

UNIVERSITÀ DEGLI STUDI DI PADOVA



FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA SPECIALISTICA IN STATISTICA ED INFORMATICA

Tesi di laurea

**Un motore di ricerca
per un flusso di etichette di immagini**

Relatore: Prof. MASSIMO MELUCCI

Laureando: ANDREA GRASSETTO

Matricola: 569762

Anno accademico 2008/2009

Indice

1. Introduzione	5
2. Il progetto: Visualizzazione di etichette nel tempo	11
2.1. <i>Il progetto di Yahoo</i>	<i>11</i>
2.2. <i>Il progetto di questa tesi</i>	<i>15</i>
2.2.1. <i>Ricerca - Creazione file XML</i>	<i>16</i>
2.2.2. <i>Organizzazione delle fotografie</i>	<i>18</i>
2.2.3. <i>Reperimento di documenti dal World Wide Web</i>	<i>22</i>
2.2.4. <i>Visualizzazione</i>	<i>26</i>
3. La visualizzazione passo passo	33
4. Software utilizzati	37
4.1. <i>Microsoft Server SQL 2005Express - SQL</i>	<i>37</i>
4.2. <i>Netbeans IDE 6.5 – Java</i>	<i>37</i>
4.3. <i>Macromedia Flash 8 Mx – ActionScript</i>	<i>38</i>
5. Conclusioni	41
Bibliografia	43

1. Introduzione

Il continuo sviluppo delle tecnologie informatiche e la progressiva espansione della quantità crescente di informazioni, rendono necessario un continuo aggiornamento dei programmi da parte di chi ha realizzato il software in modo da rendere il prodotto sempre attuale ed efficiente.

Le innumerevoli esigenze degli utenti provocano una richiesta di sviluppo sia in termini di hardware che di software; in questo contesto si registra un incremento di nuovi portali web che hanno l'obiettivo di soddisfare l'esigente e sempre più raffinata richiesta degli utenti, mettendo a disposizione servizi e utilità: da pochi anni nel web è stata conosciuta una nuova tipologia di trasmissione e condivisione delle informazioni, che va sotto il nome di "Social Network".

Il Social Network è uno spazio di aggregazione virtuale in cui condividere idee, proposte di lavoro, soluzioni, costituito da un gruppo di persone interconnesse tra loro tramite dei legami sociali, che vanno dalla conoscenza casuale, ai rapporti lavorativi, sportivi, religiosi, familiari.

Il Social Network nasce nel 2003, grazie alla popolarità di siti web come Friendster, Tribe.net e LinkedIn. In Italia il primo grande portale fu SuperEva nel 2004. Attualmente i social networks più gettonati sono Myspace e Facebook con milioni e milioni di utenti in tutto il mondo (rif [1]).

Una parte del Social Network sviluppa l'attività di Tagging, una espansione presente nel web che si basa sul concetto di tag.

Il "tag" (etichetta) è una parola chiave o un termine associato a un particolare dell'informazione (un'immagine, una mappa geografica, un post, un video clip ...), che descrive l'oggetto rendendo possibile la

classificazione e la ricerca di informazioni, basate su parole chiave. I tag hanno la peculiarità di non essere codificati in schemi preconfezionati, ma creati dall'inventiva, dall'intuizione, dalla logica e dalla fantasia degli autori/creatori dell'oggetto dell'indicizzazione (rif [1]).

I tag possono anche essere usati in modo errato, ovvero fornire indicazioni riguardo all'opinione che qualcuno ha di un'opera e quindi essere correlati al consumatore del contenuto e non al contenuto in sé.

Da queste osservazioni si evince come il semplice associare tag non sia sufficiente a dare un livello semantico alla rete, sebbene alcuni includano il meccanismo del tagging nel web semantico.

La classificazione, attraverso le etichette, ha portato allo sviluppo della cosiddetta folksonomia. La folksonomia è un neologismo, deriva dall'inglese folksonomy, una categorizzazione collaborativa di informazioni mediante l'utilizzo di parole chiave (o tag) scelte liberamente (rif [1]). Il termine è formato dall'unione di due parole, folk e tassonomia; una folksonomia è, pertanto, una tassonomia creata da chi la usa. In maniera più semplice e concreta, questo termine si riferisce alla metodologia utilizzata da gruppi di persone che collaborano spontaneamente per organizzare in categorie le informazioni disponibili attraverso internet.

Questo fenomeno, in contrasto con i metodi di classificazione formale (in particolare con la tassonomia classica), cresce soprattutto in comunità non gerarchiche legate ad applicazioni web, attraverso le quali vengono diffusi contenuti testuali e/o multimediali.

I due portali di maggior spicco, che utilizzano questa metodologia di classificazione e di reperimento, sono Delicious (del.icio.us) e Flickr (www.flickr.com).

Il primo nasce nel 2003 grazie a Joshua Schachter, ingegnere americano, che creò del.icio.us, come servizio per condividere links

con i suoi amici, una sorta di linkblog basato su un ingegnoso sistema di tagging: ad ogni bookmark (segnalibro) salvato si può associare un tag ed una descrizione. Questa funzionalità consente di accelerare enormemente i tempi di ricerca dei links e di organizzare in maniera ordinata tutti i propri preferiti (rif [1], [2]).

A questo si aggiunge il supporto per la syndacation RSS che espande ancor più le capacità di condivisione del servizio. RSS (Really Simple Syndication) è uno dei più popolari formati per la distribuzione di contenuti Web, basato su XML.

Nel 2004 nasce il portale Flickr ([ww.flickr.com](http://www.flickr.com)). Si tratta di un sito web multilingua che permette agli iscritti di condividere fotografie e video personali con chiunque abbia accesso a Internet, in un ambiente web 2.0 (rif [1], [3]).

Il sito, di proprietà del gruppo Yahoo!, ha una libreria in continua crescita contando ogni minuto più di duemila nuove foto inserite da parte dei suoi oltre venti milioni di utenti.

Flickr è stato sviluppato dalla Ludicorp, una compagnia canadese di Vancouver fondata nel 2002. Nel marzo del 2005, sia la Ludicorp che Flickr sono stati comprati da Yahoo! I server si sono quindi trasferiti dal Canada agli Stati Uniti. Il 16 maggio 2006 Flickr conclude la fase di sviluppo beta, definendosi in stato "Gamma", ossia non più in fase di prova ma in stato di perpetua evoluzione.

Il 29 dicembre 2006 sono stati ritoccati i limiti portando a 100 MB al mese (dai precedenti 20 MB) per gli account gratuiti ed a 2 GB al mese per gli utenti a pagamento. Nel gennaio 2007 Flickr ha completato la migrazione alla piattaforma Yahoo! facendo trasferire forzatamente i vecchi utenti, con account aperto prima dell'acquisizione del 2005, ad un account della piattaforma Yahoo.

Nel giugno 2007 Flickr ha concluso anche la sua fase Gamma, sostituendo la dicitura Gamma con "Flickr LOVES YOU". Nello stesso mese viene anche inaugurato un sistema di censura, guidato dagli utenti, per filtrare le fotografie considerate potenzialmente controverse.

Questo innovativo portale permette di:

- caricare le foto dal tuo desktop, via email o attraverso un cellulare
- organizzare le proprie foto e video usando le raccolte, i set e i tag
- condividere le proprie foto e video per mezzo di gruppi
- condividere con gli altri il luogo nel quale si è scattata la foto o il video grazie a delle cartine virtuali
- creare biglietti da visita, photo-book, stampe incorniciate, copertine per cd/dvd, ecc.

Aggregando tutte queste informazioni e prendendo spunto dall'articolo "Visualizing Tags over Time" tratto dalla conferenza World Wide Web 2006, sono state create le basi necessarie per sviluppare la tesi (rif [4]).

L'articolo spiega il progetto di un'applicazione eseguita in Flash: una serie di tag sono visualizzati nello schermo e gli utenti hanno la possibilità di interagire con essi. Vengono fornite, in output, informazioni sotto forma multimediale con foto riferite al tag stesso.

Nella tesi viene rielaborata questa idea con una presentazione grafica più semplice ed intuitiva. La novità sviluppata consiste di offrire agli utenti la possibilità di consultare, tramite l'indicazione di indirizzi web riportati sotto ogni etichetta, documentazioni testuali ricavate da una ricerca del tag in Wikipedia.

In particolare il lavoro svolto consiste nell'abbinare ai tag i contenuti multimediali trovati in prima battuta (immagini) e gli indirizzi internet a cui far riferimento per consultare velocemente il web senza bisogno di ulteriori ricerche.

La presentazione è organizzata come segue.

Nel Cap. 2 sono sviluppati i seguenti argomenti:

- indicazioni che riguardano il funzionamento del progetto di Yahoo! Research
- descrizione del lavoro realizzato nella tesi nel suo sviluppo organico
- esposizione di come viene effettuata la ricerca e la creazione del file XML
- presentazione del risultato ottenuto con l'organizzazione delle fotografie e con il reperimento della documentazione testuale
- indicazioni sullo sviluppo software e sul linguaggio utilizzato
- visualizzazione grafica di tutte le informazioni ottenute.

Nel Cap. 3 viene mostrata una presentazione passo passo del risultato ottenuto.

Nel Cap. 4 vengono citati i programmi e i linguaggi di programmazione utilizzati.

2. Il progetto: Visualizzazione di tag nel tempo

2.1. Il progetto di Yahoo! Research

Nel maggio 2006 ad Edinburgo, alla conferenza del World Wide Web, Yahoo! Research ha presentato un nuovo approccio basato su una categorizzazione dei tag più interessanti, recuperati all'interno di Flickr (flickr.com), associati ad un intervallo di tempo di scorrimento. Uno dei servizi fondamentali del portale Flickr è la possibilità, da parte degli utenti iscritti, di poter etichettare le proprie foto mediante una o più parole chiavi. Negli ultimi anni gli utenti hanno aggiunto, in media, oltre un milione di etichette ogni settimana (rif [4]).

Comprendere l'evoluzione di questi tag nel tempo è quindi un compito impegnativo. L'animazione in Flash, fornita tramite un browser Web, permette all'utente di osservare, di interagire con il tag interessante e di verificarne l'evoluzione nel tempo.

Sono stati creati nuovi algoritmi e strutture dati per garantire l'efficienza di questa generazione di visualizzazione.

Yahoo! Research ha coniato una nuova soluzione, in grado di produrre visualizzazioni casuali a scale su un grande set di dati in tempo reale.

Osservando, nel corso del tempo, il comportamento degli utenti in questi ambienti sociali, Yahoo! Research è stata in grado di esplorare i cambiamenti della comunità. Ha quindi studiato un metodo per visualizzare l'evoluzione del tag interessante in Flickr. La ricerca dei tag avviene considerando un periodo di tempo di un anno; per ogni giorno vengono scelti i tag più rappresentativi (rif [5]).

La visualizzazione ha determinati requisiti:

- fornire una visione di evoluzione temporale, con una grande quantità di dati facilmente visibile in ciascun periodo di tempo (giorno)
- permettere all'utente di interagire con la presentazione, al fine di visualizzare in dettaglio un particolare risultato
- rimanere "situato", nel senso che l'utente deve essere sempre consapevole dell'attuale punto di tempo presentato
- fornire un accesso casuale in streaming, in modo che l'utente possa posizionarsi in qualsiasi punto in qualsiasi momento.

Un programma che segue questi requisiti è Flash MX, grazie all'apporto di un linguaggio di programmazione singolare come l'Actionscript.

La visualizzazione, presentata da Yahoo! Research, si avvale di due regioni schermo. Nella parte superiore della pagina è presente una linea del tempo che identifica il giorno relativo ai tag più interessanti rappresentati nel resto della pagina. Questa barra orizzontale può essere anche spostata dall'utente in entrambe le direzioni consentendo un accesso diretto ai giorni che definiscono l'intervallo.

Questi tag entrano ed escono dalla pagina in due situazioni differenti chiamate metafore:

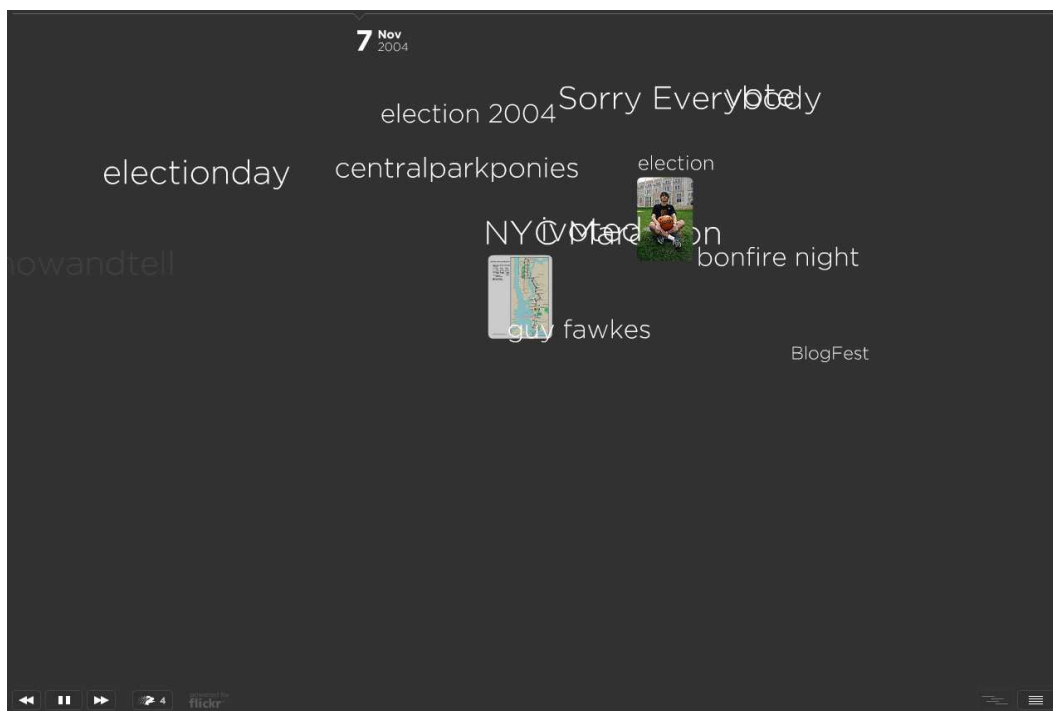
- la metafora a "fiume", in cui i tag seguono un flusso che va da destra a sinistra dello schermo; ogni tag ha al suo fianco una foto relativa da Flickr. Se il tag è "catturato" usando il puntatore del mouse durante il suo viaggio, vengono visualizzate più foto di Flickr
- la metafora a "cascata", in cui vengono visualizzati centralmente i primi 8 tag più interessanti in 8 righe. A lato di questi tag, sono presenti le relative foto recuperate da Flickr. Se

un tag persiste per più giorni consecutivi, viene visualizzato nello stesso slot del giorno precedente aggiungendo le nuove foto. Questa metafora è utile per studiare i tag che persistono per più giorni

Le visualizzazioni sono disponibili all'indirizzo:

<http://research.yahoo.com/taglines>.

