

Pick your movie



Sommario:

1) INTRODUZIONE

1.1) IDEA

1.2) OBIETTIVO DELL'APPLICAZIONE

1.3) DESCRIZIONE DELL'APPLICAZIONE

2) STATO DELL'ARTE

3) GRAFICA E MULTIMEDIALITÀ

4) FONTI

5) DATABASE

5.1) COMPOSIZIONE

5.2) PROBLEMI AFFRONTATI

6) TECNOLOGIA UTILIZZATA

7) ACCESSIBILITÀ E USABILITÀ

8) CONCLUSIONI: LIMITI E POTENZIALITÀ

Introduzione

L'applicazione web “**Pick your movie**” è stata realizzata da Roberta Voleri e Maurizio Ricci come progetto per l'esame di Laboratorio e Progettazione Web, secondo modulo del corso da 12 CFU di Basi di Dati e Laboratorio Web, tenuto dalle docenti Anna Monreale e Giovanna Rosone (modulo di Basi di Dati), Andrea Marchetti e Angelica Lo Duca, coadiuvati dall'assistente Federico Martinelli (modulo di Laboratorio Progettazione Web).

Idea:

Come utenti di siti web che offrono raccolte, classifiche e giudizi di film prefissati, ci siamo chiesti in quale modo avremmo potuto migliorarne ed ottimizzarne il funzionamento. Abbiamo quindi pensato che sarebbe potuto essere interessante realizzare un'applicazione web che potesse estendere le funzionalità dei siti già esistenti, lasciando all'utente un maggior controllo sul risultato finale.

Obiettivo dell'applicazione:

“**Pick your movie**” è un'applicazione web nata con lo scopo di offrire agli utenti vari servizi specifici, tra cui: la possibilità di cercare un film adatto alle proprie esigenze, cercare tra i film più belli del proprio attore preferito, rivivere i migliori film di sempre, osservare l'evoluzione del cinema scorrendo i titoli per anno oppure lasciare che la dea bendata lo guidi nella scelta.

Descrizione dell'applicazione

Inizialmente all'utente viene presentata la pagina principale, **Home**, nella quale è possibile cliccare su 3 diverse icone per recarsi nelle varie sezioni dell'applicazione. Una volta scelta la sezione desiderata, è possibile navigare all'interno di quella pagina o spostarsi orizzontalmente nel sito mediante un menù a comparsa laterale. Le sezioni del sito sono 4:

1. **Suggest Me a Movie**: sezione dell'applicazione web destinata a quella categoria di utenti che ha voglia di guardare un film ma non possiede un'idea precisa su quale scegliere. Cliccando sull'icona è possibile farsi consigliare un film in maniera casuale, ma non troppo, infatti per poter concorrere all'estrazione, i film devono possedere un giudizio medio non inferiore a 3/5.

2. **Find A Movie**: sezione dell'applicazione web destinata a quella categoria di utenti che ha un'idea più o meno chiara circa il tipo di film desiderato. Cliccando sull'icona, l'utente si reca in una pagina web che gli dà la possibilità di applicare vari filtri alla propria ricerca, ossia: titolo, attore, tag, genere e anno. I vari filtri possono essere combinati tra loro, oppure no. Nel caso in cui l'utente non selezioni alcun filtro, gli verrà mostrato tutto l'elenco dei film, suddiviso in varie pagine, con un certo numero di elementi.
3. **Best Movies Ever**: sezione dell'applicazione web destinata a quella categoria di utenti che vuole il meglio del meglio, infatti permette di ottenere la classifica dei film presenti nel database secondo alcuni criteri, ossia: migliori film di sempre, migliori film per attore, migliori film per anno, migliori film per genere.
4. **Database Composition**: sezione dell'applicazione web che offre uno sguardo più tecnico sulla bontà dei risultati. Qui l'utente trova una panoramica sul numero dei titoli presenti e sulle recensioni globali. Sono presenti due grafici: il primo elenca il numero di film presenti suddivisi per anno, il secondo elenca il numero di recensioni presenti suddivise per anno. Dai due grafici presenti nella pagina è possibile notare l'uniformità dei dati, in quanto seguono lo stesso andamento. Sono infatti presenti due picchi negli anni 1994 – 1996, sia nel primo che nel secondo grafico, anche se dopo un'attenta analisi è stato dimostrato che non sono dovuti a qualche particolare ragione.

