

# ALIGNMEET: A Comprehensive Tool for Meeting Annotation, Alignment, and Evaluation

Peter Polák, Muskaan Singh, Anna Nedoluzhko, Ondřej Bojar  
 <surname>@ufal.mff.cuni.cz

Charles University, Faculty of Mathematics and Physics  
 Institute of Formal and Applied Linguistics  
 Prague, Czech Republic



## Introduction

Summarization is a challenging problem and more so for spontaneous discussions. ALIGNMEET is a comprehensive tool for meeting annotation, alignment, and evaluation. It provides fast annotation while mitigating the risk of introducing errors. The evaluation mode enables a comprehensive quality evaluation of meeting minutes.

## What is an ALIGNMENT?

An  $n$ -to-1 mapping capturing which dialogue acts are associated with a particular summary point.

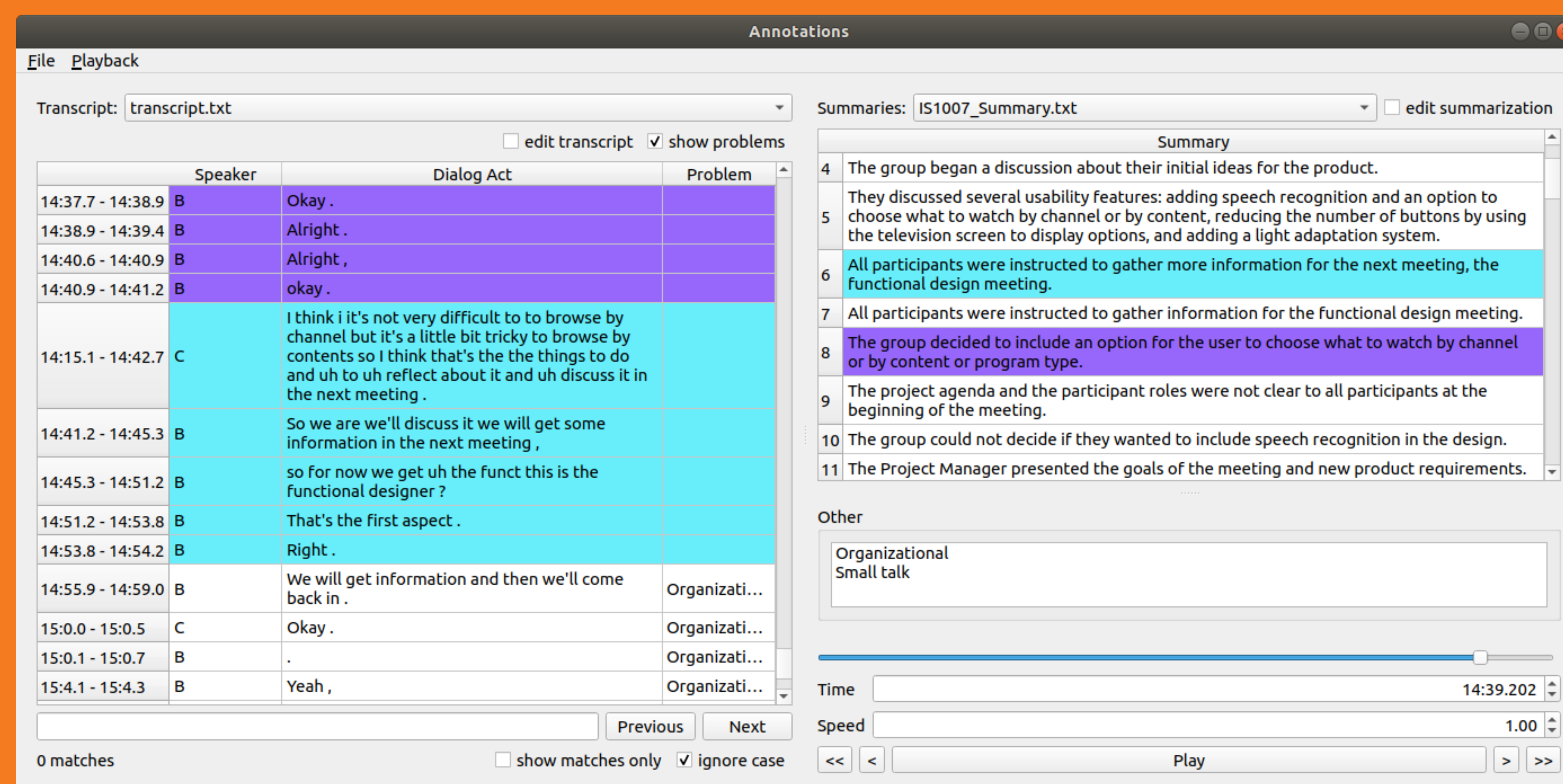
## Why do we need the ALIGNMENT?

- Visual aid for annotators.
- Annotation error check (no left-out transcript).
- Data for the training of automatic minuting.
- More reliable minutes evaluation (via aggregating meeting-level scores across hunks of the alignment).

## Evaluation

- Annotator evaluates on a summary-point-level rather than on document-level.
- More fine-grained and less mentally challenging.
- *Adequacy, grammaticallity and fluency.*
- We aggregate the metrics using plain average to get document-level scores, and we compute coverage of the transcript by aligned summary points.
- We additionally ask annotators to provide a doc-level adequacy.

# User-friendly tool for creating and curating meeting transcripts and summaries, aligning between them, and evaluating the summaries.

