

# Sistemi di ricerca

*Corso di Web Design - F. Pittarello*

*Nota: il materiale contenuto in questo documento è disponibile solo per uso interno nell'ambito del corso.*

## Quando implementare un sistema di ricerca?

- Quando è opportuno implementare un sistema di ricerca?
  - Quando c'è molta informazione da percorrere attraverso sistemi di browsing
  - Quando è costituito da componenti frammentati (es. siti distinti per unità di business sviluppate indipendentemente)
  - Quando si ha bisogno di imparare dagli utenti dall'analisi dei search-logs
  - Quando gli utenti si aspettano che ci sia
  - Quando il sito è molto dinamico

# Zone di ricerca

---

- Indicizzare tutto il contenuto di un sito per permettere la ricerca non è sempre opportuno
- La creazione di search-zones riduce l'effetto **apple-and-oranges**, cioè la ricerca effettuata su aree eterogenee di contenuti che può essere un ostacolo al ritrovamento dell'informazione.



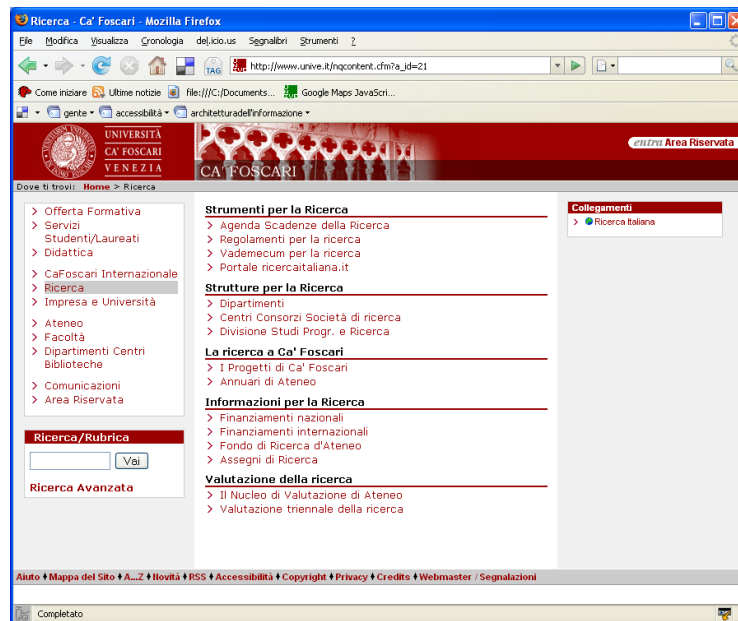
- Le zone ricerca (searching zones)
  - Sono parti di sito che sono state indicizzate separatamente per ottenere migliori risultati nel reperimento dell'informazione
  - Si possono creare suddividendo fisicamente o logicamente i documenti
  - Gli schemi organizzativi e le strutture organizzative possono essere la base per la creazione di zone di ricerca
  - Attenzione ... è un'arma a doppio taglio, si aggiunge un livello di complessità alla ricerca

# Pagine di destinazione e di navigazione

---

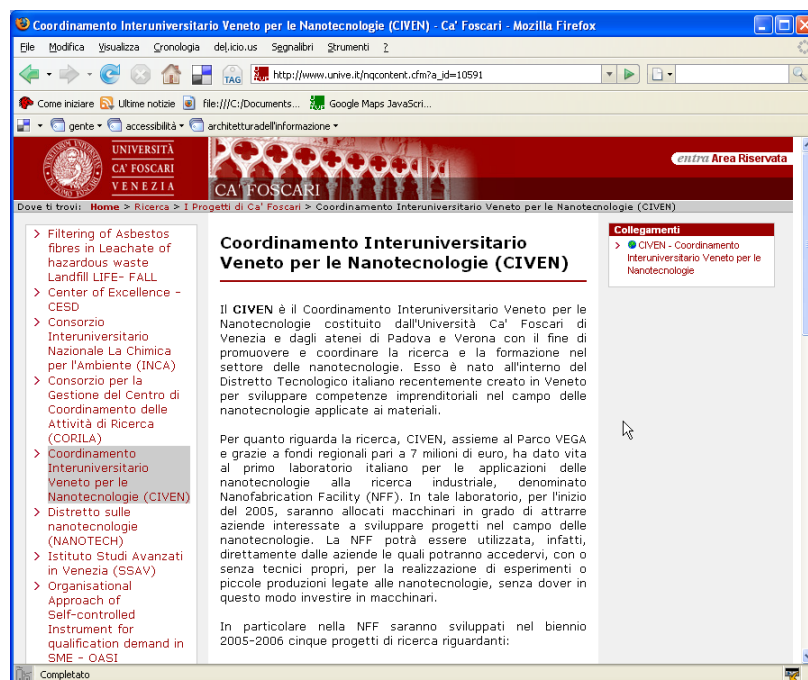
- Molti siti contengono almeno due categorie di pagine:
  - **Pagine di destinazione**, che ospitano il contenuto vero e proprio
  - **Pagine di navigazione**, come la home page, il cui scopo primario è condurre alle pagine di destinazione
- In alcuni casi le pagine appartengono ad entrambe le categorie
- Nel caso in cui le pagine appartengano esclusivamente alla seconda categoria è bene escluderle dall'insieme di pagine da indicizzare