Tutti i film risultanti dalle varie sezioni offrono la locandina del film ed il titolo. Con un click sulla locandina del film è possibile scoprire altri dettagli, quali: anno, genere e voto medio. Nel caso il film stuzzichi l'interesse dell'utente, è possibile ottenerne una breve trama facendo passare il mouse sopra la locandina del film in questione.

Stato dell'arte:

IMDb o Internet Movie Database: è un sito web di proprietà di Amazon.com che gestisce informazioni su film, attori, registi, personale di produzione, programmi televisivi, e anche videogiochi. Su IMDb è presente un'ampia gamma di informazioni su ciascuna opera che comprende informazioni essenziali quali quelle sul cast di attori e registi, il sunto delle trame e le recensioni. Inoltre è possibile reperire informazioni su argomenti più specifici, come l'elenco del cast tecnico coinvolto nella produzione e nello sviluppo del film, gli errori durante il girato, le incongruenze storiche o temporali, l'elenco degli eventuali brani della colonna sonora, il formato, le eventuali versioni

alternative e altro ancora. Gli attori, i registi e gli sceneggiatori, così come altri membri della troupe, hanno pagine a loro dedicate nelle quali sono elencati i film e i programmi ai quali hanno lavorato e, spesso, anche una biografia. È presente anche una pagina dedicata ai "personaggi", con una loro descrizione e con l'elenco dei film e delle serie TV in cui sono comparsi e gli attori da cui sono stati interpretati. Ora, avendo un sito così preciso e dettagliato a disposizione, perché l'utente dovrebbe scegliere il nostro? La risposta è semplicissima. IMDb possiede due tipi di classifiche: **Top 250** una classifica dei primi 250 film di tutti i tempi votati dagli utenti registrati al sito, **Bottom 100** una classifica negativa realizzata con i medesimi criteri della Top 250 ma per i film col voto più basso. Il nostro sito invece permette agli utenti di personalizzare le classifiche in base ai propri criteri.

MYmovies.it: è un magazine e database italiano online di cinema edito da Mo-Netj S.r.l, società del Gruppo Editoriale L'Espresso. Le principali sezioni del sito sono: Database (tra cui dizionari, film, premi), Cinema (tra cui Oggi al cinema, prossimamente), Programmi TV (tra cui digitale terrestre, Sky, Tv News), Dvd & CD (tra cui novità in dvd, colonne sonore), Attualità (tra cui box office), Multimedia (tra cui trailer), Community. I limiti di questo sito rispetto al nostro sono relativi al fatto che possiede un'interfaccia poco intuitiva e che la ricerca è poco mirata, trattandosi di articoli.

Nientepopcorn.it: forse l'unico sito che si avvicina alla struttura del nostro, in base alle classifiche. L'unico limite rispetto al nostro è che la navigazione è molto più lenta, per ottenere un'informazione, l'utente è costretto a dover aprire varie pagine.

Suggest Me Movie: è un'applicazione web molto simile alla nostra, infatti in esso è possibile farsi consigliare un film in modo casuale, oppure cercarne uno applicando alcuni filtri. Ha un vantaggio rispetto al nostro, ossia quello di avere un collegamento video a youtube che permette di vedere il trailer di ogni film, però ha anche un limite che sta nel fatto che non offre un servizio di visualizzazione delle classifiche.

In generale, esistono moltissimi siti internet e applicazioni web che hanno i film come tema di fondo, però nessuno di essi è completo di: suggerimento casuale, ricerca mirata, classifiche ordinate per criteri e informazioni sui dati.

