

Sistemi per il recupero delle informazioni

Gabriele Pozzani

A.A. 2012/2013

**Corso di Laurea Magistrale in
Editoria e Giornalismo**

Sistemi per il recupero delle informazioni multimediali

Multimedia

- Un'informazione multimediale è una qualunque informazione digitale
 - Testo
 - Audio
 - Video
 - ...
- Solitamente poco o non strutturata

3

Recupero delle informazioni multimediali

- Consiste nel recupero di testo, audio, video in risposta ad un'interrogazione di un utente e il loro ordinamento in base ad un qualche grado di rilevanza
- Come può avvenire una ricerca multimediale?
 - Descrivendo il suo contenuto
 - tecnicamente, simbolicamente, semanticamente (?)
 - La semantica è esterna al dato e la sua definizione dipende dal media
 - tutta l'immagine, l'audio o alcuni componenti
 - Fornendo un esempio dell'informazione desiderata
 - Immagine/audio simile
 - proprietà simili (sketch)
 - Attraverso i metadati

4

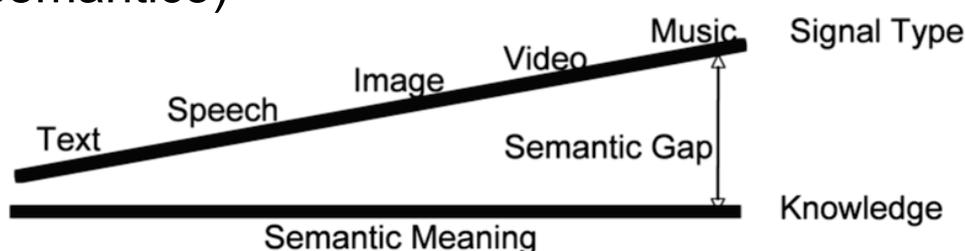
Caratteristiche dei sistemi di information retrieval multimediali

- Informazioni
 - di natura semplice ma di significato complesso
 - il concetto di struttura è variabile e dipende dal media (immagine, filmato, audio)
 - elevate dimensioni per ogni esemplare di dato
- Interrogazioni
 - scarsa corrispondenza tra forma e significato
 - specifiche "sintattiche", ricerca per somiglianza
 - iterazioni successive, analisi di rilevanza
- Aggiornamenti
 - non frequenti, generalmente fuori linea
- Problemi specifici
 - la rappresentazione codificata non contiene il significato del dato
 - la codifica e la rappresentazione influiscono sui sistemi di gestione

5

Semantica

- Il principale problema con i dati multimediali è l'associazione ad essi di una semantica
 - i dati multimediali hanno un contenuto informativo molto ricco ma parte della informazione non è esplicita e può essere estratta solo analizzando un contesto più ampio
 - Vi è un gap tra il dato e la sua semantica (gap semantico)



6

Gap semantico (I)

- Un problema fondamentale è il riconoscimento degli oggetti in immagini, video, e audio
 - Non solo trovarli, ma anche interpretarli!!!
 - Una persona può guardare un'immagine e identificare facilmente visi ed altri oggetti qualunque sia la loro forma, colore, ecc...
 - Alcuni contenuti sono soggettivi ed “emozionali”
 - La velocità e il tono di un parlato
 - L'espressione di un viso
 - È un problema in generale ancora irrisolto

7

Gap semantico (II)

- Attualmente i SRI multimediali usano testi prodotti da esseri umani ed associati ai dati multimediali
 - Versioni iniziali e base di Google Images, Bing, ecc...

The image shows a side-by-side comparison of Google and Bing image search results for the query "leonardo da vinci". On the left is the Google search interface, displaying a grid of various images related to Leonardo da Vinci, including portraits and his famous drawings. On the right is the Bing search interface, also showing a grid of images. A blue callout bubble with white text points to the search results, stating: "leonardo da vinci" è ricercato nel testo associato all'immagine. This highlights the semantic gap where the text label is not directly associated with the visual content of the images.

8

Gap semantico (III)

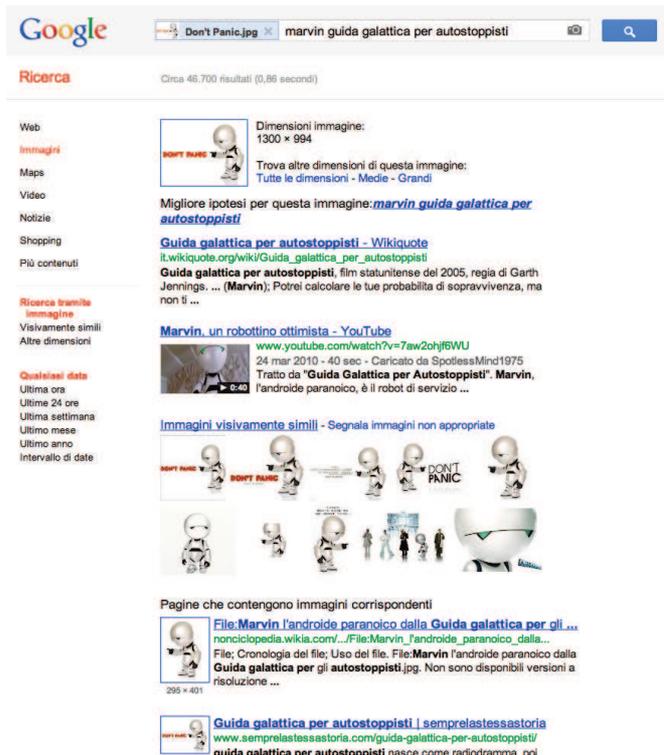
- Il contenuto è poco o per nulla usato
- Sono possibili solo alcuni filtri su caratteristiche tecniche

