

HQSYN16 - Task #4131

Task # 3672 (Assigned): RA1d - Automatic cleaning of speech corpora

Task # 3690 (Closed): Annotation error detection

Task # 3899 (Closed): Submit a paper on anomaly-based annotation errors detection (Jimp)

Task # 4128 (Closed): Final listening test based evaluation of annotation error detection

Select utterances containing units with detected annotation errors

03.01.2017 15:36 - Matoušek Jindřich

Status:	Closed	Start date:	03.01.2017
Priority:	Normal	Due date:	11.01.2017
Assignee:	Matoušek Jindřich	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	RA1: Analysis of artifacts in synthetic speech		
Description			
<ol style="list-style-type: none">1. Analyze logs of synthesized utterances and grep utterances that contain unit(s) with detected annotation error2. Sort the filtered utterances according to the number of units with detected error3. Select the following utterances:<ol style="list-style-type: none">1. 20 utterances with the most units that contain an error2. 20 utterances with just one unit containing an error3. 20 utterances with something in between (depending on the result of logging)4. Store texts and waveforms of the selected utterances			
Related issues:			
Blocked by HQSYN16 - Task #4130: Prepare words with detected annotation errors		Closed	03.01.2017 06.01.2017
Blocked by HQSYN16 - Task #4129: Synthesize & log a large portion of text by ...		Closed	03.01.2017 06.01.2017
Blocks HQSYN16 - Task #4132: Synthesize the selected utterances by TTS system...		Closed	03.01.2017 13.01.2017

History

#1 - 03.01.2017 15:36 - Matoušek Jindřich

- Blocked by Task #4130: Prepare words with detected annotation errors added

#2 - 03.01.2017 15:37 - Matoušek Jindřich

- Blocked by Task #4129: Synthesize & log a large portion of text by TTS system with annotation errors added

#3 - 03.01.2017 15:42 - Matoušek Jindřich

- Blocks Task #4132: Synthesize the selected utterances by TTS system with/without the annotation errors added

#4 - 05.01.2017 14:47 - Matoušek Jindřich

Words detected as containing annotation errors (and being really misannotated) are attached [here \(#4130\)](#).

#5 - 09.01.2017 13:43 - Tihelka Dan

- File `annot-errors.detected.with_stats.txt` added

- Status changed from New to Resolved

- Assignee changed from Tihelka Dan to Matoušek Jindřich

In [annot-errors.detected.with_stats.txt](#), the words detected as containing annotation errors (in the file ...) were extended with the number representing how many units from the words were used during the synthesis of the large tests (see [#4129](#)). The list of words was sorted according to the number of selected units.

#6 - 09.01.2017 14:29 - Matoušek Jindřich

- Status changed from Resolved to Feedback

- Assignee changed from Matoušek Jindřich to Tihelka Dan

The absolute numbers of units are fine but it might be better to specify also the average number of units (containing annotation errors) per synthetic phrase.

#7 - 10.01.2017 14:10 - Tihelka Dan

- File [annot-errors.detected.with_rels.txt](#) added

Attachment [annot-errors.detected.with_rels.txt](#) contains the similar statistics as [annot-errors.detected.with_stats.txt](#), with the difference that the first number represents $W/(P+1)$, where:

- W is the number of selections from the given word and
- P is the number of phrases it was used in.

The second column in [annot-errors.detected.with_rels.txt](#) corresponds to the first column in [annot-errors.detected.with_stats.txt](#).

#8 - 10.01.2017 14:11 - Tihelka Dan

- Status changed from *Feedback* to *Resolved*

- Assignee changed from *Tihelka Dan* to *Matoušek Jindřich*

#9 - 10.01.2017 15:26 - Matoušek Jindřich

- File [least_frequent.txt](#) added

- File [mean_frequent.txt](#) added

- File [most_frequent.txt](#) added

- Status changed from *Resolved* to *Closed*

The following items selected:

- 20 "most frequent" items (a combination of the most frequently selected units in absolute and relative numbers) -- [most_frequent.txt](#)
- 20 "least frequent" items (a combination of the least frequently selected units in absolute and relative numbers) -- [least_frequent.txt](#)
- 20 "mean frequent" items (a combination of the moderately frequently selected units in absolute and relative numbers) -- [mean_frequent.txt](#)

Files

annot-errors.detected.with_stats.txt	13.9 KB	09.01.2017	Tihelka Dan
annot-errors.detected.with_rels.txt	17.3 KB	10.01.2017	Tihelka Dan
least_frequent.txt	1000 Bytes	10.01.2017	Matoušek Jindřich
mean_frequent.txt	954 Bytes	10.01.2017	Matoušek Jindřich
most_frequent.txt	1015 Bytes	10.01.2017	Matoušek Jindřich