CC CCExtractor GUI 0.85 — □ ×

| About & Save | Input files | Input options | Advanced input options | Decoders | Output (1) | Output (2) | Credits | Debug | HDHomeRun | Burned-ir ◄ ► |

○ Process input via UTP, listen on port `1235`

● Process files from box below (drag them into box)   Clear list

```
C:\Users\wulff\Desktop\▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓3_2245.ts
```

CCExtractor binary version

● Without OCR (no external dependencies, should work everywhere)   ○ With OCR (use if you need to convert bitmaps to text, such as DVB)

Command line

C:\Program Files (x86)\CCExtractor\ccextractorwin.exe --gui_mode_reports -autoprogram -out=srt -bom -latin1 [+input files]

CCExtractor GUI 0.85

About & Save | Input files | **Input options** | Advanced input options | Decoders | Output (1) | Output (2) | Credits | Debug | HDHomeRun | Burned-in ◀ ▶

Input type

- ◉ Autodetect the correct format
- ○ Transport Stream
- ○ Hexadecimal Dump
- ○ Program Streams
- ○ Elementary Stream
- ○ Microsoft WTV
- ○ DVR-MS
- ○ .raw (McPoodle's)
- ○ M2TS
- ○ .bin (CCExtractor's)
- ○ MP4

Split type

- ○ These are individual, unrelated files. Produce an output subtitle file for each input file.
- ○ These files are part of the same video. They were cut with a video tool.
- ◉ These files are part of the same video. They were cut by a generic tool.

Some input types (MP4, hexadecimal dump and WTV) don't support cut files and each file will be processed as a separate video file.

Mode

- ◉ This is a complete recording
- ○ This is live. Continue reading file until no new data arrives for [0] seconds.
(0 = no limit)

Start of processing
- ◉ Process the input from the start
- ○ Start at [0] [0] [0]

End of processing
- ◉ Process the input up to the end
- ○ End at [0] [0] [0]

"Screenfuls" limit
- ◉ No limit
- ○ Stop after [0] screenfuls

Elementary Stream

- ◉ Select the stream most likely to contain data
- ○ Process the stream with this PID [   ]
- ○ Process the first stream with this type [   ]
- Process the stream as if its MPEG type is [   ]

Command line

C:\Program Files (x86)\CCExtractor\ccextractorwin.exe --gui_mode_reports -autoprogram -out=srt -bom -latin1 [+input files]

CCExtractor GUI 0.85

About & Save | Input files | Input options | **Advanced input options** | Decoders | Output (1) | Output (2) | Credits | Debug | HDHomeRun | Burned-in ◄ ►

Note: While you can't break anything (other than the output subtitles) by playing with these options, they are usually not needed.

MythTV

○ This recording have the closed captions in the VBI data, so use MythTV's code to read it

○ Do not use MyTV's code even if there is VBI data

◉ Autodetect this and use MyTV's code only when needed

Miscellaneous options

☐ Use a clock frequency of 90090 instead of 90000 (needed for some DVDs)

☐ Fix incorrect padding data (by replacing 00 with 80 where appropiate)

☐ This recording has very large GOPs, don't discard GOP timing anyway

☐ Input file was originally a WTV file that has been converted to DVR-MS and CCExtractor finds no captions.

☐ File was captured with a Hauppage card (try to extract captions in Hauppage's propietary encoding)

☐ Use the pic_order_cnt_lsb in AVC/H.264 data streams to order the CC information

☐ In MP4, process video track even if a dedicated closed captions track is present

☐ Ignore SCTE-20 data

Clock

◉ Auto   ○ GOP   ○ PTS

There are two possible sources of timing information for CCExtractor. One comes is the PTS (Presentation Time Stamp) and the other is the information for each GOP (Group of Pictures). In general CCExtractor knows which clock to use, but if you see an output file in which the clock seems stopped, use these options to switch.

Multiprogram Transport Streams

○ File is known to contain more than one program. Process this one: [   ]

◉ If the TS contains more than one program, select the first one found with suitable data

Command line

C:\Program Files (x86)\CCExtractor\ccextractorwin.exe --gui_mode_reports -autoprogram -out=srt -bom -latin1 [+input files]

About & Save | Input files | Input options | Advanced input options | **Decoders** | Output (1) | Output (2) | Credits | Debug | HDHomeRun | Burned-ir ◄ ►

608 Decoder

Field
- ☑ Extract captions from field 1
- ☐ Extract captions from field 2

Channel
- ◉ Extract captions from channel 1
- ○ Extract captions from channel 2

In general, if you want English subtitles you don't need to use these options as they are broadcast in field 1, channel 1. If you want the second language (usually Spanish) you may need to try other combinations.