In figura viene riportato un frame della visualizzazione definita “a fiume”. In alto, nello schermo, è visibile la data, mentre al centro, scorrono le etichette.



Per definire l'importanza di un tag, Yahoo! Research ha introdotto degli algoritmi più o meno complessi basati su delle combinazioni di intervalli predefiniti che sono stati calcolati precedentemente e salvati in una struttura “Indice”. È stato perfezionato il ben noto algoritmo di soglia (Threshold Algorithm),

con la creazione e lo sviluppo degli algoritmi additivi e sottrattivi, i quali inglobano o suddividono i vari intervalli predefiniti (rif [7]).

I dati sono stati ottenuti da Flickr. Nella struttura dati sono presenti le seguenti caratteristiche: data, identificativo della foto, identificativo dell'utente che scrive l'etichetta alla foto, e i tag. Nella tabella sono state contate ben 86.8M righe, ogni riga rappresenta un esempio di foto con le caratteristiche elencate precedentemente. Di questi tag circa 1.26M sono unici, quindi ogni tag viene ripetuto, in media, 70 volte. Sono stati inseriti 472 giorni di foto e tag, a partire dal 3 giugno 2004, in media più di 1.2M tag a settimana.

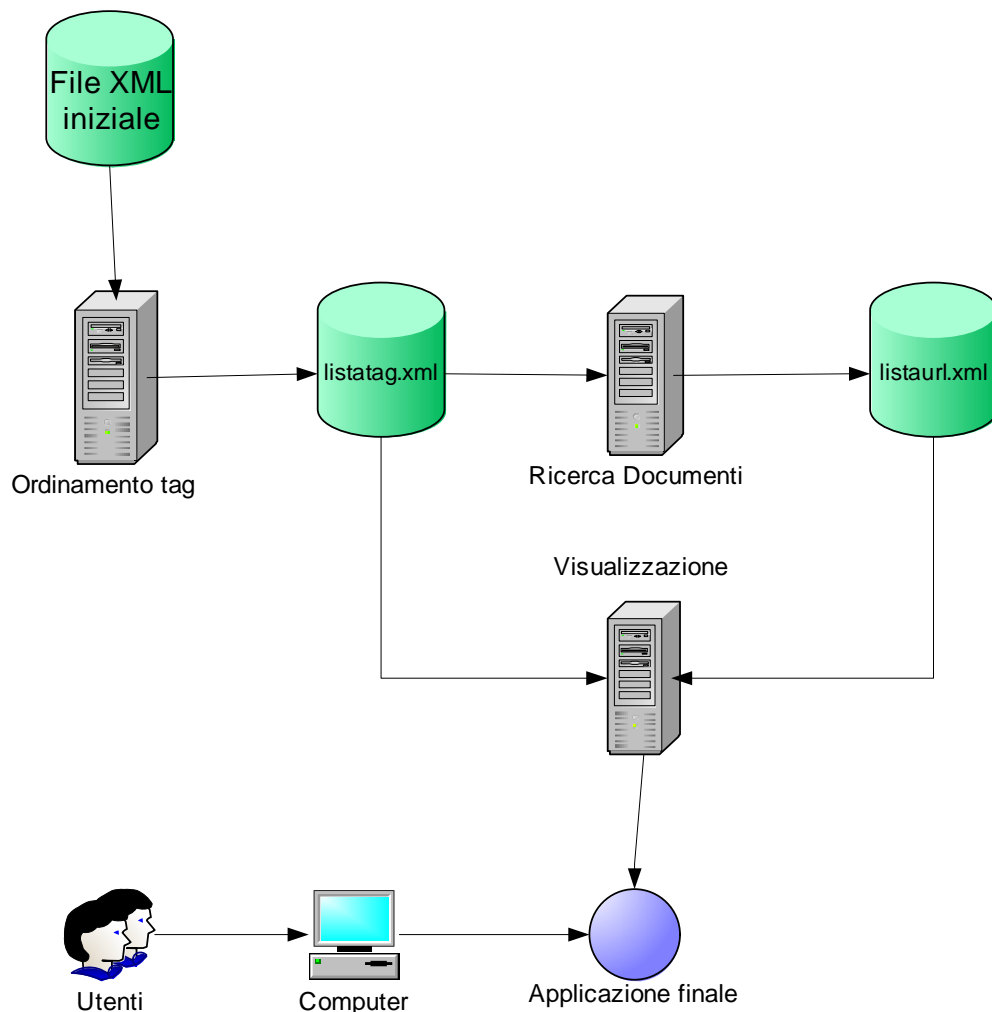
In primo luogo, Yahoo! Research ha effettuato alcune semplici operazioni di pulizia dei dati in questa tabella. È stato estratto il tag grezzo e rimossi gli spazi bianchi, i segni di punteggiatura, e convertiti in minuscolo eventuali caratteri in stampato maiuscolo (rif [4]).

2.2. Il progetto di questa tesi

Prendendo spunto dal progetto di Yahoo! Research, è stata implementata la visualizzazione grafica con il reperimento dei documenti più interessanti, ricavati dal sito di Wikipedia.

In questa sezione vengono illustrate le varie procedure di analisi e sviluppo del lavoro con descrizioni tecniche dettagliate e frammenti di codice ritagliati. In particolare:

- creazione della struttura dati sottoforma di file XML
- ordinamento dei tag e delle foto in base al giorno
- ricerca dei documenti più interessanti
- visualizzazione dei vari contenuti



2.1.1. Ricerca – Creazione del file XML

La creazione della collezione di foto e di tutte le sue informazioni è stata effettuata manualmente. Si tratta di una collezione di scarse dimensioni, creata appositamente senza dare molta importanza all'entità numerica dei dati inseriti; il database è sufficiente per ottenere le condizioni necessarie allo sviluppo delle funzioni e delle strutture rappresentate nell'applicazione finale.

Nel Febbraio 2007 è nata CoPhIR (Content-based Photo Image Retrieval), una collezione di foto scaricate dal portale Flickr; è stato creato un file XML di dimensioni pari a 4.5GB con 300 milioni di foto distinte (rif [8]).

Avvalersi di una banca dati più estesa (la struttura della raccolta è stata presa come modello), esistente presso l'Istituto di Scienza e Tecnologie (ISTI), istituto del Consiglio Nazionale delle Ricerche (CNR) di Pisa, non avrebbe cambiato il risultato dell'applicazione.

In questo file, oltre all'indirizzo url della foto nel sito di Flickr, vengono evidenziate molte altre informazioni: titolo e descrizione, identificazione e locazione dell'autore, i tag che etichettano la foto, i commenti, il numero di volte che essa è stata visualizzata, il numero di utenti che hanno aggiunto la foto tra i loro preferiti, la data di inserimento della stessa nel sito, e altri dati.

Non avendo a disposizione questi dati, è stato creato un file con una struttura XML simile a quella che l'ISTI avrebbe potuto fornire.

Per XML (eXtensible Markup Language) si intende un metalinguaggio di demarcazione che fornisce un formato per la descrizione di dati strutturati. Ciò consente di dichiarare con maggiore intuitività il contenuto dei dati e di ottenere risultati più significativi nelle ricerche eseguite in più piattaforme. Inoltre, il linguaggio XML

renderà possibile una nuova generazione di applicazioni Web per la visualizzazione e la manipolazione di dati (rif [1], [9]).