Grafica e multimedialità:

L'applicazione è rappresentata nel 65% dello schermo, non viene sfruttato quindi l'intero spazio. La prima e l'ultima sezione sono costituite da due

banner (uno per l'header e uno per il footer) che rappresentano rispettivamente il nome dell'applicazione, i dovuti crediti ai creatori di essa e sono state realizzate entrambe attraverso l'uso di Photoshop. La sezione centrale è il cuore dell'applicazione in quanto cambia nelle diverse pagine e ospita i vari contenuti:

- **Home**: troviamo 3 icone che rappresentano le varie sezioni dell'applicazione che si possono utilizzare. Le icone sono state realizzate con Photoshop a partire da un'immagine in jpeg che dopo essere stata scontornata e convertita in png è stata completata con le rispettive scritte che riportano il nome delle varie sezioni. Passando il mouse sopra le icone si ha una breve animazione realizzata con transition, una proprietà di css3. Lateralmente abbiamo una sorta di menù a comparsa che porta l'utente nella 4^a sezione del sito che abbiamo deciso di non inserire nelle 3 icone principali perché non è nata con lo scopo di essere uno strumento dell'applicazione ma come approfondimento sul database.
- **Suggest Me A Movie**: troviamo in posizione centrale il film casuale che l'applicazione ha scelto. Passando il mouse sulla locandina del film appare una breve descrizione il cui sfondo semi-trasparente permette comunque di continuare a vedere la locandina del film, mentre cliccandoci sopra appaiono: anno, genere, voto medio sotto di essa. Lateralmente abbiamo un vero e proprio menù a comparsa che permette di navigare nelle varie sezioni dell'applicazione oppure di ottenere un nuovo suggerimento cliccando su "Another One!".
- **Find A Movie**: troviamo un form con vari campi da compilare: titolo, attore e tag, poi abbiamo un menù a tendina per il genere e l'anno ed i tasti per il reset e la ricerca che contengono una piccola animazione che si ha al passaggio del mouse: i tasti inizialmente sono bianchi con bordo blu, passando il mouse sopra si riempiono del colore del bordo. Dopo la ricerca otteniamo i risultati che sono locandine di film con il loro titolo e qualora il numero di film sia superiore a 40, sul fondo della pagina si crea un menù a tendina che permette di scorrere le varie pagine. Passando il mouse sulla locandina del film appare una breve descrizione il cui sfondo semi-trasparente permette comunque di continuare a vedere la locandina del film, mentre cliccandoci sopra appaiono: anno, genere, voto medio sotto di essa. Lateralmente abbiamo un vero e proprio menù a comparsa che permette di navigare

nelle varie sezioni dell'applicazione oppure di effettuare un'altra ricerca cliccando su "Torna su".

- **Best Movies Ever:** troviamo immediatamente un menù a tendina ed una lista di film presentati sempre attraverso le loro locandine e titoli. Il menù a tendina è composto da:
 - **Migliori film;**
 - **Migliori film per attore:** attraverso la sua selezione si apre un form da compilare ed i tasti reset e ricerca che hanno la stessa animazione citata precedentemente in "Find A Movie";
 - **Migliori film per anno:** attraverso la sua selezione si apre un altro menù a tendina che permette di selezionare l'anno che vogliamo scegliere ed i tasti reset e ricerca che hanno la stessa animazione citata precedentemente in "Find A Movie";
 - **Migliori film per genere:** attraverso la sua selezione si apre un altro menù a tendina che permette di selezionare il genere che vogliamo scegliere ed i tasti reset e ricerca che hanno la stessa animazione citata precedentemente in "Find A Movie".

Passando il mouse sulla locandina del film appare una breve descrizione il cui sfondo semi-trasparente permette comunque di continuare a vedere la locandina del film, mentre cliccandoci sopra appaiono: anno, genere, voto medio sotto di essa. Lateralmente abbiamo un vero e proprio menù a comparsa che permette di navigare nelle varie sezioni dell'applicazione oppure di effettuare un'altra ricerca cliccando su "Torna su".

