

JTubeSpeech: Corpus of Japanese speech collected from YouTube

This repository provides a list of YouTube videos with Japanese subtitles.

Description

`{YYYYMM}.csv` lists as follows.

	videoid	auto	sub	channelid
0	0017RsBbUHK	True	True	UCTW2tw0Mhho72MojB1L48IQ
1	00PqfZgiboc	False	True	UCzoghTgl4dvlW9GZF6UC-BA
---	---	---	---	---

- `YYYYMM`: Year and month when we collect data
- `videoid`: YouTube video ID. Its YouTube page is <https://www.youtube.com/watch?v={videoid}>.
- `auto`: The video has an automatic subtitle or not.
- `sub`: The video has a manual (i.e., human-generated) subtitle or not.
- `channelid`: YouTube Channel ID. Its YouTube page is <https://www.youtube.com/channel/{channelid}>.

Statistics

	202103.csv	{YYYYMM}.csv
#videos-sub-true	xxx	(TBA)
#videos-auto-true	xxx	(TBA)

Contributors

- [Shinnosuke Takamichi](#) (The University of Tokyo, Japan) [main contributor]
- [Ludwig Kürzinger](#) (Technical University of Munich, Germany)
- [Takaaki Saeki](#) (The University of Tokyo, Japan)
- [Sayaka Shiota](#) (Tokyo Metropolitan University, Japan)
- [Shinji Watanabe](#) (Carnegie Mellon University, USA)

Scripts

- coming soon

Reference

- coming soon

Link

- [youtube-dl](#)