# Pagine di destinazione e di navigazione



- [www.unive.it](http://www.unive.it), pagina di navigazione

# Pagine di destinazione e di navigazione



- [www.unive.it](http://www.unive.it), pagina di destinazione

## Componenti di una pagina

---

- La chiara identificazione dei diversi elementi della pagina permette di selezionare gli elementi da indicizzare o meno ad un livello di granularità più accurato rispetto a quello della pagina
- L'avvento di sistemi di gestione dei contenuti (Content Management System) rende più facile l'identificazione di questi elementi (es. il contenuto informativo, i menu di navigazione, la pubblicità, i disclaimers, ecc.), e la loro inclusione o esclusione dall'indicizzazione.

## Presentare i risultati - Quali

---

- Quali componenti del contenuto vanno mostrati?
  - Una semplice linea guida è quella di mostrare come risultato all'utente che sa quello che sta cercando (**known-item search**) meno informazione rispetto a quella presentata all'utente che non sa quello che sta cercando (**exploratory search**)
  - Agli utenti che sanno quello che stanno cercando saranno sufficienti alcuni elementi rappresentativi del contenuto, come il **titolo** o l'**autore**
  - La seconda categoria di utenti beneficerà di contenuti descrittivi, come un **sommario**
  - In alcuni casi all'utente viene consentito di **scegliere quali elementi visualizzare** nella risposta ad una ricerca; utile quando l'utente in momenti diversi svolge sia known-item searches che exploratory searches

# Presentare i risultati - Quanti

---

- **Quanti** documenti vanno mostrati?
  - Dipende da due fattori principali:
    - Dalla **quantità di componenti informative** visualizzate per ogni componente
    - **Dalle caratteristiche del dispositivo** che si utilizza (la risoluzione del monitor, la velocità di connessione)
  - E' opportuno partire dalla **presentazione di un gruppo limitato di documenti** (es. 10 per pagina) lasciando all'utente la libertà di configurare il sistema per avere un maggior numero di risultati
  - In ogni caso è opportuno visualizzare il numero complessivo di documenti ritrovati e di fornire un sistema di browsing per permettere una rapida navigazione tra i risultati

# Elencare i risultati

---

- In quale ordine devono essere presentati i risultati all'utente? Esistono due metodi:
  - **Sorting**: i risultati vengono ordinati utilizzando una delle componenti del documento (es. la data di creazione o modifica, l'iniziale alfabetica, ecc.); la tecnica è utile quando l'utente deve prendere una decisione rispetto ad una componente specifica (es. il prezzo)
  - **Ranking**: i risultati vengono ordinati utilizzando criteri come la rilevanza del documento, dedotta da un'analisi dei contenuti, o la sua popolarità; la tecnica è utile quando l'utente ha bisogno di capire o imparare qualcosa

# Elencare i risultati

- Ranking per rilevanza
  - Gli algoritmi di ranking solitamente considerano:
    - Quanti termini della query si trovano nel documento
    - quante volte i termini della query vengono trovati nel documento
    - La vicinanza dei termini (es. adiacenti, nella stessa frase o paragrafo)
    - Dove si trovano questi termini (es. nel titolo, nelle intestazioni, ecc.)

