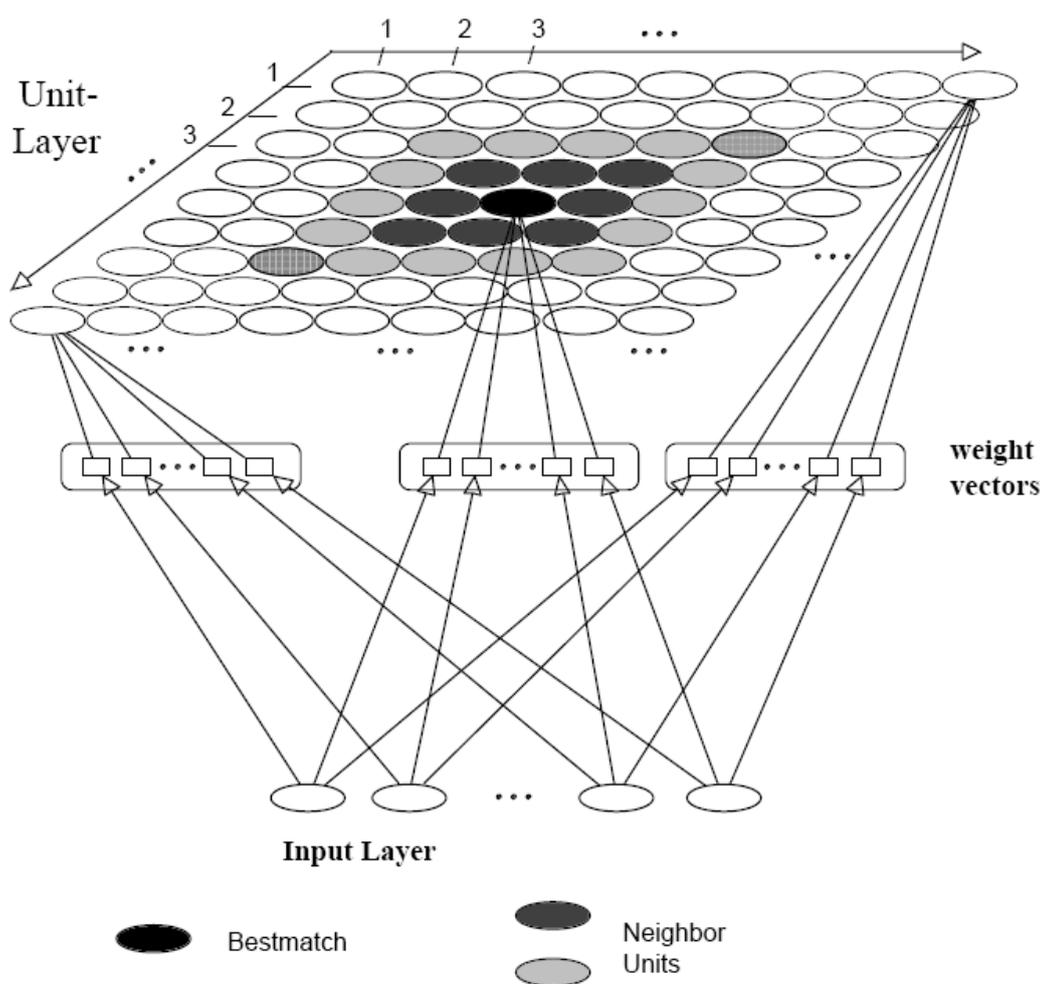
- quando un pattern precedentemente memorizzato (ossia un pattern "usuale") viene "visto" dal sistema, esso viene amplificato, rispondendo con una versione più forte del pattern di input;
- quando un pattern inconsueto viene "visto", esso viene attenuato, e la risposta della macchina viene inibita. Questo è un caso di risposta inconsueta;
- quando solo una parte di un pattern noto viene "visto", il sistema risponde "riempiendo" le parti mancanti. Questo meccanismo costituisce un tipo di paradigma di richiamo nel quale la parte costituisce un suggerimento per il *retrieval*, mentre l'azione di riempimento costituisce una sorta di processo di ricostruzione della memoria;
- quando un pattern simile ad uno di quelli archiviati viene individuato, il sistema risponde distorcendo il pattern di input verso quello archiviato. In questo caso abbiamo una sorta di risposta assimilativa, nella quale input simili vengono assimilati a eventi simili che sono stati memorizzati;
- infine, se un certo numero di pattern simili è stato archiviato, il sistema risponderà in maniera forte alla tendenza centrale del pattern archiviato, anche se la tendenza centrale stessa non è mai stata archiviata.

