



UNIVERSITÀ DEGLI STUDI DI SALERNO

DAVIMUS

DISCIPLINE DELLE ARTI VISIVE, DELLA MUSICA E DELLO SPETTACOLO

Corso di **INFORMATICA PER LE ARTI DELLO SPETTACOLO**

“Libreria Netflix”

Relazione sulle potenziali applicazioni del linguaggio XML nella catalogazione di contenuti multimediali

Professore: Francesco Colace

Studenti: Arianna De Vito, Danilo Fusco

INTRODUZIONE

Questa relazione si pone come obiettivo quello di validare i passaggi effettuati nella presentazione PowerPoint precedentemente svolta, ovvero, la catalogazione di contenuti multimediali attraverso l'uso di un documento in linguaggio XML (Extensible Markup Language) e relativo DTD (Document Type Definition).

In un mondo oramai del tutto digitalizzato, dove le piattaforme streaming si presentano di continuo sul mercato e crescono sempre di più, abbiamo scelto di prendere in esame, come esempio di contenuto multimediale, la catalogazione di una generica libreria Netflix.

I film presi in esame sono i seguenti:

- *“The Irishman”*, di Martin Scorzese con Al Pacino e Robert DeNiro, distribuito sulla piattaforma di streaming nel Dicembre 2019;
- *“Her”*, di Spike Jonze con Joaquin Phoenix, distribuito nelle sale cinematografiche nel 2014 e inserito nel catalogo di Netflix Italia a Novembre 2019.

Il primo passo è stato quello di definire, insieme alle loro caratteristiche specifiche, cosa siano un linguaggio XML e un linguaggio DTD.

L'**XML** è un metalinguaggio che consente la rappresentazione di documenti e dati complessi su supporti digitali. Esso presenta alcune caratteristiche molto importanti, come quello di essere un *documento chiaro, facilmente intuibile dall'essere umano e case sensitive* (nei nomi dei tag distingue tra maiuscole e minuscole).

Il **DTD** è, invece, un linguaggio che definisce la grammatica; descrive la composizione degli elementi costituenti una certa classe di documenti XML ed è quindi volto a giustificare l'esistenza.

- Il DTD è stato programmato in questo modo:

```
<!ELEMENT Libreria Netflix (Film+)>
<!ELEMENT Film (Titolo, Regista, Attore+, Genere*)
<!ELEMENT Titolo (#PCDATA)>
<!ELEMENT Regista (#PCDATA)>
<!ELEMENT Attore (#PCDATA)>
<!ELEMENT Genere (#PCDATA)>
<!ATTLIST Film cod classificazione ID #REQUIRED
cod catalogazione #REQUIRED
anno CDATA #IMPLIED
formato (HD|STANDARD) #REQUIRED
originale netflix (SI|NO)>
```

Come si può notare, siamo in presenza di una libreria Netflix come *elemento radice*, contenuto all'interno di due *tag* (o coppia di marcatori). Questa libreria può contenere al suo interno uno o più film, ciò viene definito dalla dicitura **Film+** ; ogni film è descritto da un Titolo, un Regista, uno o più Attore (**Attore+**) e zero o più Genere (**Genere***). Per ogni elemento è importante mantenere il singolare.

Un film è poi dotato di alcune proprietà, ovvero *gli attributi*, quali: un **codice di classificazione** di tipo ID, cioè di tipo identificatore; questo attributo è **#REQUIRED**, ciò ci indica che esso deve essere obbligatoriamente presente; è di tipo **#REQUIRED** anche il **codice di catalogazione** in quanto, nell'industria cinematografica, i film vengono classificati in base al loro contenuto attraverso il visto censura (in inglese Rating), un provvedimento volto a stabilire il pubblico adatto alla loro visione. Abbiamo poi l'**anno** di distribuzione, il quale è **#IMPLIED**, ovvero, un attributo opzionale e, ancora, il **formato** al quale sono stati assegnati due valori opzionali: High Definition (HD) e Standard.

In fine, l'ultimo attributo è volto a definire se esso sia un film originale Netflix; dall'analisi del DTD possiamo dedurre che esso può esserlo o meno, in quanto non abbiamo specificato un valore di default.

- Analizziamo ora il documento XML:

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE Liberia Netflix SYSTEM "liberianetflix.dtd"
<Libreria Netflix>
  <Film cod classificazione="TI123" cod catalogazione="VM14" anno="2019" formato="HD" originale netflix="si"
    <Titolo>"The Irishman"</Titolo>
    <Regista>"Martin Scorzese"</Regista>
    <Attore>"Robert DeNiro, Al Pacino"</Attore>
  </Film>
  <Film cod classificazione="H456" cod catalogazione="T" formato="HD" originale netflix="no"
    <Titolo>"Her"</Titolo>
    <Regista>"Spike Jonze"</Regista>
    <Genere>"drammatico"</Genere>
  </Film>
</Libreria Netflix>
```

Ogni documento XML contiene due parti: *un prologo*, che contiene la dichiarazione XML ed introduce il documento e *un corpo*, cioè il documento vero e proprio.

Il prologo è:

```
<?xml version="1.0" encoding="utf-8"?>
```

Dove **<?xml...?>** corrisponde all'elemento che dichiara il documento conforme alle specifiche XML; **version="1.0"** indica la versione di riferimento, che attualmente è la 1.0; **encoding="utf-8"** corrisponde alla codifica del linguaggio del documento.

Successivamente abbiamo la Doctype Declaration che ci rimanda alla dichiarazione del DTD:

```
<!DOCTYPE Liberia Netflix SYSTEM "liberianetflix.dtd"
```

La Doctype Declaration è la grammatica, in questo caso contenuta in un file locale, da utilizzare per la validazione del documento.

Il corpo è:

```
<Libreria Netflix>
  <Film cod classificazione="TI123" cod catalogazione="VM14" anno="2019" formato="HD" originale netflix="si"
    <Titolo>"The Irishman"</Titolo>
    <Regista>"Martin Scorzese"</Regista>
    <Attore>"Robert DeNiro, Al Pacino"</Attore>
  </Film>
  <Film cod classificazione="H456" cod catalogazione="T" formato="HD" originale netflix="no"
    <Titolo>"Her"</Titolo>
    <Regista>"Spike Jonze"</Regista>
    <Genere>"drammatico"</Genere>
  </Film>
</Libreria Netflix>
```

In questo caso, i film esaminati sono due, uno originale Netflix e l'altro no. Il primo film, "*The Irishman*", originale Netflix, è definito da elementi quali Titolo, Regista e più attori; il secondo, "*Her*", non originale Netflix, è stato definito da elementi quali Titolo, Regista, un attore e un genere.

Come si può leggere dalla dichiarazione del DTD, ogni film è caratterizzato obbligatoriamente da un codice di catalogazione, uno di classificazione e da un formato. L'anno è stato inserito soltanto al primo film, in quanto era facoltativo inserirlo.

Un documento XML può essere scritto, oltre che nella sua struttura fisica, anche in una di tipo logico, chiamata "*document tree*". Dalla terminologia anglosassone, è facilmente intuibile il riferimento ad un albero.

Quello inerente all'XML preso in esame è così strutturato:

LIBRERIA NETFLIX:

FILM

Titolo, regista, attore

FILM

Titolo, regista, attore, genere