**Terrorism** remains a major preoccupation of Americans. Accordingly, the *Encyclopedia of Terrorism* would seem to be a timely addition to a library reference shelf, especially if it is up to date. The cover of the print volume features a photograph of the burning World Trade Center towers, following the September 11th attack on them. Final materials for this encyclopedia appear to have been prepared and sent off to the printers in December 2001. Accordingly, quite a few entries address aspects of the September 11th attack, including a 27-page section on "U.S. and International Reactions to September 11, 2001, Day by Day." On the other hand, the aftermath of September 11th, and American responses to it, is an evolving story, and, in this sense, some of the entries become dated. For many purposes, then, those interested in this story are more likely to turn to the media and the Internet, rather than an encyclopedia. The authors of this encyclopedia (and they are billed as authors, not editors) have each published books on terrorism. Eleven other contributors are listed, but only a few are identified as one of the primary authors at the University of North Carolina at Charlotte. Readers are only provided with their disciplinary or departmental affiliations, and it is unclear whether they have any expertise in topics related to terrorism. No bylines (or even author initials) are listed with any of the entries. This book opens with a brief introductory essay, but the bulk of the book is devoted to conceptual and defining controversies surrounding the term "terrorism." They state that the objective of this encyclopedia is to stress contemporary forms of terrorism, with entries providing a brief description and analysis of the topic at hand. Although they note that the encyclopedia is necessarily selective, they do not provide any criteria. Most of the entries in this encyclopedia are quite short, with a few being more descriptive over in-depth analysis. For many of the briefer entries, no bibliographic references are provided; where such bibliographic references are provided they might include a book, several articles, and one or more Websites. Cross-references are also included for some of the entries. In at least some cases, the length of the entry seems to be dictated more by the special interest or knowledge of a contributor, rather than the relative importance of the phenomenon or event. For example, a long entry is included on the somewhat obscure Wilmington, North Carolina Coup and Massacre of 1898 and the entry on animal rights organizations is much longer than the entry on anarchism. Furthermore, the entries on Josef Stalin and Adolf Hitler are quite brief and superficial. At least some of these allocations of space seem to be somewhat arbitrary. The entries in this encyclopedia seem to fall principally in the following categories: terrorist organizations; terrorist incidents; terrorist modes of operation; forms of response to terrorism; key individuals in the history of terrorism; countries, in relation to their history as sites of terrorism, perpetrators of terrorism, or responders to terrorism; and conventions, accords, and treaties relating to terrorism. This encyclopedia commendably addresses state terrorism or terrorism carried out on behalf of, or in the name of, states. This major form of terrorism which by many measures is the most consequential is often neglected in discussions of terrorism. On the other hand, some questionable entries are included (e.g., the one on Charles Manson). This encyclopedia closes with appendices providing some charts and tables relating to aspects of terrorism, a chronological listing of major acts of terrorism from 1941 to 2000, and a chronology of the U.S. and international responses to September 11th through the end of November 2001. A selective bibliography and an index are also included. Altogether, this encyclopedia should certainly be helpful to those seeking some basic, timely information about terrorism, and as a starting point for more substantial research on aspects of terrorism.

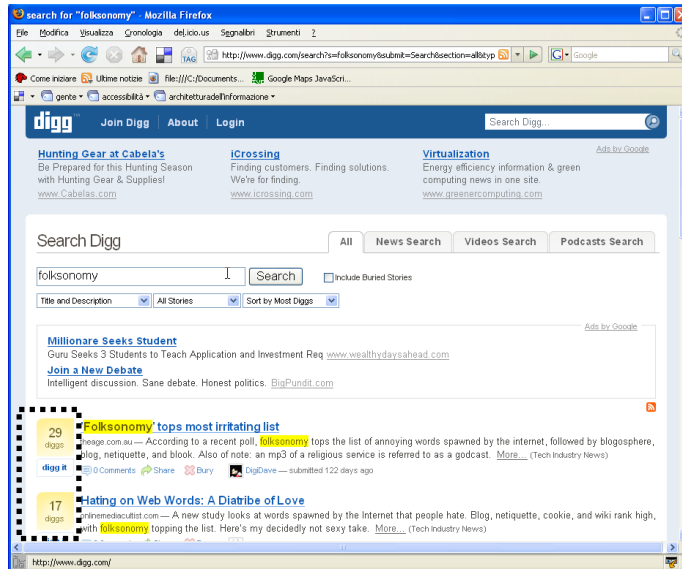
# Elencare i risultati

- Ranking per popolarità
  - La popolarità è la fonte della popolarità di Google
    - Google utilizza un algoritmo denominato PageRank, nel quale si considerano il numero di link che puntano da un sito esterno ad un determinato documento per determinarne la popolarità;
    - ... ma per calcolare la rilevanza Google usa oltre altri 100 criteri in aggiunta a PageRank