- **Application Details:** è forse la sezione più diversa del sito in termini grafici, perché presenta unicamente i due grafici citati già nella descrizione dell'applicazione, realizzati con Highcharts, che prevedono una selezione random dei colori, infatti ogni volta che la pagina viene ricaricata, il grafico assume un colore diverso.

Fonti:

Il dataset è stato fornito da grouplens.com. GroupLens è un laboratorio di ricerca del Dipartimento di Informatica e Ingegneria presso l'Università del Minnesota, città specializzata in recommender systems, community online, tecnologie mobili e onnipresenti, biblioteche digitali e sistemi informativi territoriali locali. Il loro obiettivo è avanzare la teoria e la pratica del social

computing attraverso la costruzione e la comprensione dei sistemi usati da persone reali.

Link: <http://grouplens.org/datasets/movielens>

Le locandine dei film, le loro descrizioni e l'elenco degli attori sono stati ottenuti mediante scraping da IMDb, sito già precedentemente citato nella sezione "Stato dell'arte".

Link: <http://www.imdb.com/>

Database:

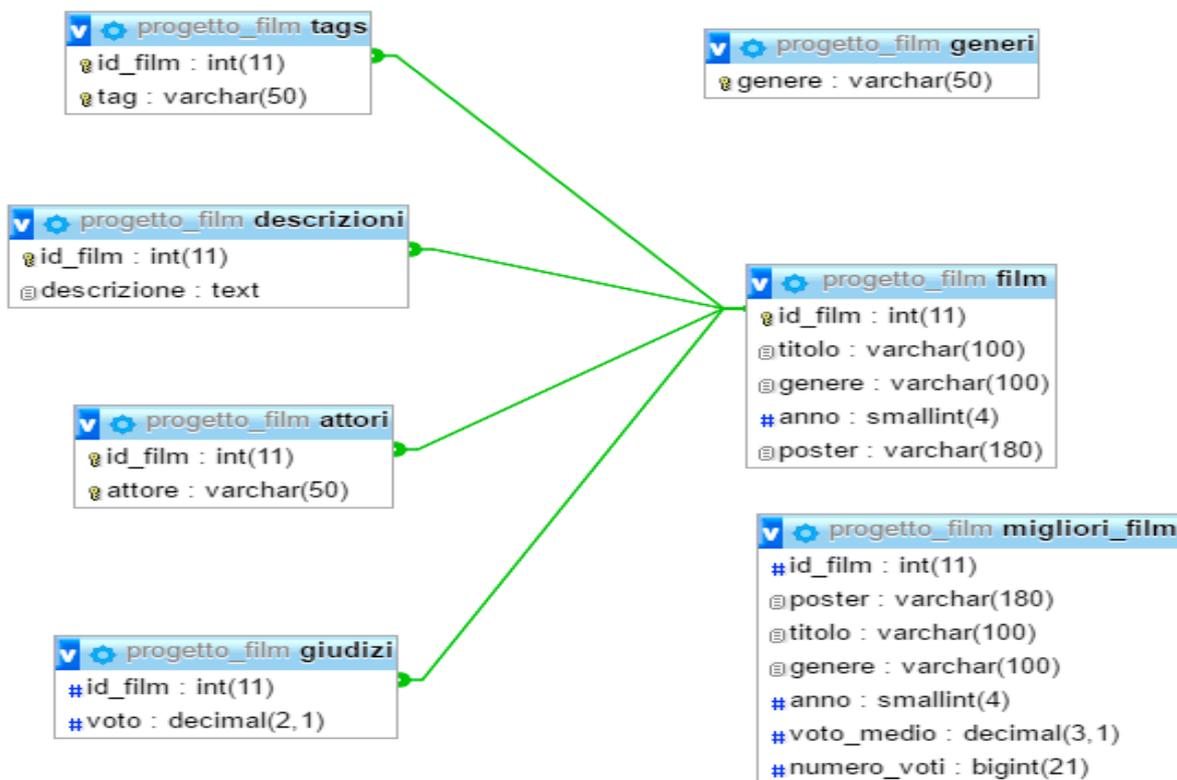
Per quanto riguarda il dataset originale ottenuto da GroupLens, offriva 4 tabelle:

