Filmoteca

Rodrigo G. López

rgl@antares-labs.eu

Antares Telecom Laboratories
Albatera, Alicante

ABSTRACT

This document is a work in progress and is subject to change by the author at any time. It shows the ideas that are being implemented in a 3TB multimedia dataset distributed among two hard drives, for the author's personal consumption.

Introduction

The filmoteca is an attempt at organizing motion picture data like movies, tv shows and documentaries in a formalised file hierarchy that is easy to use and maintain. We take influence from the Plan 9 operating system and its file servers to design an interface that makes it simple to script any operation using standard Unix tools.

1. The Tree

The file hierarchy is structured with the first-level directories named after the titles they keep. Any collision is resolved by appending a space and a roman numeral between parentheses to the title, similar to how IMDb does it [IMDb], except we favor historical correctness over time of insertion into the database. This means that if we add Rush (2012) first and later we find about Rush (1991), we will add it under the title *Rush (I)* and the latest one will become *Rush (II)*. The inner levels are thoroughly explained in Section 2.

1.1. Movies

```
title/
title/release
title/synopsis
title/cover
title/video
title/history
title/sub/
title/sub/en
title/sub/es
title/sub/...
title/dub/
title/dub/es/
title/dub/es/video
title/dub/...
title/extra/
title/extra/Opening.mp4
title/extra/Ending.mp4
title/extra/Deleted Scenes.mp4
title/extra/...
title/remake/
title/remake/1984/
title/remake/1984/video
title/remake/1984/...
title/remake/y...
```

1.2. Multipart Movies

```
title/
title/release
title/synopsis
title/cover
title/video1
title/video2
title/videon...
title/history
title/sub1/
title/sub1/en
title/sub1/...
title/subn...
title/dub1/
title/dub1/es/
title/dub1/es/video
title/dub1/...
title/dubn...
title/extra/...
title/remake/y...
```

1.3. Series

```
title/
title/release
title/synopsis
title/cover
title/history
title/s/
title/s/1/
title/s/1/1/
title/s/1/1/video
title/s/1/1/sub/
title/s/1/1/sub/en
title/s/1/1/sub/...
title/s/1/1/dub/
title/s/1/1/dub/es/
title/s/1/1/dub/es/video
title/s/1/1/dub/...
title/s/1/n...
title/s/n...
title/extra/...
title/remake/y...
```

2. Walking Down the Tree

2.1. The `release` file

The *release* file stores the date on which the work was first released. In the case of a movie (including multipart) there is only one line, whereas in a series there is one line per season. If there is some season lacking between two other seasons, an empty string is used instead. The date string has two different formats: `yyyy` and `ddmmmyyyy`, e.g. `2016`, `1932`, `23mar1997`, `1jun1978`, etc.

I am going to show you two examples, to clarify the structure of a hollowed file. First we have *Black Mirror*, with two seasons, the first and the third:

```
$ cat 'Black Mirror'/release
2011

2016
$ xd -c 'Black Mirror'/release
0000000   2  0  1  1 \n \n  2  0  1  6 \n
000000b
```

Then a extreme case, *The X Files*, where there is only one season, the tenth:

```
$ cat 'The X Files'/release
```

```
2016
$ xd -c 'The X Files'/release
0000000  \n \n \n \n \n \n \n \n \n  2  0  1  6 \n
000000e
```

You can then use the following script to figure out the folders in *s* (explained in Section 2.8):

```
$ <'Black Mirror'/release awk '
        BEGIN{ s = 0; }
        /^$/ { s++; }
        /^[0-9]+$/ { print ++s }
        '
1
3
```

2.2. The `synopsis` **file**

Contains a brief summary of the plot.

2.3. The `cover` **file***

The *cover* file is the poster of the movie or tv show encoded in well-known image formats such as JPEG, PNG, TIFF or GIF.

2.4. The `video[n]` **file**

The *video* file is the fruit of the tree, it stores the movie or a series episode under some multimedia container format, the most common being MP4, Matroska (MKV) and AVI.

2.5. The `history` **file**

In an attempt to educate the user, a *history* file is provided containing an explanation, facts and references for movies that claim to be based on/inspired by true events. You might be surprised.

_____
* This file will be renamed to *poster* in the future.

2.6. The `sub[n]` folder

Files within *sub* are SubRip (SRT) subtitle files, each named using a two-letter country code* for the language they provide.

2.7. The `dub[n]` folder

The *dub* folder stores revoicings and provides one folder per language, using the same conventions as the subtitles. Inside each of these, there is a video file (see Section 2.4)†.

2.8. The `s` folder‡

S is a three-level directory containing the episodes of a series. The first level are seasons by number, if there is some season that is lacking, we completely avoid it (signaled by the *release* file, Section 2.1), there are no empty folders. The second level are episodes following the same naming rules. The structure of an episode folder is similar to that of a movie, except it only has a *video*, and the *sub* and *dub* folders.

2.9. The `extra` folder

The *extra* folder holds a bunch of random files that make up the featurettes and other behind-the-scenes content, such as the making-of, auditions, the director's commentary, additional posters, OSTs, etc.

2.10. The `remake` folder

Remake is a two-level directory, with the first level holding directories named after the year the remake was released on, and inside each of these, the exact same file structure you would find in a movie (see Section 1.1).

---

* This will be changed to use a reasonable subset of the two/three-letter IANA language subtag registry [IANALang].

† Although under best conditions it should be just an audio file one could swap into the stream. This is an idea that is going to affect the video file as well, since one could compose their own media streams into a single experience, for example grabbing the original video, adding japanese audio and english subtitles; a very common setup for watching anime.

‡ It really is a leftover from a previous design, since it means ''season'', but there is no *e* directory containing ''episodes''.

References

[IMDb]   https://help.imdb.com/article/imdb/discover-watch/what-do-the-roman-numerals-like-i-and-
ii-after-people-s-names-mean/GA827M8GK5KVH8TC?ref_=helpart_nav_33#,   last checked March 19, 2019
[IANALang]   https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry