

CONTINUOUS
IMPROVEMENT



System Improvement

Intelligent Conversational Bot

YUN-NUNG (VIVIAN) CHEN

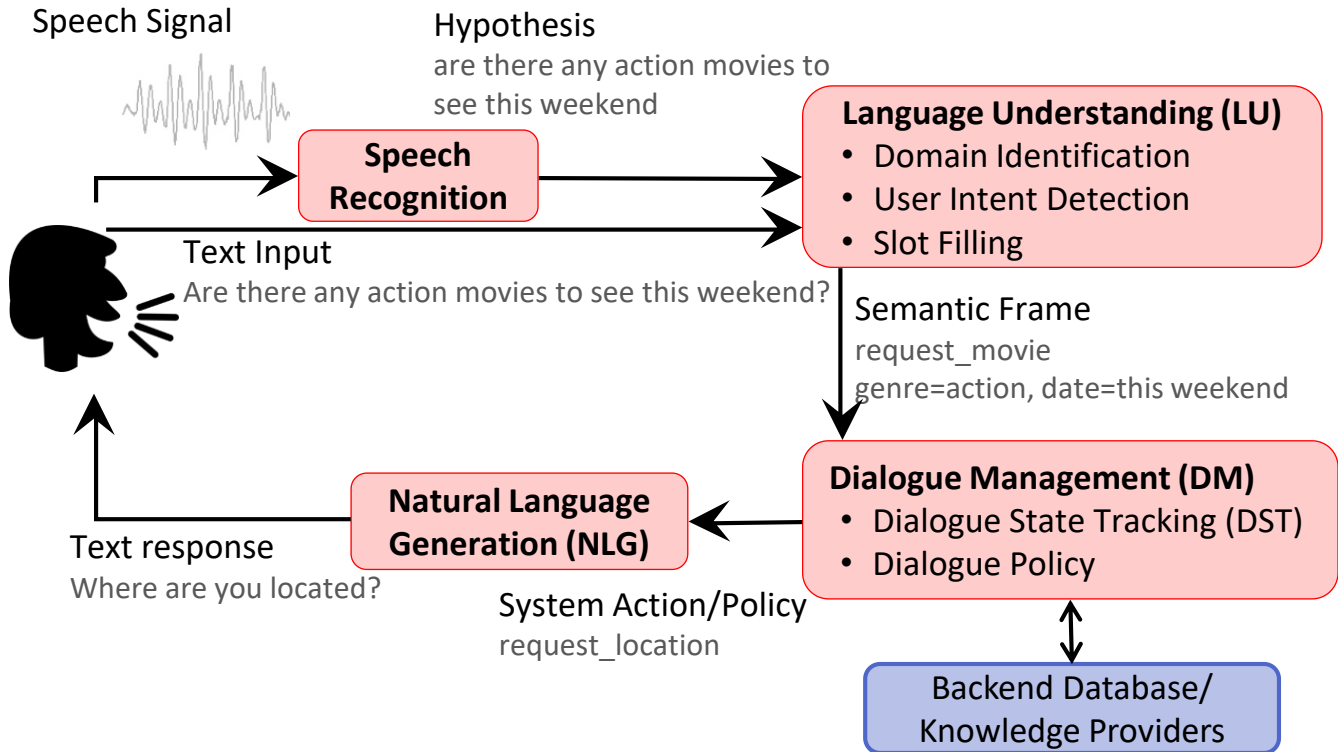
WWW.CSIE.NTU.EDU.TW/~YVCHEN/S105-ICB



國立臺灣大學
National Taiwan University

Framework

2



Demo System Release (Beta)

3

- Fill the form by testing other systems [[link](#)]
 - ▣ Record failed dialogues
 - ▣ Report to the owner team
 - X: cannot work; F: fail; S: success
 - ▣ Bonus if you test all other systems (due 6/9 Fri 23:59:59)
- **Important!**
 - ▣ Guide your users by showing some possible examples to start
 - ▣ Improve your systems by the failed dialogues
 - Final system scores will be judged using part of prior failed interactions

System Improvement

4

- **Ontology:** check whether all columns in the table can be searched as the target
- **LU:** evaluate the LU to see the *coverage* of the understanding module
 - ▣ Testing data should come from real human
 - Provide the system link to collect more dialogues and then annotate them for evaluation
- **DM:** add multi-turn interactions into the simulator for training the RL agent
 - ▣ The RL agent should handle misunderstanding better than the rule-based agent
 - Check whether the agent can handle *misrecognized* texts or *misunderstanding*
 - ▣ If the RL agent performs worse than the rule agent, increase your system complexity
 - More functionality/backend databases, more complex simulated interactions
 - Please check the strategies [this agent](#) applied to make sure your RL agent has increasing performance trend
- **NLG:** improve diverse and interesting responses
- **Multimodality:** try richer multimodality for interesting interactions
 - ▣ Emotion recognition, speaker recognition, etc for better greeting

Final Score

5

- System functionality
 - ▣ #tables, #slots, #intents
- System success performance
 - ▣ Human testing performance evaluated by TAs
 - ~30 dialogues
 - If the failed dialogues are fixed, we use the refined performance
- Evaluation
 - ▣ Correctness and reasonability
 - Testing data should be from real human instead of generated patterns
- **Creativity**
 - ▣ Multimodality usage (e.g. emotion)
 - ▣ Diverse/interesting responses
- **The poster template will be provided**



Milestone Score

6

- ❑ Regrade if updating your system to support the failed interactions
- ❑ Required documentations/programs can be re-submitted and regraded (max 80%)
- ❑ TA will be evaluated via the teams' feedback