WebVTT

This specification defines WebVTT, the **Web Video Text Tracks** format. Its main use is for marking up external text track resources in connection with the HTML <track> element. WebVTT files provide captions or subtitles for video content, and also text video descriptions, chapters for content navigation, and more generally any form of metadata that is time-aligned with audio or video content.

Table of Contents

WebVTT is a simple caption file basically

The main use for WebVTT files is captioning or subtitling video content. Here is a sample file that captions an interview:

```
WEBVTT

00:11.000 --> 00:13.000
<v Roger Bingham>We are in New York City

00:13.000 --> 00:16.000
<v Roger Bingham>We're actually at the Lucern Hotel, just down the street

00:16.000 --> 00:18.000
<v Roger Bingham>from the American Museum of Natural History

00:18.000 --> 00:20.000
<v Roger Bingham>And with me is Neil deGrasse Tyson

00:20.000 --> 00:22.000
<v Roger Bingham>Astrophysicist, Director of the Hayden Planetarium

00:22.000 --> 00:24.000
<v Roger Bingham>at the AMNH.

00:24.000 --> 00:26.000
<v Roger Bingham>Thank you for walking down here.

00:27.000 --> 00:30.000
<v Roger Bingham>And I want to do a follow-up on the last conversation we did.

00:30.000 --> 00:31.500 align:right size:50%
<v Roger Bingham>When we e-mailed—

00:30.500 --> 00:32.500 align:left size:50%
<v Neil deGrasse Tyson>Didn't we talk about enough in that conversation?

00:32.000 --> 00:35.500 align:right size:50%
<v Roger Bingham>No! No no no no; 'cos 'cos obviously 'cos

00:32.500 --> 00:33.500 align:left size:50%
<v Neil deGrasse Tyson><i>Laughs</i>

00:35.500 --> 00:38.000
<v Roger Bingham>You know I'm so excited my glasses are falling off here.
```

Caption cues with multiple lines

These captions on a public service announcement video demonstrate line breaking:

```
WEBVTT

00:01.000 --> 00:04.000
Never drink liquid nitrogen.

00:05.000 --> 00:09.000
— It will perforate your stomach.
— You could die.

00:10.000 --> 00:14.000
The Organisation for Sample Public Service Announcements accepts no liability for the content of this
advertisement, or for the consequences of any actions taken on the basis of the information provided.
```

The first cue is simple, it will probably just display on one line. The second will take two lines, one for
each speaker. The third will wrap to fit the width of the video, possibly taking multiple lines. For example,
the three cues could look like this:

```
          Never drink liquid nitrogen.

      — It will perforate your stomach.
            — You could die.

    The Organisation for Sample Public Service
    Announcements accepts no liability for the
    content of this advertisement, or for the
     consequences of any actions taken on the
        basis of the information provided.
```

If the width of the cues is smaller, the first two cues could wrap as well, as in the following example. Note
how the second cue's explicit line break is still honored, however:

```
      Never drink
    liquid nitrogen.

  — It will perforate
      your stomach.
    — You could die.

  The Organisation for
  Sample Public Service
  Announcements accepts
  no liability for the
     content of this
  advertisement, or for
   the consequences of
  any actions taken on
    the basis of the
  information provided.
```

Also notice how the wrapping is done so as to keep the line lengths balanced.

Styling captions

CSS style sheets that apply to an HTML page that contains a video element can target WebVTT cues and regions in the video using the ::cue, ::cue(), ::cue-region and ::cue-region() pseudo-elements.

```
WEBVTT

STYLE
::cue {
  background-image: linear-gradient(to bottom, dimgray, lightgray);
  color: papayawhip;
}
/* Style blocks cannot use blank lines nor "dash dash greater than" */

NOTE comment blocks can be used between style blocks.

STYLE
::cue(b) {
  color: peachpuff;
}

hello
00:00:00.000 --> 00:00:10.000
Hello <b>world</b>.

NOTE style blocks cannot appear after the first cue.
```

Comments in WebVTT

Comments are just blocks that are preceded by a blank line, start with the word "NOTE" (followed by a space or newline), and end at the first blank line.

```
WEBVTT

NOTE
This file was written by Jill. I hope
you enjoy reading it. Some things to
bear in mind:
- I was lip-reading, so the cues may
not be 100% accurate
- I didn't pay too close attention to
when the cues should start or end.

00:01.000 --> 00:04.000
Never drink liquid nitrogen.

NOTE check next cue

00:05.000 --> 00:09.000
— It will perforate your stomach.
— You could die.

NOTE end of file
```

List of program can open .vtt files

Product Name	Company	Actions
Atlantis Word Processor	The Atlantis Word Processor Team	open
GOM Player Plus	GOM & Company	Add to GOM Player Plus, open
PotPlayer	Kakao	Add to PotPlayer playlist, open, Play with PotPlayer
VisionTools Pro-e	Crestron Electronics, Inc	open

Metadata Tracks

Metadata Tracks are used to convey any additional information (such as base64 encoded images, JSON, additional text or any additional text-based file format) the developer needs to include in the page based on time indexes. A web app can listen for cue events, extract the text of each cue as it fires, parse the data and then use the results to make DOM changes (or perform other JavaScript or CSS tasks) synchronised with media playback.

sample_metadata_tracks.vtt

```
WEBVTT - Example metadata track containing JSON payload

multiCell
00:01:15.200 --> 00:02:18.800
{
"title": "Multi-celled organisms",
"description": "Multi-celled organisms have different types of cells that perform specialised functions.
  Most life that can be seen with the naked eye is multi-cellular. These organisms are though to have evolved
around 1 billion years ago with plants, animals and fungi having independent evolutionary paths.",
"src": "multiCell.jpg",
"href": "http://en.wikipedia.org/wiki/Multicellular"
}

insects
00:02:18.800 --> 00:03:01.600
{
"title": "Insects",
"description": "Insects are the most diverse group of animals on the planet with estimates for the total
  number of current species range from two million to 50 million. The first insects appeared around
  400 million years ago, identifiable by a hard exoskeleton, three-part body, six legs, compound eyes
  and antennae.",
"src": "insects.jpg",
"href": "http://en.wikipedia.org/wiki/Insects"
}
```

sample_metadata_tracks2.vtt

```
WEBVTT

NOTE
Thanks to http://output.jsbin.com/mugibo

1
00:00:00.100 --> 00:00:07.342
{
 "type": "WikipediaPage",
 "url": "https://en.wikipedia.org/wiki/Samurai_Pizza_Cats"
}

2
00:07.810 --> 00:09.221
{
 "type": "WikipediaPage",
 "url" :"http://samuraipizzacats.wikia.com/wiki/Samurai_Pizza_Cats_Wiki"
}

3
00:11.441 --> 00:14.441
{
 "type": "LongLat",
 "lat" : "36.198269",
 "long": "137.2315355"
}
```

Good References

- Technical Specs: https://www.w3.org/TR/webvtt1/
- Metadata format can contain image, description, and its hyper link (href): https://www.w3.org/wiki/VTT_Concepts
- WebVTT Example in HTML 5 implemented by Ian Devlin: https://www.iandevlin.com/html5test/webvtt/html5-video-webvtt-sample.html
- Plugins supported: plyr.io, playr, Flowplayer, jwplayer, MediaElement.js, LeanBack Player, SublimeVideo, Video.js, Radiant Media Player. You can also have good information at https://videosws.praegnanz.de/ that shows HTML5 Video Player Comparison.