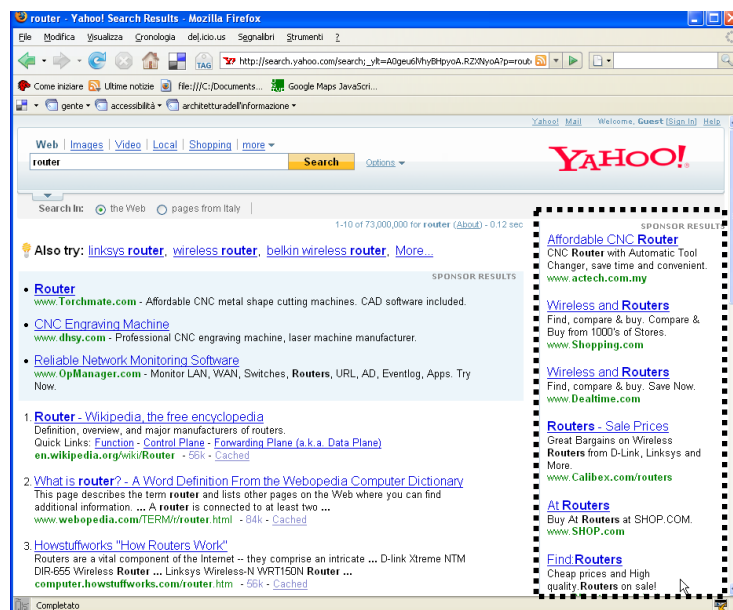
# Elencare i risultati

- Ranking attraverso il giudizio di utenti o di esperti
  - In questo caso gli utenti (o gli esperti) valutano il valore dell'informazione



# Elencare i risultati

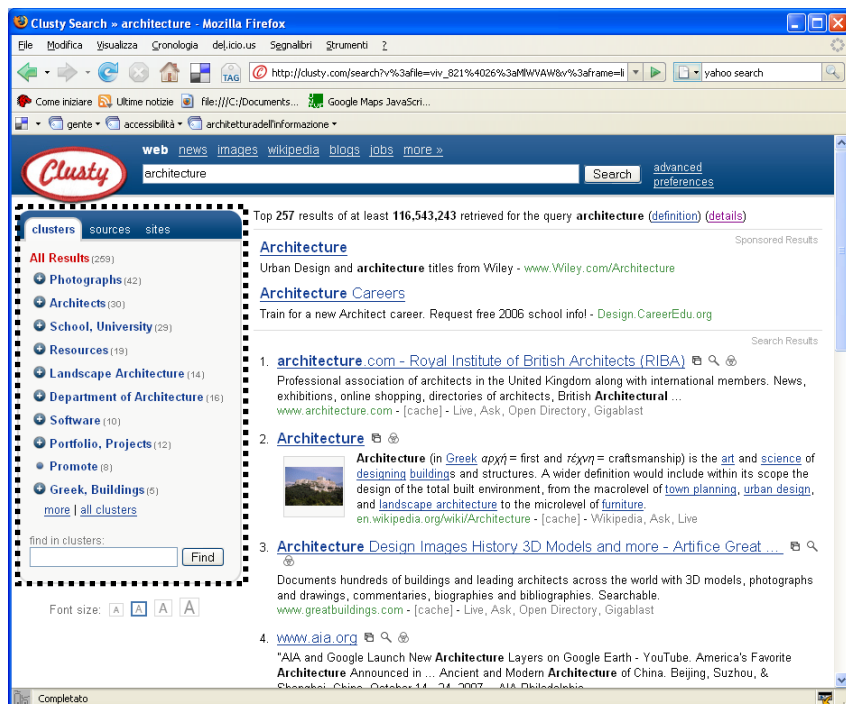
- Ranking attraverso il modello pay-for-placement
  - Somiglianza con gli annunci evidenziati nelle pagine gialle



# Raggruppare i risultati

- Una tecnica alternativa o complementare al sorting e al ranking consiste nel raggruppare i risultati in base a qualche aspetto comune
- Come si possono raggruppare i risultati?
  - Attraverso l'uso di metadati esistenti (come il tipo di documento, la data di creazione/modifica)
  - Attraverso l'uso di metadati applicati manualmente (coe l'audience, l'argomento, ecc.)
  - Dal momento che l'applicazione manuale di metadati è molto dispendiosa, alcuni tools cercano di derivare automaticamente alcuni cluster considerando i contesti possibili a partire da una keyword inserita dall'utente

# Raggruppare i risultati



clusty.com



# Esportare i risultati

---

- La possibilità di stampare, inviare via e-mail o salvare i risultati è un utile complemento per gli utenti che hanno trovato un'informazione e vogliono conservarla o condividerla
- La selezione di un sottoinsieme di risultati può essere utile quando si vuole salvare un insieme di elementi interessanti (anziché uno)
- Il salvataggio di una ricerca rappresenta un'alternativa utile per i domini dinamici, per i quali la riesecuzione dell'interrogazione in un altro momento può portare a nuovi risultati

# Esportare i risultati

---

Google Alerts Manage your Alerts [Settings](#) | [FAQ](#) | [Sign out](#)

Create a Google Alert Sending HTML emails. [Switch to text emails.](#)

