Package 'transcribeR'

August 3, 2015

Version 0.0.0

Date 2015-08-02

Type Package

Title Automated Transcription of Audio Files Through the HP IDOL API

Author Christopher Lucas, Dean Knox, Dustin Tingley, Thomas Scanlan, Shiv Sunil, Michael May, Angela Su

Maintainer Christopher Lucas <clucas@fas.harvard.edu>

Description Transcribes audio to text with the HP IDOL API. Includes functions to upload files, retrieve transcriptions, and monitor jobs.

VignetteBuilder knitr

Imports httr

Suggests knitr

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2015-08-03 18:53:03

R topics documented:

printLanguages *Convenient user-facing function to print language codes for HP API.*

Description

Prints the language codes used by HP IDOL OnDemand's Speech Recognition API. The default in `sendAudioGetJobs` is "en-US".

Usage

```
printLanguages()
```

Examples

```
## Prints the language codes used by HP IDOL OnDemand's Speech Recognition API.
printLanguages()
```

retrieveText *Gets Text From Returned JSON Object*

Description

Gets asynchronous call results from jobs previously posted to HP IDOL OnDemand's Speech Recognition API. Calls `getRequestResults` to extract the transcription from the JSON object. Returns a dataframe which was used to update a `transcribeR` CSV file.

Usage

```
retrieveText(job.file, api.key)
```

Arguments

`job.file`	the CSV file that `retrieveText` checks to determine the indices that need transcriptions and then updates
`api.key`	the API Key used to authenticate requests to HP IDOL OnDemand; one can be obtained from making an account on HP IDOL OnDemand's website https://www.idolondemand.com/

Examples

```
## Not run:
## Adds transcripts to a transcribeR CSV


retrieveText(CSV_LOCATION, API_KEY)

## End(Not run)
```

sendAudioGetJobs *Posts to HP IDOL OnDemand Speech Recognition API and Store the Resulting Job ID*

Description

Main user function to POST to HP IDOL OnDemand Speech Recognition API and write Job ID to a transcribeR CSV, a filename. It posts all compatible files in the provided directory. It avoids posting any file that has already been posted according to the transcribeR CSV.

Usage

```
sendAudioGetJobs(wav.dir, api.key, interval = "-1",
                 encode = "multipart", existing.csv = NULL, csv.location,
                 language = "en-US", verbose = FALSE)
```

Arguments

wav.dir	the directory in which the WAV files of interest are located
api.key	the API Key used to authenticate requests to HP IDOL OnDemand; one can be obtained from making an account on HP IDOL OnDemand's website https://www.idolondemand.com/
interval	optional argument; specifies segmentation length for transcripts in milliseconds; -1 indicates no segmentation and 0 indicates segmentation by words; default is "-1"
encode	optional argument; passed directly to POST in the httr package; defaults to "multipart"
existing.csv	optional argument; specifies CSV to post; defaults to NULL
csv.location	output tag for CSV file
language	optional argument; provides language code for HP IDOL OnDemand's Speech Recognition API; defaults to "en-US"
verbose	optional argument; if TRUE, prints out messages indicating the percentage of audio files posted to HP to the console

Details

When two files share the same basename but are different file formats, only one of the files will be posted to the Speech Recognition API because after the first file is posted, the basename is logged into the transcribeR CSV. Hence, all subsequent files with the same basename will not be posted. The first file that is posted is system-dependent because the ordering of filepaths is system-dependent. Functionally, however, this still avoids redundancy in posting files to the API.

Examples

```
## Not run:
## Reads files from the specified directory and creates a new CSV at the specified path

fnames <- c('boxer.wav', 'merkley.wav') # Example with 2 files
urls <- paste("http://christopherlucas.org/transcribeR/", fnames, sep = '')
lapply(urls, function(x) download.file(x, destfile = basename(x), mode = 'wb'))

WAV_DIR <- getwd()

sendAudioGetJobs(wav.dir = WAV_DIR,
                 API_KEY",
                 csv.location = CSV_LOCATION)

## End(Not run)
```

Index