Nel file, chiamato “listimmagini.xml”, sono presenti una serie di informazioni:

- necessarie ai fini del progetto, come: indirizzo url della foto, nome della diapositiva (foto con dimensioni ridotte), data di inserimento, lista di tag e numero di visualizzazioni
- non necessarie ai fini del progetto, come: autore, nome e descrizione della foto, commenti degli utenti.

Il progetto sviluppato prevede che, per il funzionamento dell'applicazione, la raccolta debba avere i seguenti requisiti:

- per ogni giorno ci sono almeno cinque tag
- per ogni giorno ci sono almeno tre foto che rappresentano un tag

Segue un estratto del file.

```
<image>
<link>http://www.flickr.com/photos/todorrovic/3117216307/
</link>
<thumbnail>3117216307_8f88e03369_t.jpg</thumbnail>
<metadata>
  <author>Dragan*</author>
  <title>Merry Christmas!</title>
  <location></location>
  <tags>
    <tag>Merry Christmas</tag>
    <tag>Christmas pictures</tag>
    <tag>Christmas images</tag>
    <tag>Xmas</tag>
    <tag>Xmas pictures</tag>
    <tag>Christmas pics</tag>
    <tag>Blue</tag>
    <tag>Silver</tag>
    <tag>Blue light</tag>
    <tag>Ball ornament</tag>
    <tag>Christmas ornament</tag>
  </tags>
<comments></comments>
```

```
<views>1103</views>
<date>2008-12-18</date>
</metadata>
</image>
```

Queste informazioni sono state trascritte direttamente dal sito di Flickr. Da segnalare che la thumbnail (diapositiva) è un'immagine di dimensioni ridotte rispetto a quella originale ed è salvata direttamente nel computer. Questa immagine sarà necessaria nella visualizzazione finale.

2.2.2 Organizzazione delle fotografie

In questa sezione viene descritto come vengono ordinate le foto e i relativi tag, tramite delle interrogazioni SQL.

Dato in ingresso il file “listaimmagini.xml”, lo scopo è quello di creare un nuovo file XML, contenente una serie di etichette ordinate per data di introduzione nel portale Flickr.

Il file iniziale viene caricato in una variabile locale con queste istruzioni (rif [10], [11], [12]):

```
set @xmlDoc = (
  select Q.BulkColumn from
  OPENROWSET(BULK 'listaimmagini.xml', SINGLE_CLOB) as Q
);
```

Lo step successivo consiste nel recuperare tutte le foto che quel giorno sono state identificate con quel determinato tag, raggrupparle e ordinarle per numero decrescente di visualizzazioni effettuate.

Vengono recuperate dal file le singole etichette e le relative date; viene salvato in una tabella il numero di volte che un particolare tag compare in una determinata data:

```

select tag, date, count(*) as number into listatag
from OPENXML (@hdoc, '/root/image/metadata/tags/tag', 2)
with (
    date varchar(10) ' ../../date',
    tag varchar(100) 'text()'
)
group by date, tag order by number desc;

select distinct date into listadate
from listatag;

```

Per agevolare i compiti di ricerca e di ordinamento, si è ritenuto più semplice constatare, come tag rilevanti, quelle parole chiavi che vengono ripetute il maggior numero di volte nello stesso giorno, mentre come foto più rilevanti, quelle foto con più alto numero di visualizzazioni.

Ogni foto può avere un numero n di tag, con n numero arbitrario; può verificarsi la presenza di una foto sprovvista di etichette. Ogni tag può essere presente su un numero k di foto, con k numero casuale maggiore di uno (se il tag è presente deve essere stato per forza associato almeno ad una foto).

Quindi, con un ciclo while, si ricavano i cinque tag più importanti di ogni specifico giorno:

```

while @lengthdate>0
BEGIN
insert into listatagtop(tag,date,number)
select top 5 listatag.tag,listatag.date,listatag.number
from listatag, listadate
where listatag.date = listadate.date and
      listadate.id = @lengthdate
order by listadate.date, number desc;

set @lengthdate = @lengthdate -1;
END

```

Nel passo successivo si sono recuperate le informazioni necessarie:

- link: link della foto su Flickr

- thumbnail: link della diapositiva sul computer
- date: data di inserimento dell'immagine in Flickr
- views: numero di visualizzazioni della foto
- tag: parola chiave che identifica la foto

```
select * into listaimg
from OPENXML (@hdoc, '/root/image/metadata/tags/tag', 2)
with (
    link varchar(100) '...../link',
    thumb varchar(100) '...../thumbnail',
    date varchar(10) '.../date',
    views int '.../views',
    tag varchar(100) 'text()'
);
```

L'ultima operazione è quella di scrittura del file XML.

Utilizzando la clausola FOR XML seguita dalla direttiva AUTO è possibile ottenere un file XML con elementi nidificati: per ogni tabella specificata nella clausola FROM della query, viene creato un nuovo livello nella struttura XML; l'ordine di nidificazione dei dati è determinato dall'ordine delle colonne dichiarate nella SELECT (rif [13]). Quindi:

```
select tag.tag as tname, tag.date as tdate, link,
       thumb, views
from listaimg as foto, listatagtop as tag
where foto.tag = tag.tag
order by tdate, number desc, tname, views desc
FOR XML Auto, ROOT('root');
```

Nell'estratto del file di output si vede come ogni tag sia descritto con un nome (tname) e una data (tdate).

```
<tag tname="roma" tdate="2008-10-25">
  <foto link="..." thumb="3196903700_f708961ffc_s.jpg"
        views="995" />
  <foto link="..." thumb="3201926698_9476f319f2_t.jpg"
        views="753" />
```

```

    <foto link="..." thumb="3195123200_7b9f0717b4_t.jpg"
      views="75" />
    <foto link="..." thumb="3184054757_df01f70fec_t.jpg"
      views="51" />
    <foto link="..." thumb="3184054761_d18275a62c_t.jpg"
      views="16" />
  </tag>
  <tag tname="rome" tdate="2008-10-25">
    <foto link="..." thumb="3196903700_f708961ffc_s.jpg"
      views="995" />
    <foto link="..." thumb="3201926698_9476f319f2_t.jpg"
      views="753" />
    <foto link="..." thumb="3195123200_7b9f0717b4_t.jpg"
      views="75" />
    <foto link="..." thumb="3184054757_df01f70fec_t.jpg"
      views="51" />
    <foto link="..." thumb="3184054761_d18275a62c_t.jpg"
      views="16" />
  </tag>
  <tag tname="italy" tdate="2008-10-25">
    <foto link="..." thumb="3196903700_f708961ffc_s.jpg"
      views="995" />
    <foto link="..." thumb="3201926698_9476f319f2_t.jpg"
      views="753" />
    <foto link="..." thumb="3184054761_d18275a62c_t.jpg"
      views="16" />
  </tag>
  <tag tname="kathmandu" tdate="2008-10-25">
    <foto link="..." thumb="3162153392_f80f14ed58_t.jpg"
      views="33" />
    <foto link="..." thumb="3104716340_17afab188d_t.jpg"
      views="20" />
    <foto link="..." thumb="3183191081_345f809216_t.jpg"
      views="16" />
  <tag tname="temple" tdate="2008-10-25">
    <foto link="..." thumb="3184054757_df01f70fec_t.jpg"
      views="51" />
    <foto link="..." thumb="3183191081_345f809216_t.jpg"
      views="16" />
    <foto link="..." thumb="3184054761_d18275a62c_t.jpg"
      views="16" />

```

Questa parte di file si riferisce alla data del 25 Ottobre 2008. Ogni tag è ordinato per numero di foto, infatti i primi due tag (“roma” - “rome”) rappresentano cinque foto, mentre gli ultimi tre (“italy” - “kathmandu” - “temple”), ne identificano tre. Ogni foto, minimo tre per ogni tag, è ordinata secondo la sua visibilità (campo views).

