

## Tvheadend - Bug #5174

### subtitle scraping wrong (double subtitle) if text contains apex character '

2018-08-13 10:43 - g siviero

<b>Status:</b>	New	<b>Start date:</b>	2018-08-13
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Adam Sutton	<b>% Done:</b>	0%
<b>Category:</b>	EPG - Grabbers	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Affected Versions:</b>	
<b>Found in version:</b>	4.3-1292		

#### Description

In some circumstances the subtitle is "doubled". This seems to happen when the subtitle is contained between apex characters (but this could be an unrelated coincidence).

issue:

```
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: svc='DMAX', ch='DMAX', eid=21815, tbl=52, running=0, start=2018-08-25;01:55:00(+0200), stop=2018-08-25;02:45:00(+0200), ebc=0x123ed28
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: eit: dtag 4D dlen 225
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 69 74 61 0E 41 20 6D 61 72 69 20 65 73 74 72 65 ita.A m
ari estre
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 6D 69 CE 53 74 2E 31 20 45 70 2E 31 20 27 49 6E mi.St.1
Ep.1 'In
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 20 6D 61 72 65 20 61 70 65 72 74 6F 27 20 2D 20 mare a
perto' -
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 4C 61 20 70 65 73 63 61 20 69 6E 20 6D 61 72 65 La pesc
a in mare
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 20 61 70 65 72 74 6F 20 65 27 20 75 6E 6F 20 64 aperto
e' uno d
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 65 69 20 6C 61 76 6F 72 69 20 70 69 75 27 20 64 ei lavo
ri piu' d
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 69 66 66 69 63 69 6C 69 20 69 6E 20 47 72 61 6E ifficil
i in Gran
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 20 42 72 65 74 61 67 6E 61 2E 20 4C 61 20 76 69 Bretag
na. La vi
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 74 61 20 61 20 62 6F 72 64 6F 20 6E 6F 6E 20 65 ta a bo
rdo non e
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 27 20 63 6F 6D 70 6C 69 63 61 74 61 20 73 6F 6C ' compl
icata sol
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 6F 20 70 65 72 20 6C 65 20 6D 61 74 72 69 63 6F o per l
e matrico
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 6C 65 3A 20 74 75 74 74 61 20 6C 61 20 70 72 65 le: tut
ta la pre
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 73 73 69 6F 6E 65 20 72 69 63 61 64 65 20 73 75 ssione
ricade su
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 6C 6C 6F 20 73 6B 69 70 70 65 72 20 50 68 69 6C llo ski
pper Phil
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: 2E
```

#### Season extraction (ok with modified code)

```
Aug 13 10:23:42 tvheadend[1453]: epgrab: pattern "[?St\\.([0-9]+)\]?" matches '1' from 'St.1 E
p.1 'In mare aperto' - La pesca in mare aperto e' uno dei lavori piu' difficili in Gran Bretagna.
La vita a bordo non e' complicata solo per le matricole: tutta la pressione ricade sullo skipper P
hil.'
```

```
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: extract season number 1 using eit
```

#### Episode extraction (ok with modified code)

```
Aug 13 10:23:42 tvheadend[1453]: epgrab: pattern "[?Ee]p\.\.? ?([0-9]+)" matches '1' from 'St.1
```

Ep.1 'In mare aperto' - La pesca in mare aperto e' uno dei lavori piu' difficili in Gran Bretagna . La vita a bordo non e' complicata solo per le matricole: tutta la pressione ricade sullo skipper Phil.'

```
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: extract episode number 1 using eit
```

#### Subtitle extraction

```
Aug 13 10:23:42 tvheadend[1453]: epgrab: pattern "Ep[.] ?[0-9]+[A-Za-z]? -? ?'([^\']*(' [^A-Z0-9-])?('[^ '])?)+'" matches 'In mare apertoIn mare aperto' from 'St.1 Ep.1 'In mare aperto' - La pesca in mare aperto e' uno dei lavori piu' difficili in Gran Bretagna. La vita a bordo non e' complicata solo per le matricole: tutta la pressione ricade sullo skipper Phil.'
```

```
Aug 13 10:23:42 tvheadend[1453]: tbl-eit: scrape subtitle 'In mare apertoIn mare aperto' from 'St.1 Ep.1 'In mare aperto' - La pesca in mare aperto e' uno dei lavori piu' difficili in Gran Bretagna. La vita a bordo non e' complicata solo per le matricole: tutta la pressione ricade sullo skipper Phil.' using eit
```

the issue seems to come from lines 141-146 in src/epgrab/module/eitpatternlist.c

```
for (matchno = 2; ; ++matchno) {
    if (regex_match_substring(&p->compiled, matchno, matchbuf, sizeof(matchbuf)))
        break;
    size_t len = strlen(buf);
    strlcat(buf, matchbuf, size_buf - len);
}
```