Search terms	Type	How often	
<input type="text" value="Google Guide"/>	<input type="text" value="Groups"/>	<input type="text" value="as-it-happens"/>	<input type="button" value="Create Alert"/>

Your Google Alerts

Search terms	Type	How often	
<a href="#">"bird flu" site:whyfiles.org</a>	Web	once a week	<a href="#">edit</a>   <a href="#">delete</a>
<a href="#">"Google Guide"</a>	News & Web	as-it-happens	<a href="#">edit</a>   <a href="#">delete</a>
<a href="#">Uzbekistan</a>	News	once a day	<a href="#">edit</a>   <a href="#">delete</a>

- Google Alerts

# L'interfaccia di ricerca

---

- Il modulo per la ricerca
  - Mantenere l'**interfaccia di ricerca iniziale più semplice possibile**
  - Alcune assunzioni comuni dell'utente
    - Non è necessario che utilizzi operatori logici come AND, OR oppure NOT
    - Non si deve preoccupare di utilizzare un vocabolario controllato
    - La ricerca viene fatta sull'intero sito
  - Educare l'utente quando è pronto ad essere educato
    - ... quando la prima ricerca è stata fatta e l'utente ha raggiunto un punto di indecisione o frustrazione
    - Inserendo un collegamento ad una pagina di consigli (tips)
    - Inserendo la possibilità di restringere la ricerca ad una search-zone
    - Inserendo un collegamento ad un modulo di ricerca avanzata (è possibile dire di no a questa richiesta? Probabilmente no, ma le complicazioni che lo caratterizzano dovrebbe suggerire al designer di cercare di renderlo non necessario per la maggior parte degli utenti)

# L'interfaccia per i risultati

---

- **Supporto alla revisione**
  - Ripetizione del modulo di ricerca compilato nella pagina dei risultati, per permettere all'utente di modificarlo e di eseguirlo senza inserire tutto da zero
  - Spiegazione da dove vengono i risultati (ad es. se vengono da una search-zone specifica del sito)

# L'interfaccia per i risultati

---

- **Supporto alla revisione**

- Spiegazione delle caratteristiche della ricerca e della presentazione dei risultati
  - Descrivere gli eventuali filtri applicati alla ricerca
  - Mostrare gli operatori booleani e di altro tipo implicitamente applicati alla ricerca
  - Mostrare tutte le caratteristiche utilizzate per la presentazione, come il tipo di sorting utilizzato per i risultati
  - Evidenziare il numero dei risultati ottenuti dalla ricerca

## Quando gli utenti vengono sopraffatti ...

---



- Che cosa fare quando i risultati della ricerca sono nulli oppure ci sono moltissimi risultati?
- Nel primo caso **evitare dead-end policies** fornendo:
  - Strumenti per revisionare la ricerca
  - Istruzioni per l'uso del sistema di ricerca
  - Collegamenti ai sistemi di navigazione del sito
  - Un contatto con un umano, se il resto non funziona ....
- Nel secondo caso è possibile fornire:
  - Strumenti per restringere la zona di ricerca
  - Permettere ricerche più specifiche all'interno dell'insieme di documenti ritrovati con la prima ricerca (search within)

# Interfacce di ricerca basate sulla classificazione sociale

- Anche se molto spesso le risorse associate a tags vengono trovate attraverso la navigazione (es. attraverso tag clouds), in alcune situazioni sono disponibili **sistemi di ricerca** dei tags.
- Questo succede generalmente per le situazioni in cui i tags costituiscono il sistema prevalente di classificazione dell'informazione o comunque nei sistemi caratterizzati dalla presenza di molti tags associati alle risorse.
- Importante: è importante, nella presentazione delle form di ricerca per tags, rendere consapevole l'utente che **la ricerca** delle informazioni in questo caso **avviene sull'insieme dei tags e non sull'insieme delle parole che compongono la pagina web** (come avviene invece per i sistemi di ricerca, interni o esterni al sito, che indicizzano tutti i contenuti delle pagine web).