2.2.3. Reperimento di documenti dal World Wide Web

L'implementazione effettuata al lavoro di Yahoo! Research consiste nella ricerca e nel reperimento di pagine web relative al tag preso in considerazione.

In particolare, il lavoro eseguito nella tesi, studia e realizza una ricerca nel portale Wikipedia, tramite il motore di ricerca Google.

Wikipedia è un'enciclopedia online, multilingue, a contenuto libero, redatta in modo collaborativo da volontari e sostenuta dalla Wikimedia Foundation, un'organizzazione senza fine di lucro (rif [1]).

La scelta è caduta su questa enciclopedia on-line, in quanto, attualmente, risulta essere tra le più utilizzate e proposte dal web per la vasta gamma di argomenti trattati.

La ricerca consiste nell'inserire, nella casella di testo di Google, la parola chiave della quale si vuole avere dei riferimenti web.

Si aggiunge, dopo la parola chiave, la dicitura "*site:*" seguita dall'indirizzo del sito nel quale si effettua la ricerca (nel nostro caso "*en.wikipedia.org*"). Google produce come risultato una lista di pagine web trovate solo ed esclusivamente nel sito *en.wikipedia.org*.

Con un programma Java, dal file di ingresso costruito al punto 2.2.2, chiamato "listatag.xml", è stato prodotto un altro file che contiene, per ogni tag selezionato, una lista dei primi tre indirizzi web reperiti, con i relativi snippet che sono dei frammenti ed esempi di codice sorgente delle varie pagine web ricercate (vedi immagine a pagina seguente) (rif [10]).



Web

Rome - Wikipedia, the free encyclopedia - [Traduci questa pagina]
Rome's history as a city spans over two and a half thousand years, as one of the founding cities of Western Civilisation. Even outside of the history of the ...
en.wikipedia.org/wiki/Rome - 249k - Copia cache - Pagine simili

Rome (TV series) - Wikipedia, the free encyclopedia - [Traduci questa pagina]
Rome is a BAFTA, Golden Globe-nominated and Emmy-winning historical drama television series co-created by John Milius, William J. MacDonald, ...
en.wikipedia.org/wiki/Rome_(TV_series) - 106k - Copia cache - Pagine simili

Ancient Rome - Wikipedia, the free encyclopedia - [Traduci questa pagina]
Ancient Rome was a civilization that grew out of a small agricultural community founded on the Italian Peninsula as early as the 10th century BC. ...
en.wikipedia.org/wiki/Ancient_Rome - 256k - Copia cache - Pagine simili

Mediante l'utilizzo di alcune procedure standard della JDOM API, quali le funzioni *build()*, *getRootElement()*, *getChildren()* e *getAttributeValue()*, si carica il file "listatag.xml" e lo si elabora estraendo le informazioni necessarie alla ricerca su Google (rif [15]). Il successivo passaggio consiste nell'individuare il nome del tag, al quale verrà aggiunta la dicitura "site:en.wikipedia.org".

```
SAXBuilder builder = new SAXBuilder();
Document document = builder.build (new File
                                   ("listatag.xml"));
Element rootElement = document.getRootElement();
List children = rootElement.getChildren();
Iterator iterator = children.iterator();
while (iterator.hasNext()){
    Element item = (Element)iterator.next();
    String tag = item.getAttributeValue("tname");
    tags[i] = (tag + " site:en.wikipedia.org");
    i++;
}
```

Il vettore contenente la lista dei tag viene dato in ingresso a Google mediante la GOOGLE API (rif [16], [17], [18], [19]). Le funzioni *doSearch()* e *getResultElements()* sono essenziali per trovare ed estrapolare tutte le informazioni che Google è in grado di recepire dalla ricerca effettuata. Successivamente con *getURL()* e *getSnippet()*, si riesce ad identificare e salvare le indicazioni che servono successivamente.

```
GoogleSearch s = new GoogleSearch();
String [][] listaur1 = new String[10][2];
s.setQueryString(tag);
GoogleSearchResult result = null;
GoogleSearchResultElement [] resultElement = null;
try {
    result = s.doSearch();
    resultElement = result.getResultElements();
    for(int j=0;j<resultElement.length;j++){
        listaur1[j][0]=resultElement[j].getURL();
        listaur1[j][1]=resultElement[j].getSnippet();
    }
}
```

Nel frattempo è già stato creato il nuovo file e, in maniera sincronizzata, vengono inseriti gli url e i relativi snippet per ogni tag.

```
Element tag = new Element ("tag");
tag.setAttribute("searchname",listatag[i]);
listaur1 = new Search().search(listatag[i]);
for(int k=0;k<3;k++) {
    Element url = new Element ("url");
    url.setAttribute("link",listaur1[k][0]);
    url.setAttribute("snippet",listaur1[k][1]);
    tag.addContent(url);
}
```


Una parte del nuovo file XML, creato e denominato “listaur.xml”, è la seguente:

```
<tag searchname="roma site:en.wikipedia.org">
  <url link="http://en.wikipedia.org/wiki/Roma_people"
    snippet="Jump to: navigation, search. Roma
    people may refer to ... Roma (Romani subgroup).
    , a subgroup of the Romani people who live
    primarily in Central and Eastern ..." />
  <url link="http://en.wikipedia.org/wiki/A.S._Roma"
    snippet="Associazione Sportiva Roma, (BIT: ASR)
    commonly referred to as simply Roma or AS .
    Roma, is an Italian professional football club
    from Rome. ..." />
  <url link="http://en.wikipedia.org/wiki/Roma"
    snippet="Jan 22, 2009 ... Roma rice, an Italian
    cultivated variety of rice particularly
    suitable for . risotto; Roma tomato, a plum
    tomato commonly found in ..." />
</tag>
```

Ogni tag ha un attributo (searchname) che rappresenta la richiesta di reperimento nel motore di ricerca Google. Essa è rappresentata dal nome del tag succeduto dalla dicitura “site:en.wikipedia.org”. Inoltre ogni tag ha tre figli chiamati “url” che rappresentano le tre pagine web che Google ha dato come risposta all’interrogazione. Questi figli sono a loro volta rappresentati da due attributi: il primo (link) rappresenta il collegamento alla pagina web, il secondo (snippet) rappresenta lo snippet del documento stesso.

2.2.4. Visualizzazione

L'ultimo step della tesi raffigura, mediante una visualizzazione grafica, i contenuti programmati, ricercati e sviluppati nelle sezioni 2.2.1, 2.2.2, 2.2.3.

Osservando la collocazione del tempo e la dinamica del movimento proposta da Yahoo! Research, si è creata una rappresentazione dei tag nello schermo con una sequenza dell'animazione simile, adattata però alle esigenze del progetto.

Per ogni giorno, indicato in alto a sinistra, appaiono le cinque etichette più rappresentative, scelte nella fase di organizzazione delle immagini e dei tag.

Questi tag scorrono da destra verso sinistra dello schermo finché l'utente non intende interagire con essi; il passaggio del mouse, sopra una di queste parole chiave, blocca immediatamente il flusso e, accanto all'etichetta selezionata, compaiono le tre foto con maggior visibilità.

