

Tvheadend - Feature #4592

EPG: Italian parsing season/episode from Title/Subtitle/Description

2017-09-14 16:34 - g siviero

Status:	Fixed	Start date:	2017-09-14
Priority:	Normal	Due date:	
Assignee:	Adam Sutton	% Done:	0%
Category:	EPG - Grabbers	Estimated time:	0.00 hour
Target version:			
Description			
<p>I tried to modify the file <code>/usr/share/tvheadend/data/conf/epggrab/eit/scrape/uk</code> to enable Italian parsing season/episode from Title/Subtitle/Description from EIT EPG data. For EIT EPG I'm using the grabber "Over-the-air: EIT: DVB Grabber".</p> <p>On some channels here in Italy we could have:</p> <p>1) season/episode inside the title Title: "A casa dei Loud [St.1] Ep. 4"</p> <p>2) season/episode inside the subtitle Subtitle: "St.5 Ep.12 - Joe Teti e Matt Graham hanno metodi di sopravvivenza completamente diversi. Come si comporteranno quando verranno abbandonati insieme nei luoghi piu' inospitali della Terra?"</p> <p>If I add/edit the eit/scrape file with</p> <pre>"season_num": ["\\[?St\\. ([0-9]+)\\]? "], "episode_num": [" ?[Ee]p\\.? ? ([0-9]+) "],</pre> <p>These regular expression should match both cases, but tvheadend scrapes correctly the season/episode only when this information is in the subtitle/description (case 2), while it seems to ignore it if it is inside the title (case 1).</p> <p>Any suggestion? Is "case 1" (season/episode) not yet enabled?</p>			

Associated revisions

Revision c12a80a5 - 2017-09-24 07:05 - Em Smith

eit: Also scrape eit episode information from title and description (#4592)

Previously we only searched the summary for scraping episode information, but several countries also put the information in the title or description. So we search each one in turn with the same regex and merge the results.

Issue: #4592

Revision ae18f5f6 - 2017-09-24 07:05 - Em Smith

eit: Add extra eit episode scrape configurations (#4592)

We include the regex from the opentv configuration for scraping episode information for Italy, Australia and New Zealand with minor changes to allow parsing by the Python test harness. Also added additional Italian regex from the bug report.

Issue: #4592

History

#1 - 2017-09-14 19:23 - Ludi K.

- File `epg1.png` added

- File `epg2.png` added

I get the epg by using the OpenTV grabber Sky Italia module and selecting the channel named "Giallo" for several minutes.

The season and episode information is included in the description of the event in a different format:

```
4' Stagione Ep.20 - 'Madre' Nella Foresta Incantata del passato, Regina
si reca alla tomba di Daniel in occasione dell'anniversario della morte
dell'amato. Intanto Lilith si ricongiunge con la madre.
```

Should tvheadend look for the season and episode information also in this other format?

However, I wonder why some created timers have season and episode information in the appropriate column, while others don't have that information, though the season and episode information is in the description. Does OpenTV send the season and episode information also in a field separate from the description field, which would explain why the timer in epg1.png does not have that information in the season/episode column?

In any case, I just created an RFE bug report asking for the possibility to allow the user to manually add the season and episode information to the timer. (Issue [#4593](#))

#2 - 2017-09-14 19:52 - saen acro

Ciao leggi qui

<https://tvheadend.org/issues/4509#change-23305>

You need to create new file with command

```
touch /usr/share/tvheadend/data/conf/epggrab/eit/scrape/it
```

This will create file "it"
with you will use in EPG parser

Vorrei il successo

#3 - 2017-09-14 23:31 - Em Smith

@siviero

There is a patch here to also search the title (case 1):

[\[\[https://tvheadend.org/issues/4287#note-12\]\]](https://tvheadend.org/issues/4287#note-12)

The patch is not yet submitted.

[Ludi K.](#)

The opentv grabber is different to the dvb grabber. You are probably using the configuration files in data/conf/epggrab/opentv/prov/skyit. I don't know much about it, but it looks to me like epg1 and epg2 have the same "Stagione Ep." format. Perhaps it has slightly different spacing and not matching the regex in that file?

#4 - 2017-09-15 10:16 - g siviero

- File Screenshot_2017-09-15.png added

I think I will wait for the "title" patch to be officially submitted.
Anyway at the moment season/episode from subtitle/description are working ok.

#5 - 2017-09-16 22:04 - Ludi K.

That was strange: the column with the season and episode information was blank on the timer page, but the recorded file had the season and episode information in the filename.

#6 - 2017-09-30 23:07 - Ludi K.