Si possono costruire architetture connessioniste di tipo più avanzato, così come si possono programmare algoritmi di apprendimento più complessi. Tutto questo indica che è corretto pensare alla risoluzione di problemi archeologici in termini di versioni speciali di problemi inversi. Risolveremo un problema archeologico, sia esso semplice o complesso, attraverso il calcolo della particolare connessione tra l'input descrittivo e l'output esplicativo. La forza computazionale delle reti neurali indica che questo problema inverso può essere risolto nella maggior parte dei casi ... ammesso che si abbiano a disposizione dati sufficienti sull'associazione causale che stiamo studiando.

Quando non disponiamo di questa conoscenza rimane comunque una possibilità, ossia quella di costruire uno spazio concettuale sulla base di un enorme numero di istanze non note del concetto. Non abbiamo bisogno di sapere come l'input è casualmente legato all'output. L'unica cosa di cui abbiamo bisogno è un insieme esaustivo di input. La qualità ed il valore esplicativo dello spazio concettuale dipenderanno dalla quantità, dalla varietà e dalla esaustività degli input che sono stati utilizzati per costruirlo. Ricordate quando Hodder o Binford hanno fatto riferimento a quadri di riferimento generale o a dizionari universali per i simboli umani [8][9]?

L'Intelligenza Artificiale e le Reti Neurali ci permettono di costruire una memoria associativa attraverso l'integrazione di molti frammenti differenti ed individuali di conoscenza.

Abbiamo pertanto bisogno di un tipo differente di Rete Neurale, ovvero di una memoria auto-organizzata senza supervisione (*Unsupervised Self-Organized Memory*). L'idea sembra essere quella di creare un modello concettuale attraverso l'aggregazione di molte esperienze. Il modello risponderà indicando che una nuova esperienza sembra essere simile ad una precedentemente sperimentata. Si tratta esattamente di un problema di partizione o di *clustering*, nel quale i dati di input servono a creare rappresentazioni geometriche dello spazio concettuale, all'interno del quale input simili attiveranno regioni simili. In questo caso la rete possiede solamente due strati: lo strato di input, che contiene la descrizione e le misure delle osservazioni reali, e lo strato di output, nel quale ogni neurone rappresenta una specifica regione dello spazio concettuale. Ogni neurone nello strato di output è connesso solamente ai suoi immediati vicini. Ogni neurone di input è connesso a tutti i neuroni nello strato di output e non è connesso a nessuno di quelli dello strato di input. Quanto più è complesso lo strato concettuale, tanti più neuroni di output sono necessari. Gli spazi concettuali saranno rappresentati in questo caso sottoforma di piani bidimensionali, sebbene si possano anche immaginare geometrie più complesse (Figura 6).



**Figura 6 - Una mappa auto-organizzata**

L'obiettivo consiste nel trasformare serie di valori di input nella corretta attivazione delle corrette regioni dello spazio concettuale. Dal punto di vista dello spazio di input, il *clustering* ne effettua una suddivisione in regioni locali, ognuna delle quali è associata ad una unità o neurone di output. Kohonen ha definito un algoritmo di competizione per ottenere questo obiettivo, in maniera tale che input simili attivino unità di output all'interno del medesimo vicinato.

L'algoritmo di *self-organizing mapping* di Kohonen [10] è così definito:

*Step 1.* Ogni neurone nel layer di Kohonen riceve una copia completa di un pattern di input.

*Step 2.* Trova il neurone vincente, che è quello con la minore distanza  $d_j$ .

$$d_j = \sum_{i=1}^m (X_i - W_{ij})^2)^{0.5}$$

*Step 3.* Per il neurone vincente e per quelli che gli sono fisicamente vicini viene utilizzata la seguente regola di apprendimento per modificare i pesi.

$$W_{ij}(t+1) = W_{ij}(t) + \alpha(t) * \gamma(t) * [X_i - W_{ij}(t)]$$

$$\gamma(t) = \exp\{-0.5 * [r_{ij}/\sigma(t)]^2\}$$

dove  $\alpha$  rappresenta la velocità di apprendimento, che decresce con il passare del tempo  $t$  (il valore iniziale è tra 0 e 1),  $r_{ij}$  è la distanza tra la cella vincente e la cella che viene aggiornata (questa distanza può essere calcolata tenendo conto, oppure no, dei bordi), mentre  $\sigma$  è il raggio di vicinanza che decresce con il passare del tempo.