Il clic del mouse, sopra una delle foto, produce l'apertura di una pagina web dal sito www.flickr.com, nella quale l'immagine stessa viene riprodotta nelle dimensioni originali assieme a tutte le sue caratteristiche.

Sotto al tag, bloccato dall'utente, risulta visibile una casella di testo nella quale sono inseriti gli url e gli snippet dei primi tre documenti reperiti con la ricerca in Wikipedia.

Questa casella di testo è selezionabile, all'utente viene offerta la possibilità di copiare l'indirizzo visualizzato o la descrizione della pagina per verificarne successivamente il contenuto.

Il lavoro è iniziato con la disposizione dei file XML creati in una cartella comune, nella quale saranno presenti anche tutte le diapositive

scaricate dal portale Flickr. E' stata quindi sviluppata la realizzazione grafica, con la creazione degli oggetti sul piano di lavoro.

Si fa riferimento principalmente a quattro oggetti:

- tag: clip filmato sotto forma testuale che scorre nella presentazione da destra verso sinistra e che rappresenta il valore dell'etichetta
- snippet: clip filmato sotto forma testuale che viene abilitato al passaggio del mouse sopra il tag e che incorpora i tre indirizzi url e snippet ricercati con l'applicazione Java
- thumb: clip filmato sotto forma fotografica che viene visualizzato alla destra del tag al passaggio del mouse sul tag stesso; mostra la diapositiva dell'immagine etichettata con quella parola chiave
- mc: clip filmato, che incorpora i tre clip filmati precedenti, utile per distinguere, abilitare o disabilitare l'intero blocco di oggetti

Nel foglio sono stati creati cinque mc indicizzati con nome differente. Le operazioni all'interno di uno di questi movie clip sono le medesime per tutti e cinque, a differenza della posizione iniziale sullo schermo.

Ora si tratta di definire il numero di fotogrammi necessari per effettuare il movimento ad ogni clip filmato e inserire istruzioni di codice ActionScrip nella sezione "Azioni" del programma.

Ogni mc è caratterizzato da tre fotogrammi. Nel primo vengono dichiarate e inizializzate le variabili necessarie.

```
_x = 900;  
_y = 80;  
_global.i = -5;  
k = 1;  
xspeed = 10;
```

```
rightedge = 900;
```

Tra il secondo ed il terzo fotogramma avviene il movimento del tag. Nel secondo fotogramma si decrementa la coordinata x del clip filmato e, qualora questo si trovasse fuori dallo schermo, rientrerebbe dalla posizione iniziale:

```
_x -= xspeed;
if (_x > 300 || _x < 100)
    _x -= xspeed;
if (_x < -300) {
    _x = rightedge;
    i += 5;
    if(k == 1){
        tag.nextFrame();
        k = 2;
    }
    else{
        tag.prevFrame();
        k = 1;
    }
}
```

Nel terzo fotogramma viene inserita solo la funzione *gotoAndPlay(2)* che fa tornare l'esecuzione del secondo fotogramma. In questo modo, come un loop infinito, la coordinata x viene continuamente spostata e si ha la percezione del movimento.

Il clip tag invece necessita di due fotogrammi uguali, in ognuno dei quali viene caricato ed elaborato il file "listatag.xml" (rif [23], [24]). Viene in questo modo estratto il tag dalla struttura dati e inserito nella casella di testo predefinita.

```
var tagXML:XML = new XML();
```

```

tagXML.ignoreWhite = true;
tagXML.load("listatag.xml");
tagXML.onLoad = function():Void{
i += 5;
if(i >= tagXML.firstChild.childNodes.length){
    i = 0;
}
tag_txt.text =
    tagXML.firstChild.childNodes[i].attributes.tname;

```

L’inserimento delle parole chiave e il movimento delle stesse da destra verso sinistra è stato completato. Ora è necessario stabilire ed uniformare gli eventi che devono essere applicati al passaggio o meno del mouse sopra una determinata etichetta.

Quindi nell’azione “*onClipEvent(mouseMove)*” del clip filmato tag si inserisce la condizione necessaria per la quale seguirà una serie di istruzioni, ovvero:

```

if (this.hitTest(_root._xmouse, _root._ymouse))

```

Se questa condizione è verificata, ovvero, se il mouse passa sopra all’oggetto tag, allora si verificheranno tre eventi fondamentali:

- tutto l’mc, e quindi il tag, viene bloccato
- sotto l’etichetta compaiono gli indirizzi e gli snippet riguardanti le parole chiavi
- a destra dell’etichetta appaiono tre immagini in successione che identificano quel tag; esse vengono caricate grazie alla funzione *loadMovie()*

```

//blocco l’mc
_root.mc.stop();

```

```

//carico gli url e gli snippet
var urlXML:XML = new XML();
urlXML.ignoreWhite = true;
urlXML.load("listaur1.xml");
urlXML.onLoad = function():Void{
    for(j=0;j<urlXML.firstChild.childNodes[i].
        childNodes.length;j++)
    {
        _root.mc.snip.snip_txt.text +=
            urlXML.firstChild.childNodes[i].
                childNodes[j].attributes.link + "\n";
        _root.mc.snip.snip_txt.text +=
            urlXML.firstChild.childNodes[i].
                childNodes[j].attributes.snippet + "\n\n";
    }
}

```

```

//carico le tre immagini
var imgXML:XML=new XML();
imgXML.ignoreWhite = true;
imgXML.load("listatag.xml");
imgXML.onLoad = function():Void{

    var img:MovieClip =
        _root.mc.thumb.createEmptyMovieClip ("img",
            _root.mc.thumb.getNextHighestDepth());
    img.loadMovie(imgXML.firstChild.childNodes[i].
        childNodes[0]. attributes.thumb);

    var img1:MovieClip =
        _root.mc.thumb1.createEmptyMovieClip ("img1",
            _root.mc.thumb1.getNextHighestDepth());
    img1._x= img._x + 30;
    img1.loadMovie(imgXML.firstChild.childNodes[i]
        .childNodes[1]. attributes.thumb);
}

```

```

var img2:MovieClip =
    _root.mc.thumb2.createEmptyMovieClip ("img2",
    _root.mc.thumb2.getNextHighestDepth());
img2._x= img1._x + 30;
img2.loadMovie(imgXML.firstChild.childNodes[i]
    .childNodes[2]. attributes.thumb);
}

```

Nel caso in cui il mouse si trovi all'interno di tutto l'mc è possibile effettuare tre operazioni:

- selezionare il tag prescelto
- selezionare l'indirizzo web o snippet preferito
- aprire una nuova pagina web cliccando sopra ad una delle tre immagini

Quest'ultima azione viene descritta con un frammento di codice riportato di seguito:

```

var tagXML:XML=new XML();
tagXML.ignoreWhite = true;
tagXML.load("listatag.xml");
tagXML.onLoad = function():Void{
    for(u=0;u<tagXML.firstChild.childNodes.length;u++)
    {
        if(_root.mc.tag.tag_txt.text ==
            tagXML.firstChild.childNodes[u].attributes.tname)
        {
            getURL(tagXML.firstChild.childNodes[u].
                childNodes[1].attributes.link);
            break;
        }
    }
}

```

Viene caricato innanzitutto il file “listatag.xml” contenente il link di ogni immagine presente nello schermo. Quindi, se la condizione è soddisfacente, viene aperta una nuova pagina web con la funzione *getURL()*.