Some channels on Skyit have a new format for the season and episode information in the description field:

```
S5 Ep18 Le scarpette di Ruby - Un flashback ci mostra Cappuccetto Rosso e Mulan ad Oz. Qui incontrano Dorothy
con cui cercano un modo per sconfiggere la Strega una volta per tutte.
```

So I added a line for it to the season and episode sections of the share/tvheadend/data/conf/epggrab/opentv/prov/skyit file; please have a look at the third line of each section; I added that line and restarted tvheadend:

```
"season_num": [
  "([0-9]+)'?a? Stagione +Ep\\. ?[0-9]+[a-z]?",
  "([0-9]+)'?a? Stagione -? ?Puntata ?[0-9]+",
```

```
"S([0-9]+) Ep[0-9]+ ",
"([0-9]+)'?a? Stagione"
],
"episode_num": [
"[0-9]+'?a? Stagione +Ep\\. ?([0-9]+) [a-z]?",
"[0-9]+'?a? Stagione -? ?Puntata ?([0-9]+)",
"S[0-9]+ Ep([0-9]+) ",
"^ *Ep\\. ?([0-9]+) [a-z]?",
"^ *Puntata ?([0-9]+)",
" Ep\\. ?([0-9]+) -"
```

Unfortunately, it still does not seem to work. Did I miss something?

Thanks in advance for any help.

PS: I am using tvheadend 4.4.20170707 on a synology NAS; in fact, I installed the package provided by dierkse.nl

#7 - 2017-10-01 00:14 - Em Smith

I can't see anything wrong with them. However, I noticed that a patch for Italy OpenTV has just gone in.

[\[https://github.com/tvheadend/tvheadend/pull/1007/commits/07ae54b96d3f2c0ac39d7b91cb1efc115f53e444\]](https://github.com/tvheadend/tvheadend/pull/1007/commits/07ae54b96d3f2c0ac39d7b91cb1efc115f53e444)

Your change looks identical to me except yours has a space at the end before the double quotation mark. Could you double-check your regex against the one on that page? Perhaps what looks like a space isn't actually a space in the description. So, try removing the space and see if it works.

If it works, let me know if you can submit the change, otherwise I'll create a submit request in a few days.

#8 - 2017-10-01 20:40 - Ludi K.

It does not work without the trailing space either.

So, I did a copy paste of the content (not only the diff, but the whole content) of the file corresponding to the link you provided and it does not work either.

Thus, I am currently wondering whether for some reason, the changes to the skyit file get ignored or whether I need a more recent version of Tvheadend, considering that meanwhile (due to the version number of the syno package, I suppose that I am running a git version from 20170707) there have also been other commits concerning the epg.

I will add a comment here, if I make some progress.

#9 - 2017-10-01 23:00 - Em Smith

The version 20170707 sounds too old, maybe it has an incorrect date since presumably you have the configuration options in EPG Grabber where you can enter "it" for config to use and you have been getting season/episodes.

If you click on "about" tab then it will probably give you the exact version (so 4.3~abcdef) where the "abcdef" is the git commit number (exact last commit). You're right there are other commits, the last one affecting this bug is eight days ago.

In the log (I don't know where it is on your box, maybe /var/log/syslog) there should be a log saying it has loaded the config ok. Something like:

```
tbl-eit: scraper uk_freesat attempt to load config "uk_freesat"
tbl-eit: scraper uk_freesat loaded config "uk"
```

If you look for "scraper" then you should see one loading the "it" config file.

#10 - 2017-10-08 18:43 - Ludi K.

Hi,

Sorry for the delay. I did not find the revision information on the previously installed tvheadend and if the information that I read on Internet are correct, tvheadend does not write logs on the Synology, yet. (This might now change with DSM6, the current synology system.)

I finally managed to cross compile current tvheadend git (revision 5be1a5a8) for my Synology NAS. (For those interested, I used the spksrc framework, but not the original, but this fork by editing the toolchain version and the version and revision of tvheadend: <https://github.com/m4tt075/spksrc/tree/tvh-4.3>).

The bad news is that tvheadend still does not recognize the new season and episode format in the OpenTV Skyit description.

Does the changes linked in message #7 of this thread, that are also present on my installation work for anybody?

Cheers

#11 - 2017-10-09 14:15 - Ludi K.

Today, the season and episode information have appeared on the timer list, without me changing anything in the meantime.

Now I wonder whether yesterday, they were not sending the information or whether tvheadend needs time to process the data.

But anyway, it is working now, which is the most important.

#12 - 2017-10-09 15:12 - g siviero

You should also check in configuration > Channel/EPG > EPG Grabber the Cron lines for the OTA grabbers.

#13 - 2017-10-09 16:13 - Ludi K.

OTA is set to run at 1:15 PM in cron. So, tvheadend tuned again to the channel with the epg data before I checked today. But I still don't know why I did not see the season and episode information already yesterday. However, it is not really relevant, as it seems to work now.

#14 - 2017-10-20 20:51 - Jaroslav Kysela

- Status changed from New to Fixed

#15 - 2017-10-31 09:10 - g siviero

Hello,

could you add the following season regexp for the italian EIT scraping?

1) for season information like "S2 Ep7 Milano" (Italian DVB-T channels: TV8, Cielo)

```
"\\[?S ([0-9]+) \\]?"
```

and

2) for season information like "Stagione 4, Ep. 13 - Una scelta difficile." (Italian DVB-T channels: Spike, POP, Cine Sony, Paramount Channel, VH1)

```
"Stagione ([0-9]+)"
```

Thanks.

Files

epg1.png	83.8 KB	2017-09-14	Ludi K.
epg2.png	83.3 KB	2017-09-14	Ludi K.
Screenshot_2017-09-15.png	68.6 KB	2017-09-15	g siviero