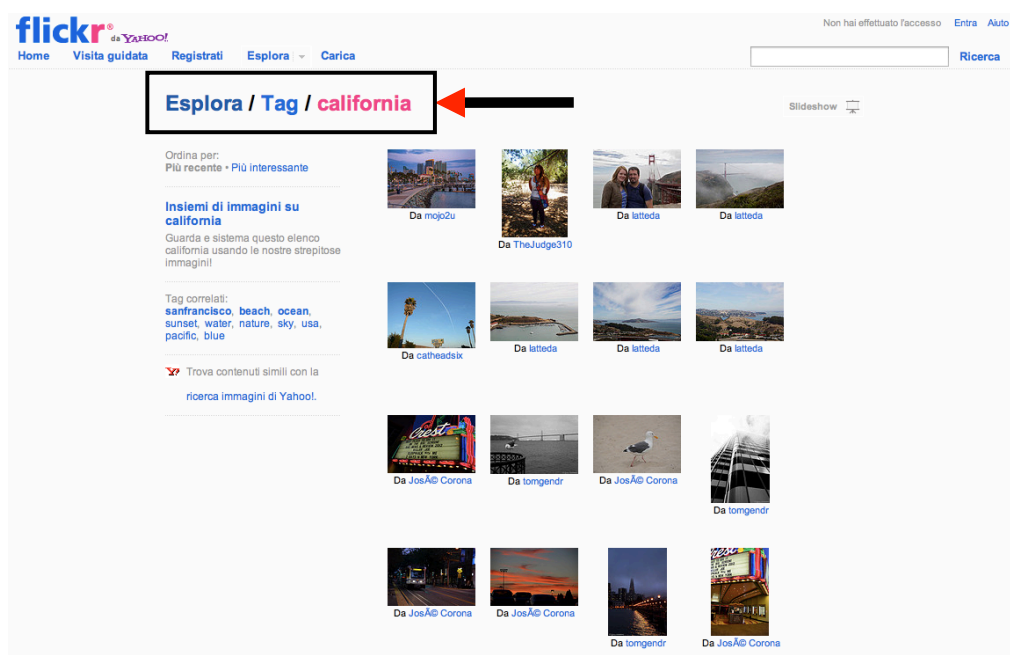
## Sistemi di ricerca di tags: Flickr



The screenshot shows the Flickr homepage interface. At the top, there is a navigation bar with links: Home, Visita guidata, Registrati, Esplora, and Carica. Below this, the main content area is divided into several sections. On the left, there is a large image titled "Rain on the sea" with a caption "Rain on the sea (Explore 1/10/12) di joeserranista.com, North Photo Tours". To the right of this image, there is a search bar with a red arrow pointing to it from the text "Modulo per la ricerca dei tags sulla home page di Flickr". Below the search bar, there is a section titled "Esplora" with a list of links: "Foto interessanti degli ultimi 7 giorni", "Visualizzazione calendario del mese in corso", "Il mappamondo", "Cerca fotocamera", "Upload più recenti", "Video su Flickr", "Gallerie", "App Garden", "Il blog di Flickr", and "Flickr per il cellulare". Below this, there is a section titled "Un anno fa oggi" with a list of links: "Fai un passo indietro nel tempo e guarda i contenuti più interessanti di 3 ottobre 2011.", "Set", "Gruppi", "FlickrToday (only 1 pic per day)", "social documentary photography", and "Paris". At the bottom, there is a section titled "Esplora Flickr attraverso i tag" with a list of tags: art, beach, blue, bw, california, canada, canon, china, city, concert, de, england, europe, family, festival, film, flower, flowers, food, france, friends, green, instagramapp, iphoneography, italy, japan, live, london, music, nature, new, newyork, night, hikon, nyc, paris, park, party, people, photography, portrait, red, sky, snow, square, squareformat, street, summer, sunset, travel, trip, uk, usa, vacation, water, wedding, white, winter.

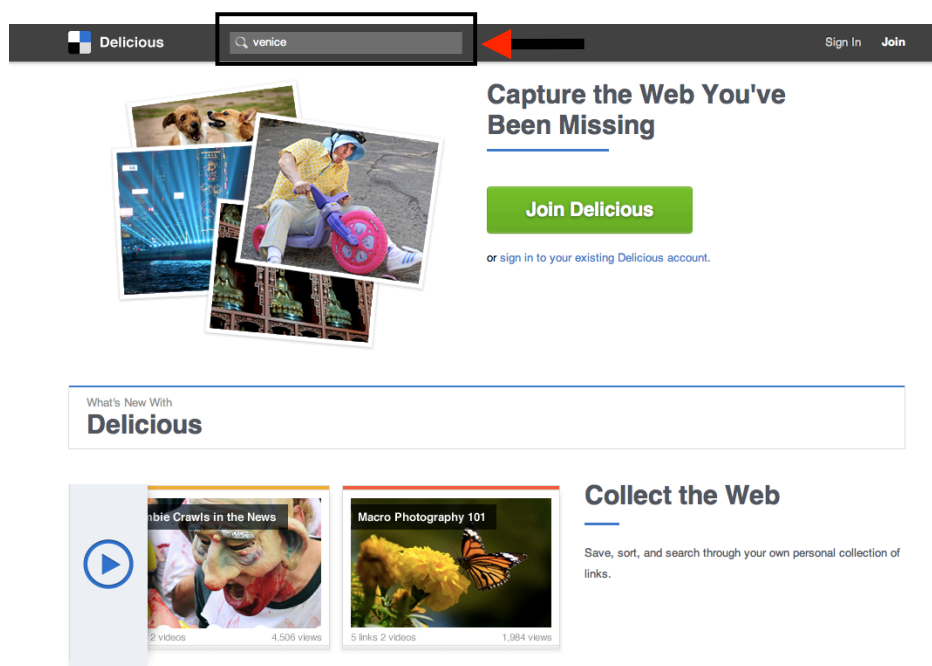
Modulo per la ricerca dei tags sulla home page di Flickr

# Sistemi di ricerca di tags: Flickr



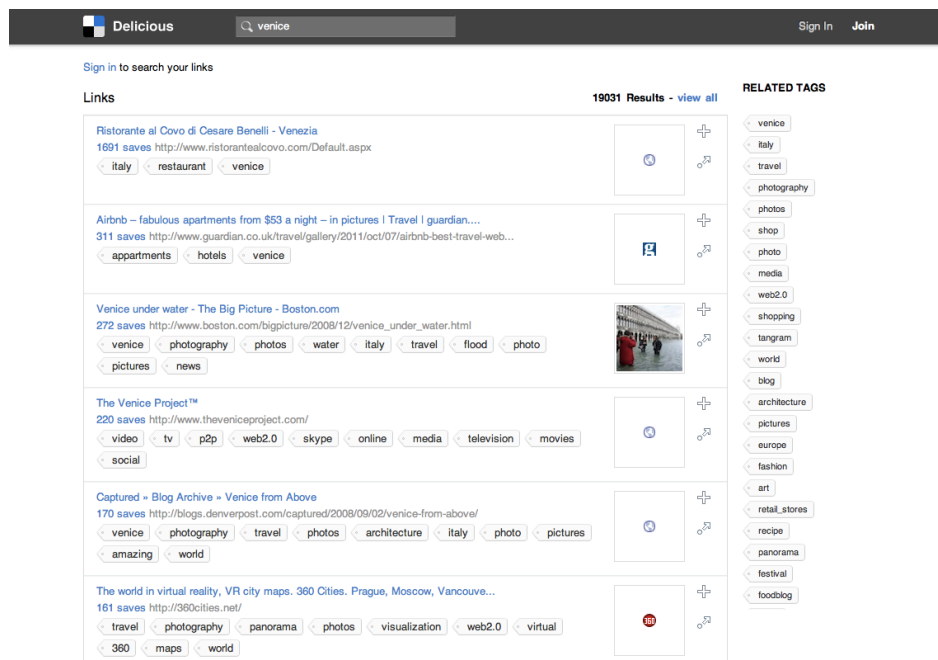
Il risultato di una ricerca per tag (california)

# Sistemi di ricerca di tags: Delicious



Home page di Delicious, servizio di bookmarking e tagging di pagine web. Viene evidenziata in rosso la form per la ricerca dei tags.