- **Movies:** tabella in cui sono salvate le informazioni sui film. Informazioni presenti: codice film, titolo e genere della pellicola.
- **Ratings:** tabella sono registrati i giudizi assegnati dagli utenti ai film. Informazioni presenti: codice utente, codice film, voto e orario in cui viene registrato il voto assegnato.
- **Tags:** tabella contenente i tag che gli utenti hanno scritto per le pellicole. Informazioni presenti: codice utente, codice film, tag e orario in cui viene registrato il tag.
- **Link:** qua sono forniti i link delle pellicole ai siti: imdb (<http://www.imdb.com/>) e tmdb (<https://www.themoviedb.org/>). Informazioni presenti: codice film, ID IMDb e ID tmdb

Il database originario integrava gli ID dei film ma non i link. Una volta scoperta la correlazione tra ID e link alla pagina web è bastato far ricorso a delle espressioni regolari per poter avere i link alle rispettive pagine web IMDb. Lo scraper è stato scritto in python e attraverso la libreria BeautifulSoup è possibile eseguire il parsing di una pagina html da cui estrarre le informazioni.

Composizione:

Table	Action	Rows	Type	Collation	Size	Overhead
attori	★ Browse Structure Search Insert Empty Drop	31,354	InnoDB	utf8_general_ci	1.5 MiB	-
descrizioni	★ Browse Structure Search Insert Empty Drop	2,116	InnoDB	utf8_general_ci	416 KiB	-
film	★ Browse Structure Search Insert Empty Drop	2,116	InnoDB	utf8_general_ci	560 KiB	-
generi	★ Browse Structure Search Insert Empty Drop	18	InnoDB	utf8_general_ci	16 KiB	-
giudizi	★ Browse Structure Search Insert Empty Drop	297,205	InnoDB	utf8_general_ci	15 MiB	-
migliori_film	★ Browse Structure Search Insert Drop	~0	View	---	-	-
tags	★ Browse Structure Search Insert Empty Drop	73,062	InnoDB	utf8_general_ci	6 MiB	-
7 tables	Sum	~405,871	InnoDB	utf8_general_ci	23.5 MiB	0 B



Il database si compone di 7 tabelle:

- **Attori:** tabella che serve a memorizzare le interpretazioni degli attori nei film. Ogni riga della tabella è una coppia attore-film. Lo stesso attore non può partecipare più volte nello stesso film.
- **Giudizi:** tabella in cui vengono salvati i voti relativi ai film. I giudizi sono espressi in una scala da 1 a 5 mentre ogni riga della tabella è una coppia film-voto. Possono esserci più giudizi per lo stesso film.

- **Film**: tabella contenente le informazioni inerenti alle pellicole. Le informazioni presenti sono: codice univoco del film, titolo, genere, anno e relativo link alla locandina della pellicola. Possono essere presenti più film con lo stesso titolo, dovranno necessariamente però avere dei codici identificativi diversi.
- **Descrizioni**: tabella contenente le descrizioni delle pellicole. Quest'ultime potranno far riferimento a una e una sola descrizione.
- **Tags**: tabella contenente i tag delle pellicole, ovvero delle brevissime parole che rappresentano al meglio il titolo in questione. Più tag possono appartenere allo stesso film.
- **Generi**: tabella in cui sono raccolti tutti i generi distinti a cui le pellicole possono appartenere. La tabella serve a permettere la ricerca di film appartenenti tutti a uno stesso genere. Per capire bene l'utilità della tabella in questione dobbiamo però osservare una cosa: i generi dei film separati da una barra verticale (pipe) sono stati inseriti all'interno della tabella **Film** dai creatori del dataset; i generi però possono essere molti per ogni titolo, così per non costringere a molte più join superflue si è scelto di lasciare i generi all'interno dei film e di creare una nuova tabella contenente i generi distinti.
- **Migliori film**: vista contenente i film ordinati in modo discendente per voto medio.

