

Podcasts Track results — (UCF\_NLP) University of Central Florida

Summary Statistics	
Run ID	UCF_NLP2
Task	Summarization
Run type	automatic
Summarization model	Supervised neural or deep learning model; Transfer learning or other pre-trained model
Summarization method	Abstractive
Number of summaries	179

Overall measures	
Summaries rated "excellent"	45 (25.1%)
Summaries rated "good"	47 (26.3%)
Summaries rated "fair"	50 (27.9%)
Summaries rated "bad"	37 (20.7%)

ROUGE score compared to episode description	
ROUGE_L P score	0.2553 (95%-conf.int. 0.24301 - 0.26686)
ROUGE_L R score	0.2242 (95%-conf.int. 0.21410 - 0.23504)
ROUGE_L F score	0.2091 (95%-conf.int. 0.19984 - 0.21846)

Qualitative judgments	
Q1: Does the summary include names of the main people (hosts, guests, characters) involved or mentioned in the podcast?	64.2%
Q2: Does the summary give any additional information about the people mentioned (such as their job titles, biographies, personal background, etc)?	38.5%
Q3: Does the summary include the main topic(s) of the podcast?	76.5%
Q4: Does the summary tell you anything about the format of the podcast; e.g. whether it's an interview, whether it's a chat between friends, a monologue, etc?	56.4%
Q5: Does the summary give you more context on the title of the podcast?	72.6%
Q6: Does the summary contain redundant information?	6.1%
Q7: Is the summary written in good English?	87.7%
Q8: Are the start and end of the summary good sentence and paragraph start and end points?	62.0%