708 Decoder

The 708 decoder is still a work in progress. Output is a lot worse than the one generated by its 608 counterpart, and it's not production ready. However for streams where only 708 captions are available it will at least generate a transcript to work with.

☐ Enable 708 decoder   Process these services: `1,2`

The service list is a comma separated list of services to process. Service 1 is always the primary language, and service 2 is a translation of the text from service 1 into another language. Rest of services are to be defined by the content supplier and can be anything. Valid values are 1 to 63, and ranges are NOT supported.

Teletext decoder

- ◉ Autodetect teletext
- ○ Disable teletext processing
- ○ Force telext processing even if not detected and extract data from this PID [ ]
- ○ Prefer DVB over teletext subtitles if both are present

Page for subtitles (blank for autodetect) [ ]

Teletext line deduplication

Some stations simulate real time captioning by adding new text to a page as it's spoken. In transcript mode, this causes duplicated lines. CCExtractor tries to remove these duplicates by removing consecutive lines under a certain levenshtein distance threshold. Check documentation for more information.

Minimum allowed distance (default 2) [ ]   Maximum allowed distance, as a % of length (default 10) [ ]

Command line

C:\Program Files (x86)\CCExtractor\ccextractorwin.exe --gui_mode_reports -autoprogram -out=srt -bom -latin1 [+input files]

CCExtractor GUI 0.85

About & Save | Input files | Input options | Advanced input options | Decoders | Output (1) | Output (2) | Credits | Debug | HDHomeRun | Burned-in ◄ ►

Output type

- ⦿ .srt (SubRip)
- ◯ .sami (Microsoft Synchronized Accessible Media Interchange)
- ◯ .txt (Transcript, no timing)
- ◯ .ttxt (Timed Transcript)
- ◯ .ttml (SMPTE Timed Text - W3C TTML)
- ◯ Null (don't generate any file)
- ◯ .g608 (608 grid)

- ◯ .raw (CC data in McPoodle's broadcast format)
- ◯ .dvdraw (CC data in McPoodle's DVD format)
- ◯ .bin (CC data in CCExtractor's format)
- ◯ spupng
- ◯ SimpleXML
- ◯ File analysis report
- ◯ .ass (SubStation Alpha)

XDS

☐ Export XDS information if the output format allows it (currently only transcripts)

Output file

- ⦿ Use the same name as the input file, replacing the extension with the correct one
- ◯ Save the ouput in this file [   ] [Browse]

  Note: When producing more than one file (i.e. when extracting the data from more than one service or language), a numeric suffix will be appended to the file.

Byte Order Mark (BOM)

⦿ Add BOM in UTF files (often required by Windows programs)   ◯ Don't add BOM in UTF files (often required by UNIX programs)

Time adjust

[0 ⬍] Delay subtitles this number of milliseconds (use negative values to make them appear sooner)

Command line

C:\Program Files (x86)\CCExtractor\ccextractorwin.exe --gui_mode_reports -autoprogram -out=srt -bom -latin1 [+input files]

CCExtractor GUI 0.85

About & Save | Input files | Input options | Advanced input options | Decoders | Output (1) | **Output (2)** | Credits | Debug | HDHomeRun | Burned-in ◄ ►

Encoding
- ◉ Latin-1
- ○ UTF-8
- ○ Unicode

Colors
- ○ Don't add any color information to the text
- ◉ Add color information when possible, but leave the default color (usually white) alone
- ○ Add color information when possible, and replace the default color with this # [   ]

Text position
- ◉ Export text in its original position by adding spaces where needed (this makes all lines 32 characters long)
- ○ Center text by removing left and right spaces
  - ☐ Add a dash (-) when the speaker changes (only for .srt and with -trim)

Sentence case
- ◉ Do not modify case. Leave letters as they are (possibly all caps)
- ○ Apply standard capitalization rules and use the built-in (small) list of words that must be capitalized (such as Spain)
  - ☐ Also use the list of words in this file [   ] [Browse]

Roll-Up
- ☐ For roll-up captions, make letters appear in real time instead of one line at a time, at the expense of duplicating the screen for each new letter (produces very large files)
- ☐ For roll-up captions, don't duplicate lines (at the expense of having only one visible line at any given time, since the only way to have several lines and emulate the roll-up effect is by duplicating lines)
- Bypass number of visible lines and limit to ◉ No limit (as defined by broadcast) ○ 3 lines ○ 2 lines ○ 1 line

Typesetting
- ◉ Export typesetting tags (for italics and bold)
- ○ Do not export typesettings tags

Line terminators
- ◉ DOS/Windows (CRLF)
- ○ UNIX (LF)

Command line
```
C:\Program Files (x86)\CCExtractor\ccextractorwin.exe  --gui_mode_reports  -autoprogram  -out=srt -bom -latin1 [+input files]
```