Qualora l’utente decida di visitare un’altra etichetta non fa altro che spostare il mouse sopra un altro tag. In questo modo si esce dall’mc precedente e le immagini, indirizzi e snippet vengono disabilitati e cancellati.

```
snip.enabled=false;  
snip._visible=false;  
snip.snip_txt.text = "";  
thumb.unloadMovie("img");  
thumb1.unloadMovie("img1");  
thumb2.unloadMovie("img2");
```


3. La visualizzazione passo passo

In questo capitolo viene descritta, passo passo, la visualizzazione finale con l'ausilio di immagini di supporto: una specie di guida utile agli utenti per capire il funzionamento dell'applicazione.

All'apertura dell'eseguibile, dopo la presentazione iniziale di qualche secondo, l'utente è in grado di vedere in alto a sinistra una data e scorrere nello schermo i cinque tag giornalieri (come da figura).



Per interagire con l'applicazione, l'utente deve spostarsi con il mouse sopra uno di questi tag. Nel momento in cui ciò accade, il tag scelto si ferma, mentre gli altri tag vengono spostati verso sinistra di qualche pixel, per non avere sovrapposizioni con scritte ed immagini.

25 Oct 2008

roma

rome

italy



kathmandu

temple

<http://en.wikipedia.org/wiki/Italy>

Italy shares its northern, Alpine boundary with France, Switzerland, Austria and . Slovenia. The independent states of San Marino and the Vatican City are ...

http://en.wikipedia.org/wiki/History_of_Italy

Nov 3, 2008 ... Italy, united in 1861, has significantly contributed to the cultural and social . development of the entire Mediterranean area. ...

http://en.wikipedia.org/wiki/Italy_national_football_team

The Italian national football team is controlled by the Federazione Italiana . Giuoco Calcio (FIGC) and represents Italy in international football competition

Come descritto nella sezione precedente, accanto al tag ora sono visibili le tre immagini ordinate per il maggior numero di visualizzazioni ottenute, mentre sotto l'etichetta appaiono delle scritte che indicano l'indirizzo web e lo snippet dei primi tre documenti reperiti.

Quando i tag giungono alla fine dello schermo e scompaiono sulla sinistra, sulla destra ne compaiono altrettanti, ma con diverso nome.

Il cambiamento dei nomi del tag è determinato dal cambio della data (particolare che si nota in alto a sinistra della schermata).

17 Dec 2008

Art

artbooks

book review

christmas

star wars

I tag scorrono nello schermo continuamente e senza interruzioni, accompagnati da un sottofondo musicale di Mozart: “*Pachelbel cannon in D minor*”.

4. Software utilizzati

In questa sezione vengono elencati i programmi e i linguaggi di programmazione utilizzati.

4.1. Microsoft Server SQL 2005 Express - SQL

Microsoft SQL Server è un DBMS (Database Management System), relazionale prodotto da Microsoft. Nelle prime versioni era utilizzato per basi di dati medio-piccole, ma, a partire dalla versione 2000, è stato utilizzato anche per la gestione di basi di dati di grandi dimensioni (rif [1], [10]).

Dalla versione gratuita di SQL Server 2005 è stata estesa la versione Express, la quale comprende anche un tool di gestione grafico (SQL Server Management Studio Express). Tra le novità più interessanti di questa versione va segnato sicuramente il formato XML come oggetto di sistema riconosciuto in modo nativo. Ciò consente di mappare il formato XML su più campi e tabelle o in un singolo campo. Con il supporto di “XQuery” e “XPath”, Sql Server è in grado di recuperare singoli dati o campi interi da un file XML.

4.2. NetBeans IDE 6.5 - Java

Java è un linguaggio di programmazione orientato agli oggetti, derivato dal C++ (e quindi indirettamente dal C) e creato da James Gosling e altri ingegneri di Sun Microsystems. La piattaforma di programmazione Java è fondata sul linguaggio stesso, sulla Macchina

virtuale Java (*Java Virtual Machine* o *JVM*) e sulle API Java (rif [1], [13], [14]).

Utilizzando l'editor di testo NetBeans IDE 6.5, viene creata un'applicazione Java che effettua delle interrogazioni al motore di ricerca Google. Per interagire con il motore di ricerca è stato necessario l'utilizzo delle API di Google.

Le API (Application Programming Interface - Interfaccia di Programmazione di un'Applicazione), sono un insieme di procedure rese disponibili al programmatore, raggruppate a formare un set di strumenti specifici per un determinato compito (rif [1]).

4.3. Macromedia Flash 8 Mx - ActionScript

Flash è un software per uso prevalentemente grafico che consente all'utente di realizzare una grande varietà di creazioni, da una semplice animazione ad una complessa applicazione Web interattiva, quale un negozio online (rif [1], [20], [21]). È possibile arricchire i supporti delle applicazioni Flash con immagini, suoni e video. Flash dispone di diverse funzioni che lo rendono particolarmente efficace e sempre di facile uso, ad esempio:

- componenti dell'interfaccia utente utilizzabili tramite trascinamento
- possibilità di mantenere simboli (clip filmato, pulsanti, oggetti bitmap, ecc.) riutilizzabili in una libreria
- comportamenti incorporati che aggiungono istruzioni di ActionScript al documento
- effetti speciali che si possono aggiungere agli oggetti.

Utilizzando un linguaggio di scripting interno, denominato ActionScript e basato su ECMAScript (JavaScript), è possibile applicare comportamenti agli oggetti o ai fotogrammi dell'animazione. Inizialmente erano presenti solo poche azioni, ma allo stato attuale ActionScript è diventato uno strumento molto potente: si possono creare menu, sistemi di navigazione, interfacce grafiche, siti web completi e giochi anche complessi.

ActionScript è un linguaggio di alto livello, non compilato come il C, ma linguaggio interpretato (o pseudocompilato). La sua esecuzione è infatti affidata al compilatore nativo di Flash Player, l'Actionscript Virtual Machine (AVM) (rif [1], [22]).

5. Conclusioni

Prima di effettuare la scelta definitiva sull'argomento del progetto, è stata di grande utilità la consultazione di documenti ed articoli tratti dalle conferenze del World Wide Web (rif. [a]), Special Interest Group on Information Retrieval (rif. [b], [c], [d], [e], [f]), Computer Supported Cooperative Work (rif. [g], [h], [i]), HyperText (rif. [l], [m]), Knowledge Discovery & Data Mining (rif. [n]).

Nel perseguire l'obiettivo prefissato in questo lavoro, ritengo che la progettazione e la scelta degli strumenti software da utilizzare sia stata sicuramente l'attività più importante e teorica. E' infatti difficile conoscere a priori le difficoltà che si possono incontrare durante lo sviluppo dell'applicazione. Alla fine si è scelto di adottare i seguenti elementi:

- SQL Server 2005 Express - SQL
- NetBeans IDE 6.5 - Java
- Macromedia Flash 8 MX - ActionScript

Il dataset utilizzato, creato manualmente all'uopo, è di dimensioni ridotte, ciò nonostante non ha influito negativamente sugli aspetti concettuali e progettuali.

L'applicazione produce sicuramente un risultato più apprezzabile con l'utilizzo di una struttura dati più consistente. Può essere ottimizzata nei risultati della ricerca verificando solo la compatibilità di alcuni campi XML del file iniziale.

L'innovazione rispetto al progetto di Yahoo! Research è stata l'introduzione di un motore di ricerca in grado di reperire e abbinare al tag i documenti più rilevanti dal web attraverso il portale Wikipedia.