The image shows a Google search results page for 'leonardo da vinci'. The search bar at the top contains the text 'leonardo da vinci'. Below the search bar, there are several filter categories: 'Web', 'Immagini', 'Maps', 'Video', 'Notizie', 'Shopping', 'Libri', and 'Più contenuti'. The 'Immagini' filter is selected, and a grid of image results is displayed. Overlaid on the image are several blue speech bubbles with white text, pointing to specific filter options: 'Dimensioni' (Dimensions) points to the 'Qualsiasi dimensione' filter; 'Durata' (Duration) points to the 'LENGTH' filter; 'Risoluzione' (Resolution) points to the 'RESOLUTION' filter; 'Colore' (Color) points to the 'Qualsiasi colore' filter; and 'Tipo' (Type) points to the 'Qualsiasi tipo' filter. The search results page also shows a 'bing' search bar and a 'SafeSearch' filter.

Content-based image retrieval (I)

- Ricerca di immagini basandosi sul contenuto
 - Non il contenuto semantico
 - Ma le caratteristiche del contenuto
 - Colore e sua distribuzione
 - Forma
 - Luminosità
 - Contorni
 - L'analisi di queste caratteristiche permette anche di
 - Catalogare le parti di un'immagine (sfondo, primo-piano, ...)
 - Identificare e taggare singoli oggetti perché se ne conoscono le proprietà fondamentali
- Query-by-example
 - Viene fornita un'immagine di riferimento e si cercano immagini "simili"

Content-based image retrieval (II)



- Google permette il query-by-example
- Analizza e ricerca
 - Colore
 - Linee
 - Punti di riferimento
 - Texture
- Ritorna immagini con caratteristiche simili
- Analizza le pagine in cui tali immagini si trovano per
 - associare una descrizione testuale usata per cercare documenti “interessanti”

11

Content-based image retrieval (III)

- QBIC
 - Sviluppato da IBM
 - Abbandonato
 - Usato dall'Hermitage di San Pietroburgo per la ricerca di opere d'arte sulla base di
 - Colore
 - Layout



12

Content-based image retrieval (IV)

- Le tecniche di analisi di immagini e video e il loro recupero basato sul contenuto sono ancora ampiamente in sviluppo
 - Esistono ampie aree di ricerca
 - Le tecniche sono basate su avanzate tecnologie informatiche e matematiche
 - Anche a causa dell'enorme numero e mole di informazioni visive presenti sul web
 - Sono richieste efficienza ed efficacia

13

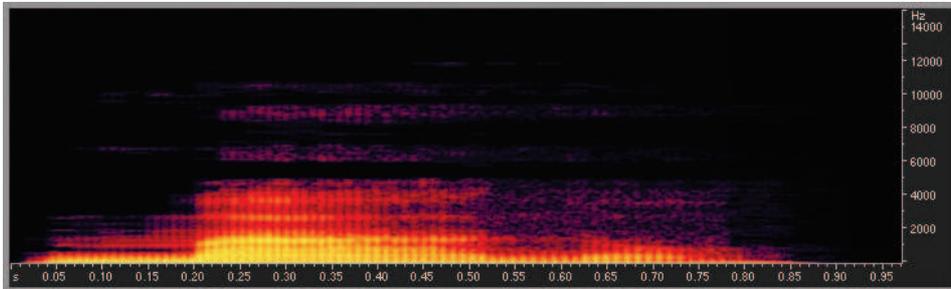
Recupero di audio e musica

- Recupero di tracce audio che risultino rilevanti rispetto ad una richiesta
 - Fingerprinting: ricerca di tracce audio a partire da un breve frammento audio (query-by-example, query-by-humming)
 - Riconoscimento del parlato: riconoscere il testo di una traccia audio contenente del parlato
 - Riconoscimento dello speaker: riconoscere le persone che parlano in una traccia audio
 - Recupero di tracce parlate: ricerca di tracce parlate rispetto ad una query fornita testualmente

14

Fingerprinting

- Applicazione di successo
- Basata sull'analisi dello spettrogramma delle tracce audio



- Complicata da eventuali rumori e distorsioni nelle tracce
- Il software più famoso è probabilmente Shazam
 - Applicazione per dispositivi mobili
 - A partire da un frammento audio di 10s ne costruisce il fingerprint e lo ricerca in DB centralizzato contenente i fingerprint di milioni di canzoni

15

Riconoscimento del parlato

- Il riconoscere le parole in una traccia audio
- Applicazioni tipiche e funzionanti sono
 - Apple Siri Personal Assistant
 - Google Voice Search e Google Voice Actions
 - Bing voice search
 - Microsoft Tellme for Windows Phone 7
- Una volta riconosciute le parole, queste possono essere usate per una ricerca o per eseguire azioni

16

Riconoscimento dello speaker

- Riconoscere chi sta parlando in una traccia senza riguardo a cosa sta dicendo
- Basato su diversi approcci matematici, tra cui
 - Speaker-dependent speech recognition: modellare e riconoscere le peculiarità nella pronuncia di ogni persona

17

Recupero di tracce parlate

- Recuperare le tracce audio che rispettano una query testuale
- Due diversi approcci basati su
 - Keyword
 - Riconoscere alcune parole chiave nell'audio
 - Limita l'utente ad usare un insieme limitato di termini nella query
 - Riconoscimento fonetico
 - Riconoscere il parlato a livello di fonemi
 - Complicato dalla similarità di alcuni fonemi

18