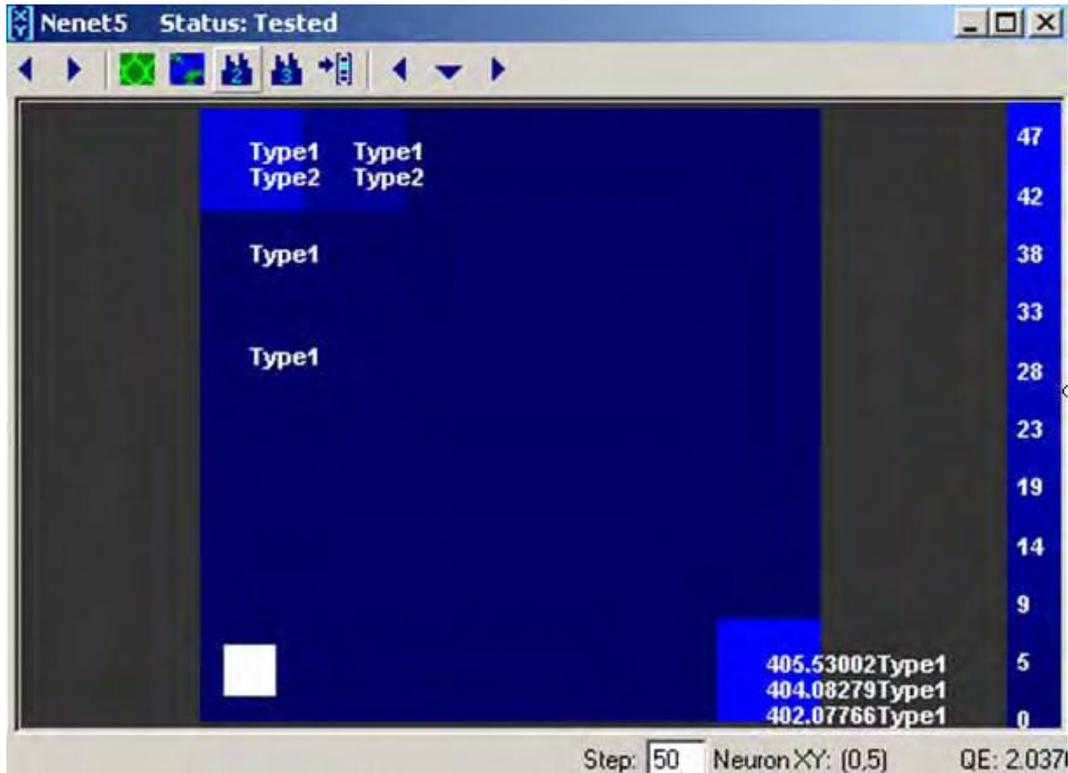
*Step 4.* Ripeti gli *Step 1-3* con un nuovo pattern di input.

*Step 5.* Ripeti lo *Step 4* fino a che tutti i pattern di input siano stati passati in rassegna (questo costituisce un ciclo di  $t$ ).

*Step 6.* Ripeti lo *Step 5* per un numero determinato di volte.

Per esemplificare, simuliamo un caso di studio. Stiamo esaminando 1900 punte di freccia ed abbiamo effettuato cinque differenti misurazioni su ciascuna di esse (lunghezza massima, lunghezza del peduncolo, larghezza massima, larghezza del peduncolo nella parte superiore, larghezza del peduncolo nella parte inferiore). Definiamo uno strato di output costituito da 6x6 unità, collegato attraverso vicinanza rettangolare. Ciò significa che ogni unità dello strato di output possiede un minimo di quattro vicini. I pesi sono inizializzati come valori casuali; effettuiamo l'addestramento della rete utilizzando le nostre 1900 punte. L'obiettivo è quello di dividere lo spazio di output in regioni locali, ognuna delle quali è associata ad un differente prototipo di punta di freccia.

Una volta addestrata, la rete può essere utilizzata per raggruppare dati ulteriori. Immaginiamo di avere a disposizione un quadro di riferimento tipologico di tipo tradizionale che vogliamo testare utilizzando le informazioni di *clustering* delle forme di punta di freccia. Abbiamo un nuovo insieme di dati con 96 punte di freccia del tipo 1. Testando la rete con questi dati, otteniamo una visualizzazione della localizzazione del *cluster* nello strato di output. Utilizzando 96 dati alternativi a questi e pertinenti a punte di freccia del tipo 2, otteniamo la mappa di Figura 7.