Questo permette all'utente, non solo di vedere graficamente la rappresentazione e visualizzazione di un flusso di etichette, interagendo con esse grazie alla visione di effetti e caratteristiche multimediali, ma anche la possibilità di usufruire di una serie di documenti testuali utili a sviluppare, se necessario, il significato o i vari significati che ogni singolo tag può avere.

Un progetto appetibile, per il futuro prossimo, è sicuramente la gestione dell'intero pacchetto on-line e in real-time; ciò consente agli utenti di verificare, su una pagina web in tempo reale, eventuali cambiamenti o nuovi inserimenti di etichette e foto effettuati nel portale Flickr.

Bibliografia

- [1] <http://en.wikipedia.org>
Vengono reperiti i contenuti testuali utili all'applicazione finale, e le definizioni di base per descrivere alcuni elementi ed oggetti descritti nella tesi.

- [2] <http://del.icio.us>
Sito di social bookmarking, è stato utile per verificare il funzionamento delle etichette applicate a indirizzi web.

- [3] <http://www.flickr.com>
Portale dal quale sono state estratte le immagini e le informazioni correlate, ed al quale l'applicazione finale si appoggia per la descrizione di alcune di queste immagini.

- [4] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P Raghavan, A. Tomkins. Visualizing Tags over Time. WWW 2006. May 23-26, 2006.
Articolo tratto dalla conferenza di Edinburgo del World Wide Web del Maggio 2006. Grazie a questo articolo è stata redatta la tesi.

- [5] <http://research.yahoo.com>
Rappresentazione dell'applicazione create e gestita da Yahoo! Reserch.

- [6] <http://www.google.it>
A questo portale si appoggia l'applicazione Java. È stato inoltre utilizzato come motore di ricerca per reperire molte indicazioni per lo svolgimento dela tesi.

- [7] <http://www2.computer.org/portal/web/csdl/doi/10.1109/TKDE.2004.60>
Accenni e spiegazioni dell'algoritmo di soglia (Threshold Algorithm)

- [8] <http://www.isti.cnr.it>
Portale utile a trovare le informazioni necessarie per la compilazione del file XML iniziale.

- [9] <http://www.fabsoft.it/default.aspx?item=xml>
Documento che tratta specifiche riguardanti l'XML.

- [10] <http://database.html.it/guide/leggi/133/guida-sql-server-2005/>
Linee guida nell'utilizzo di sql server.
- [11] <http://msdn.microsoft.com/it-it/library/ms187897.aspx>
Linee guida sull'utilizzo delle funzioni di lettura e scrittura di un file XML con SQL Server.
- [12] <http://www.dotnethell.it/articles/XML-SQL-Server-2005.aspx>
Linee guida sull'utilizzo delle funzioni di gestione di un file XML con SQL Server.
- [13] <http://java.com>
Portale di consultazione e download delle applicazioni Java.
- [14] <http://java.sun.com>
Portale di consultazione e download delle applicazioni Java.
- [15] http://www.codicefacile.it/tutorials/tutorials.php/45/JDOM_file_XML_e_Java_Introduzione_e_lettura_file/category/
Documentazione relativa all'utilizzo delle API JDOM con Java.
- [16] <http://code.google.com/intl/it-IT/apis/gdata/client-java.html>
Documentazione relativa all'utilizzo delle API Google con Java.
- [17] <http://www.stanford.edu/class/cs276a/projects/docs/googleapi/com/google/soap/search/GoogleSearchResultElement.html>
Documentazione relativa all'utilizzo delle API Google con Java.
- [18] <http://users.design.ucla.edu/~tatsuyas/tools/google/index.htm>
Documentazione relativa all'utilizzo delle API Google con Java.
- [19] <http://www.cs.usfca.edu/~wolber/SoftwareDev/Distributed/WebServices/Google/googleJavaTutorial.htm>
Documentazione relativa all'utilizzo delle API Google con Java.
- [20] www.macromedia.com
Portale di consultazione di alcuni contenuti Flash
- [21] www.adobe.com
Portale di consultazione di alcuni contenuti Flash
- [22] www.actioscript.it
Portale di consultazione di alcuni contenuti del linguaggio di programmazione ActionScript

- [23] <http://www.webmasterpoint.org/appunti/flash/115.asp>
Portale di consultazione di alcune funzioni del linguaggio di programmazione ActionScript
- [24] <http://livedocs.adobe.com/>
Portale di consultazione di alcuni contenuti del linguaggio di programmazione ActionScript

Integrazioni da conferenze informatiche internazionali

- [a] WWW 2007: P-TAG: Large scale automatic generation of personalized annotation tags for the web
P-TAG è un metodo di generazione automatica di tag personalizzabili per le pagine web. P-TAG produce delle parole chiave rilevanti sia per il suo contenuto testuale che per le informazioni residenti nel Desktop dei navigatori.
- [b] SIGIR 2006: Social Networks, Incentives, and Search
- [c] SIGIR 2008: Real-Time Automatic Tag Recommendation
Le relazioni tra documenti, tag e parole vengono schematizzati in due grafici bipartiti. Viene spiegato l'algoritmo Lanczos e lo Spectral Recursive Embedding (SRE). Viene proposto un nuovo algoritmo di ranking dei tag applicando un modello di Poisson. Esperimenti in CiteULike (sito di documenti scientifici) e delicious (sito di web page)
- [d] SIGIR 2008: Social Tagging Prediction
Data una serie di oggetti e di Tag applicati a questi oggetti dagli utenti puoi prevedere se dato un Tag, esso può o potrebbe essere applicato a un particolare oggetto? Algoritmo che verifica quali tag vengono riferiti a particolari oggetti. Si parla di entropia.
- [e] SIGIR 2008: To Tag or not to Tag – Harvesting adjacent metadata in large scale tagging system
Viene presentata una nuova tecnica di inferenza sui tag, assegnando tag a documenti precedentemente “non taggati” o estendere una serie di tag a documenti già “taggati” per propagare l'esistenza di tag da un documento ad un documento simile. È utilizzato l'algoritmo TOP-K. Esempio in CiteULike.
- [f] SIGIR 2008: Detecting Synonymis in social tagging system to improve content retrieval
Problema Sinonimi: le persone usano termini diversi per descrivere un contenuto. Si propone un nuovo metodo per identificare i sinonimi “veri” nei sistemi di contenuti sociali
- [g] CSCW 2006: Tagging, communities, vocabulary, evolution

- [h] CSCW 2006: What goes around comes around: an analysis of delicious as social space
- [i] CSCW 2006: Shared Waypoints and social tagging to support collaboration in software development
- [l] HT 2006: Tagging paper, taxonomy, Flickr, academic articles, to read
- [m] HT 2008: Understanding the efficiency of social tagging systems using information theory
Capire come funziona un sistema di social tagging, in particolare la costruzione di un deposito sociale di oggetti dei quali gli utenti trovano interessanti e dagli stessi evolvere un vocabolario condivisibile per descrivere gli oggetti stessi. Si suggerisce la teoria dell'informazione come struttura naturale nella quale evolvere l'efficienza dei vocabolari generati dai siti di social tagging. Si forniscono misure di informazioni teoretiche sia processi di codificazione che di reperimento. Viene fondato il lavoro con esempi di analisi su siti popolari di tagging come delicious.
Risultati: con l'aumento della popolarità del social tagging l'efficacia dei tag riferiti a oggetti individuali è in calo. Le misure introdotte possono essere direttamente usate per aiutare l'incremento dell'efficacia di siti di social tagging.
- [n] KDD 2006: Structure and evolution of online Social Networks