# Sistemi di ricerca di tags: Delicious



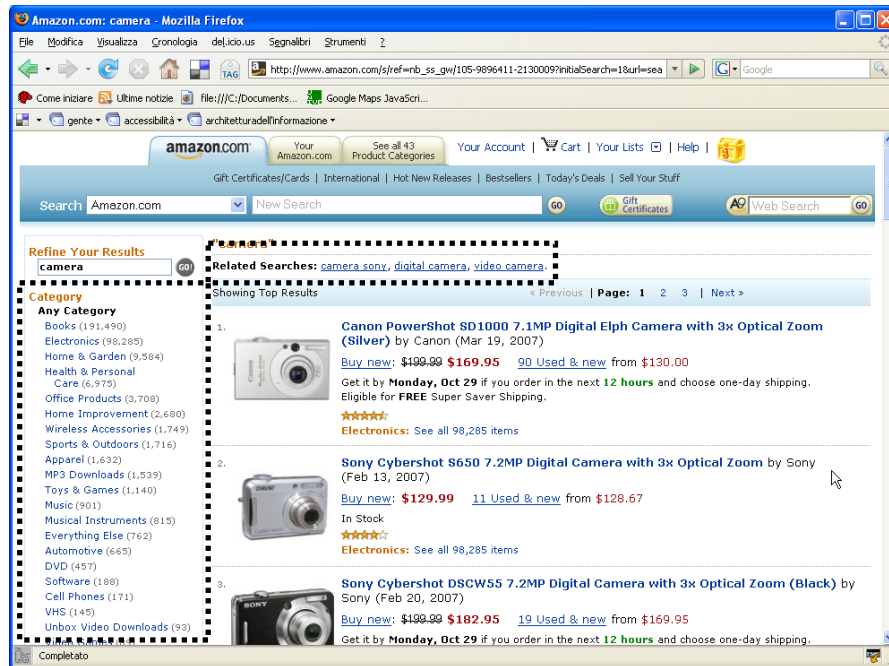
Esito della ricerca di un tag

## Trovabilità

Finding = searching + browsing

- Concludendo, è necessario integrare le due modalità di navigazione di un sito per aumentare la *trovabilità* degli oggetti informativi
- In particolare dovrebbe essere possibile passare dalla navigazione di un sito attraverso meccanismi di browsing all'esecuzione di una ricerca basata su un modulo da compilare e viceversa
- Amazon fornisce un buon esempio di integrazione delle due modalità (vedi le due slides successive)

# Trovabilità



- Amazon: dal searching al browsing

# Trovabilità



- Amazon: dal browsing al searching

## Riassumendo: Strutture informative, sistemi di navigazione e di ricerca

---

- Strutture informative
  - Definite al design time
    - Strutture gerarchiche (tassonomie) e poligerarchiche
    - Strutture sequenziali
    - Strutture ipertestuali
    - Database
  - Definite al design-time e popolate al run-time
    - Folksonomie
- Sistemi di ritrovamento dell'informazione
  - Browsing
    - Sistemi di browsing gerarchici
    - Sistemi di browsing globali (verticali e orizzontali)
    - Sistemi di browsing ad hoc/trasversali/contextuali
    - Sistemi di browsing sociali
  - Searching
    - Sistemi di searching top-down (ricerca su contenuti)
    - Sistemi di searching bottom-up top-down (ricerca su tags)

## Riassumendo: Strutture informative, sistemi di navigazione e di ricerca

---

- Anche se nessuna combinazione di strutture informative e sistemi di navigazione è ottimale per tutti i siti web, ma dipende da variabili come obiettivi del sito, complessità, utenza, ecc, per la maggior parte delle situazioni di una certa complessità la parola chiave è **integrazione**:
  - Delle strutture informative classiche definite nell'attività di design del sito (design-time) con quelle che vengono definite sempre nell'attività di design, ma popolate utilizzando il sito (run-time)
  - Dei percorsi di browsing definiti dal designer con quelli emergenti dall'attività di tagging (degli utenti, ma anche di uno o più autori del sito)
  - Dei sistemi di browsing con i sistemi di searching, per aumentare la cosiddetta findability



## Riassumendo: Strutture informative e sistemi di navigazione

---

Str. Informative > Sist. Navigaz. > Presentaz.

- Il design della presentazione (sia sotto forma di layout visuale che nelle altre forme possibili) dipende dalle scelte effettuate nella definizione delle strutture informative e dei sistemi di navigazione.

## Best practices per siti fondati sulle strutture gerarchiche I

---

- Dopo aver ribadito che nessuna soluzione è ottimale per tutti i siti web, identifichiamo alcuni punti fermi nel processo di design per i siti di una certa complessità, nei quali la struttura gerarchica svolge un ruolo primario.
- Prima di tutto vanno specificate, nell'ordine proposto, le strutture informative del sito
  - Specifica della struttura gerarchica del sito
  - Specifica delle strutture organizzative sequenziali e riconducibili a database
  - Specifica delle proprietà strutturali della folksonomia/e (area o aree informative nel quale il tagging è possibile, libertà di specifica dei tag (specifica libera, guidata, legata a vocabolario controllato), possibilità di raggruppare i tag, ecc.)
  - Specifica delle search-zones
- La specifica delle strutture informative viene utilizzata per il design del sistema di navigazione integrato, definendo nell'ordine:
  - Il sistema di browsing gerarchico
  - Il sistema di browsing globale (verticale e orizzontale)
  - Il sistema di browsing trasversale
  - Il sistema di browsing sociale, definito sulla base delle proprietà strutturali della/e folksonomia/e

## Best practices per siti fondati sulle strutture gerarchiche II

---

- Dopo aver definito il sistema di navigazione integrato, si consideri di aggiungere uno o più elementi di navigazione remota:
  - Indice dei contenuti / mappa del sito
  - Indice analitico
  - Tour guidati
  - In siti estesi si possono considerare due o più elementi (es. indice dei contenuti e analitico, che soddisfano necessità cognitive diverse)
- Specifica del sistema di searching top-down e bottom-up
- Il sistema di browsing e di searching devono essere integrati, permettendo all'utente di spostarsi a piacimento tra le due modalità, a secondo delle necessità