**Figura 7 - Test di una mappa auto-organizzata di Kohonen con dati archeologici**  
(software Nenet 1.1 <http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html>)

È facile vedere come i risultati pertinenti al tipo 2 siano simili a quelli del tipo 1 (sono all'interno dello stesso *cluster*), ma con una concentrazione maggiore. Utilizzando un approccio basato sull'attribuzione di etichette, possiamo osservare il grado di sovrapposizione delle categorie tradizionali con lo strato di output. Una mappa auto-organizzante è pertanto in grado di *individuare regolarità (regularity detection)* all'interno delle quali ogni unità apprende come rispondere a determinate proprietà dei pattern di input. Mentre nel mapping associativo la rete memorizza le relazioni tra i pattern, nella *regularity detection* la risposta di ciascuna unità ha un particolare "significato". La rete semplicemente deduce che il nuovo input è simile ad alcuni output precedentemente presi in esame. Se si interroga la rete concettuale con esempi appropriati, si capisce come questa sia organizzata.

Immaginiamo un'applicazione per le scienze sociali nella quale alcuni indici sociali siano utilizzati come input. Lo spazio concettuale è una rappresentazione dello spazio sociale. Come possiamo comprenderne la struttura? Attraverso la formulazione di domande intenzionali, ossia fornendo alla rete, per esempio, le caratteristiche sociali delle donne povere, dei lavoratori immigrati, ecc. Siamo in grado di comprendere lo spazio sociale indagando le distanze e le similarità tra individui sociali differenti.

#### 4 Conclusioni

Le Reti Neurali presentano due principali vantaggi:

- la capacità di *rappresentare qualsiasi funzione*, sia essa lineare oppure no, semplice o complicata. Le reti neurali sono ciò che i matematici definiscono "approssimatori universali";
- la capacità di *apprendere a partire da esempi caratteristici*.

La rete neurale fornisce risposte, non spiegazioni. Il modello neurale, infatti, incorpora correlazioni (come le associazioni di tipo intuitivo), non relazioni causali (spiegazioni). L'esame di una rete neurale di per sé ci mostra solo valori numerici privi di significato. Il modello neurale è una "black box".

D'altra parte questo modello, in virtù del fatto che è continuo e derivabile, si presta ad essere "esplorato", oltre che semplicemente interrogato, per determinare profili tipici o per valutare la forza esplicativa di ciascuna variabile, per definire una nuova classificazione di insiemi di esempi in modo da determinare la probabilità ad essi associata (ottenendo una "curva di Gini"), per visualizzare dati e predizioni.

Costruire un modello neurale, tuttavia, è un compito che appartiene all'analisi dei dati, non alla magia (anche se, per citare Arthur C. Clarke, una tecnologia sufficientemente avanzata non è distinguibile dalla magia). I dati devono essere esplicativi ed in quantità sufficientemente ampia.

Quando utilizziamo le reti neurali in funzione di strumenti di *inverse engineering* e di *problem solving* per la ricerca archeologica, otteniamo:

- la capacità di riconoscere elementi o *pattern* pure in presenza di rumore e distorsione o qualora siano presenti informazioni solamente parziali
- la capacità di individuare analogie complesse
- la capacità di richiamare istantaneamente informazioni rilevanti qualora siano in relazione a circostanze del tutto nuove
- la capacità di porre l'attenzione su diversi elementi dei dati empirici
- la capacità di testare concettualizzazioni differenti di fronte ad una situazione problematica
- la capacità di riconoscere proprietà empiriche tenui o vaghe.

## 5 Bibliografia

- [1] <http://foldoc.org/?query=computer+vision>
- [2] <http://foldoc.org/foldoc.cgi?natural+language+processing>
- [3] ANDERSON, J.R., 1980, *Cognitive Psychology and its Implications*, New York, W.H. Freeman.
- [4] JACKSON, K.F., 1983, *The Art of Solving Problems: Bulmershe-Comino Problem-Solving Project*, Bulmershe College, Reading (UK).
- [5] WHITE, R., 2003, The Epistemic Advantage of Prediction over Accomodation, *Mind*, 112, pp. 653-683.
- [6] PERLOVSKY, L.I., 2001, *Neural Networks and Intellect. Using Model-based Concepts*, Oxford University Press, New York.
- [7] GURECKIS, T.M., LOVE, B.C., 2003, Human Unsupervised and Supervised Learning as a Quantitative Distinction, *International Journal of Pattern Recognition and Artificial Intelligence*, 17 (5), 885-901.
- [8] HODDER, I., HUTSON, S., 2004, *Reading the Past: Current Approaches to Interpretation in Archaeology*. Cambridge University Press. Cambridge (UK).
- [9] BINFORD, L.R., 1989, *Debating Archaeology*. Academic press. New York.
- [10] KOHONEN, T., 2001, *Self-Organizing Maps* (Third Edition), Springer, Berlin.