Per i dati ottenuti mediante scraping è stato scritto uno script in python. Il programma legge in input una lista di indirizzi web di IMDb, esegue il parsing della pagina html mediante uso della libreria Beautiful Soup (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>). Il programma ottiene così le informazioni desiderate e in output stampa il file csv il quale viene importato nella tabella appropriata. Sotto è riportato una delle 3 varianti di script usati, esso serve a estrarre l'elenco degli attori dalle pagine dei film.

```

import sys
reload(sys) # Reload does the trick!
sys.setdefaultencoding('UTF8')
from bs4 import BeautifulSoup
import urllib

def readfile(fileName):
    try:
        with open(fileName) as f:
            content = f.readlines()
            return content
    except:
        return []

def writeCSV(fileName, id_film, actors):
    out_file = open(fileName, "a")
    id_film=id_film.replace('\n', ").replace('\r', ")
    for actor in actors:
        out_file.write(id_film + ";" + actor + "\n")
    out_file.close()

def writeLog(fileName,text):
    out_file = open(fileName, "a")
    out_file.write(text + "\n")
    out_file.close()

def scraper(url,actors):
    try:
        soup = BeautifulSoup(urllib.urlopen(url).read())
        table=soup.find_all('table', class_="cast_list")[0]
        elements=table.find_all('tr', {'class':['odd', 'even']})
        for el in elements:
            x=el.find_all("td", class_="itemprop")[0]
            info=x.a.find_all("span", class_="itemprop")[0].text
            actors.append(info)
    except:
        return []

def main():
    actors=[]
    link = readfile("link.txt") # carica le url dei film
    id_film = readfile("id_film.txt") # carica gli id dei film
    fatti = readfile("actorsCSV.txt")
    idUltimoFilm=fatti[len(fatti)-1].split(";",1)[0] #prendo l'id dell'ultimo film
    ultimaPosizione=0
    for c in id_film:
        if int(c)==int(idUltimoFilm): #trovo la posizione dell'ultimo id da cui ripartire
            print(1)
            break
        else:
            ultimaPosizione+=1
    ultimaPosizione+=1 #l'estremo inferiore e' incluso nel range ma non voglio farlo due volte
    m = len(link)

    for c in range(ultimaPosizione,m):
        scraper(link[c],actors)
        if actors!=[]:
            writeCSV("actorsCSV.txt",id_film[c],actors)
        else:
            writeLog("errors.txt"Problema con url: "+link[c]+" film id: "+id_film[c])
            print(str(round((c*1.0/m)*100, 2)) + "%")
            actors=[]

main()

```

Problemi affrontati:

Il database originale era eccellente, ma presentava l'inconveniente di avere l'anno dei film inserito all'interno del titolo stesso. Attraverso l'uso delle espressioni regolari l'anno è stato estratto e gli è stata dedicata una sezione a parte. Sempre usando le espressioni regolari, sono stati eliminati campi extra come quello dedicato all'utente che aveva assegnato un voto a una data pellicola. In particolare, il campo utente è stato rimosso perché per

motivi di privacy era rappresentato unicamente dal codice identificativo dell'utente all'interno del database; non esisteva quindi una tabella Utente.