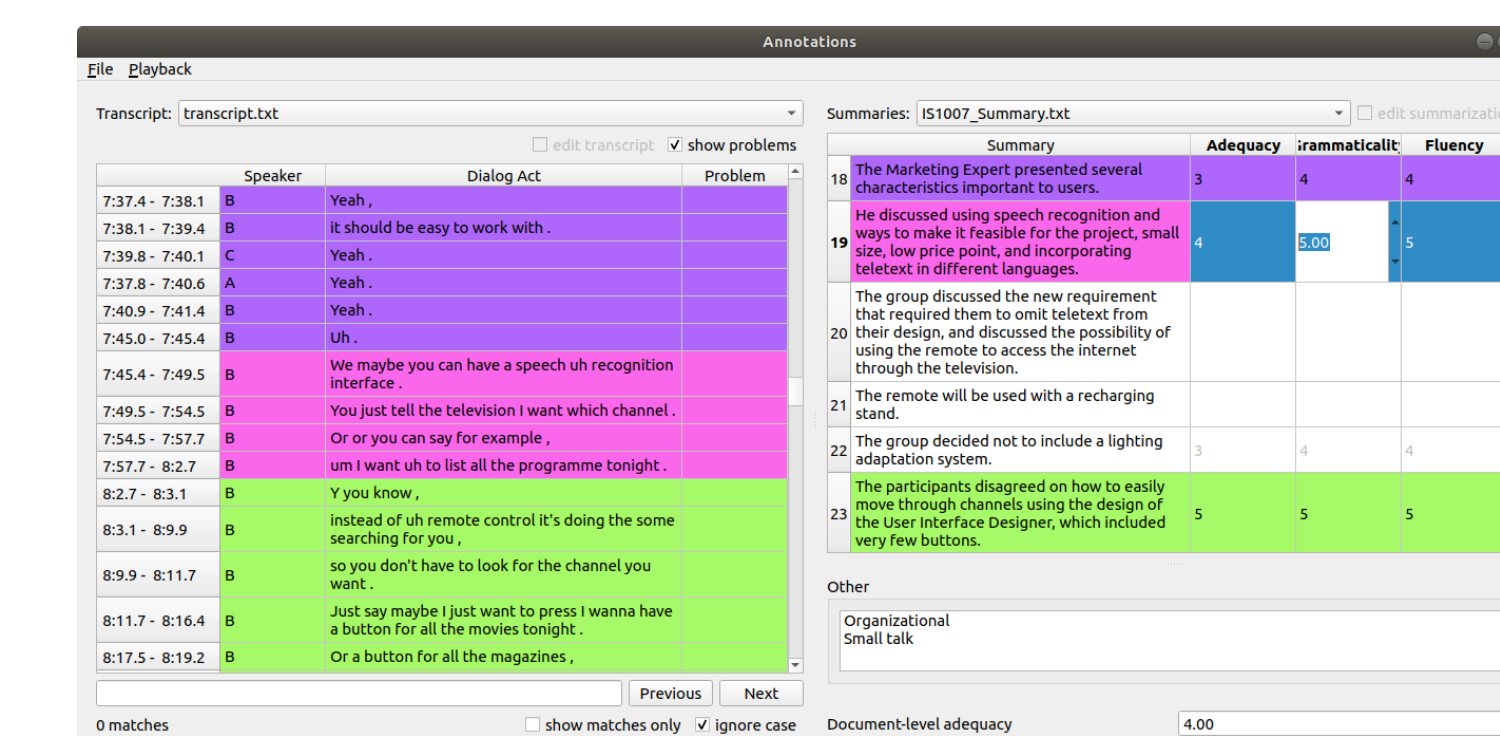


Our Github:



<https://github.com/ELITR/alignmeet>

## Evaluation mode



## Why $n$ -to-1 alignment?

- Aligning multiple summary points to a single transcript line would further increase the difficulty.
- “Summary point fragmentation” – the annotator might address the same information in separate summary points.

## Can we align everything?

- Some parts of the transcript cannot be aligned to a meeting summary point.
- They can be manually flagged as belonging to *organizational* or *small talk*.

## Pilot study

	English			Czech		
Annotator	E1	E2	E3	C1	C2	C3
Experienced	X	✓	✓	X	✓	✓
Summarization	37	45	40	23	45	31
Alignment	5	23	20	18	30	11
Evaluation	10	15	10	25	15	13
Total time	52	83	70	66	90	55

	English			Czech		
Annotator	E1	E2	E3	C1	C2	C3
Experienced	X	✓	✓	X	✓	✓
#Summary points	15	11	19	23	14	21
#Alignments	378	378	203	282	282	282
IAA	0.21*			0.63		
Avg. adequacy	3.71	3.71	3.17	3.67	4.93	4.67
Avg. grammaticality	3.86	4.21	4.08	5.00	4.13	4.67
Avg. fluency	4.71	4.07	4.92	5.00	4.53	4.53
Doc.-level adequacy	3.00	4.00	4.00	4.00	4.00	4.00
Avg. adequacy	3.53±0.31			4.42±0.67		
Avg. grammaticality	4.05±0.18			4.60±0.44		
Avg. fluency	4.57±0.44			4.69±0.27		
Doc.-level adequacy	3.67±0.58			4.00±0.00		
Coverage	1.00	0.94	0.54	0.64	0.54	0.30