Il vero problema affrontato è inerente ai dati, perché la vera difficoltà non è trovarli ma processarli in tempo accettabile. Abbiamo così deciso di alleggerire il database:

Il database originale offriva 20 milioni di giudizi e 27.000 film, mentre noi lo abbiamo ridotto a 300.000 giudizi e 2.000 film, scartando film più vecchi del 1970 e dividendo più volte in 4 la tabella dei giudizi. I film con meno di 50 recensioni sono stati rimossi ad eccezione di quelli del 2015 (abbiamo scelto di tenerli tutti poiché il dataset, essendo stato compilato a inizi 2016, offriva poche recensioni per i titoli più recenti). A questo punto si è notato che alcuni titoli più famosi avevano sempre tante recensioni (alcuni avevano 500-800-6000 recensioni), quindi abbiamo cercato di ridurre il numero di giudizi per tali film cancellando a intervalli di due le recensioni (1 riga sì e una no). Come risultato finale il numero di giudizi per film è compreso tra 100 e 200, che ci dà sempre una buona stima del voto medio della pellicola consentendoci di alleggerire il database.

Sempre per velocizzare le query, abbiamo cercato di studiare l'esecuzione di esse mediante il comando EXPLAIN, per scoprire dove poter piazzare gli indici per massimizzare le performance. È stato inoltre possibile salvare un po' di memoria impostando il tipo più appropriato di dato da memorizzare (es. il campo relativo al voto espresso per una pellicola, invece che essere *float* è stato reso *decimal* e abbiamo risparmiato circa 2Byte per ogni riga, qualche MB in generale).

Tecnologia utilizzata:

Il sito è stato sviluppato all'interno della piattaforma software gratuita XAMPP. Sono stati adoperati:

- **MariaDB:** è un fork di MySQL ed è il DBMS predefinito di XAMPP
- Phpmyadmin, applicazione web scritta in PHP che permette di gestire **MYSQL:** tramite un qualsiasi browser e che consente di creare un database, delle tabelle ed eseguire operazioni sulle stesse
- **jQuery:** libreria JavaScript che permette di manipolare facilmente il DOM per realizzare pagine dinamiche, animazioni, gestisce inoltre

facilmente gli eventi e permette di lavorare con strutture dati di tipo JSON ottenute dal server mediante richieste GET

- **PHP**: linguaggio di programmazione per l'interrogazione del database
- **HighCharts**: libreria JavaScript per la realizzazione di grafici interattivi

Accessibilità e usabilità:

L'applicazione web risponde al requisito di accessibilità perché le informazioni ed i componenti dell'interfaccia utente sono fruibili e percepibili, fornisce informazioni contestuali e di orientamento, fornisce chiari meccanismi di navigazione, il testo è leggibile e comprensibile, alcune funzionalità sono disponibili da tastiera, è compatibile con tutti i browser standard.

L'applicazione web risponde al requisito dell'usabilità perché è facilmente navigabile, le informazioni sono facilmente raggiungibili e presentate in modo efficace ed intuitivo, l'utente riesce ad orientarsi facilmente e a recuperare le informazioni in tempi brevi e modi diretti, la struttura di navigazione è ripetuta, gli argomenti sono organizzati in maniera gerarchica.

Conclusioni: limiti e potenzialità:

Il sito si pone l'obiettivo di dispensare consigli o giudizi, i quali però potrebbero non essere accurati con estrema precisione data la limitatezza dei dati.

Solo sognando un poco potremmo arrivare a immaginare di eseguire l'applicazione su server un po' più potenti in modo da poter offrire contenuti di qualità sempre più elevata all'utente.

Allora la nostra applicazione potrebbe diventare uno strumento adatto a ogni utente, ma fino ad allora il progetto rimarrà un'idea valida e già sviluppata, bloccata dalla mancanza di risorse.

Big data is better data!

Dato che l'applicazione web tratta di film, abbiamo deciso di concludere questa relazione con una perla proveniente da un grande film:

“Non permettere mai a nessuno di dirti che non sai fare qualcosa.

Se hai un sogno tu lo devi proteggere.

Quando le persone non sanno fare qualcosa

lo dicono a te che non la sai fare.

Se vuoi qualcosa, vai e inseguila.

Punto.”

- La Ricerca della Felicità