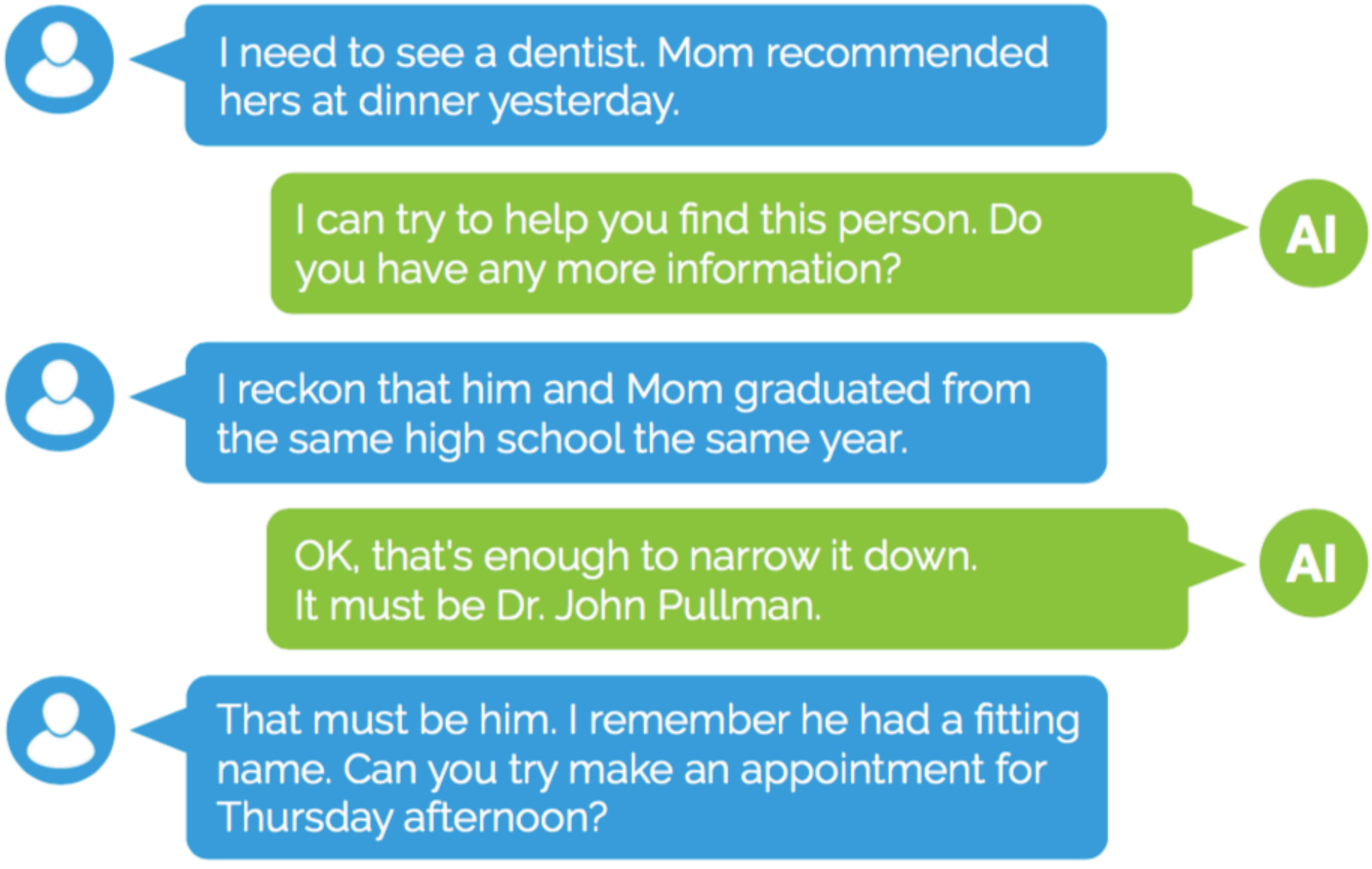
QRFA: A Data-Driven Model of Information Seeking Dialogues

Svitlana Vakulenko, Kate Revoredo, Claudio Di Ciccio, Maarten de Rijke

Conversational Search

RQ: What is the structure of an information-seeking dialogue?



Why Analyse Conversations?

Grounded theory - simple interpretable model based on data

Enable

- Human-in-the-loop
- Collaboration & Discussion
- Breakdown detection

Inform

- System design
- Data collection & cleaning
- System evaluation

https://blogs.nasa.gov/ISS_Science_Blog/2010/12/10/post_1291390819440/

Conversation Mining

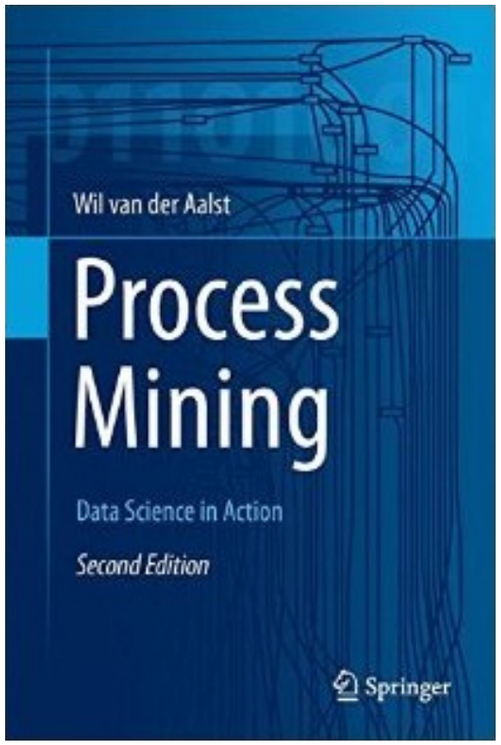
WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

1. Dialogue **annotation**
   - supervised classification

2. Process **discovery** (model "as-is")
   - frequent sequence mining

3. Process **repair** (model "to-be")

4. Process **evaluation**
   - conformance checking



Wil M. P. van der Aalst: Process Mining - Data Science in Action, Second Edition. Springer 2016

Mining Information-Seeking Dialogues

4 publicly available **datasets**

Seeker (User) - Intermediary (Agent)

- **Spoken Conversation Search** 39 dialogues 13 labels
- **Open Data Exploration** 26 dialogues 20 labels
- **DSTC1** bus schedules 15,866 dialogues 37 labels
- **DSTC2** restaurant reservation 2,118 dialogues 21 labels

QRFA Annotation Schema

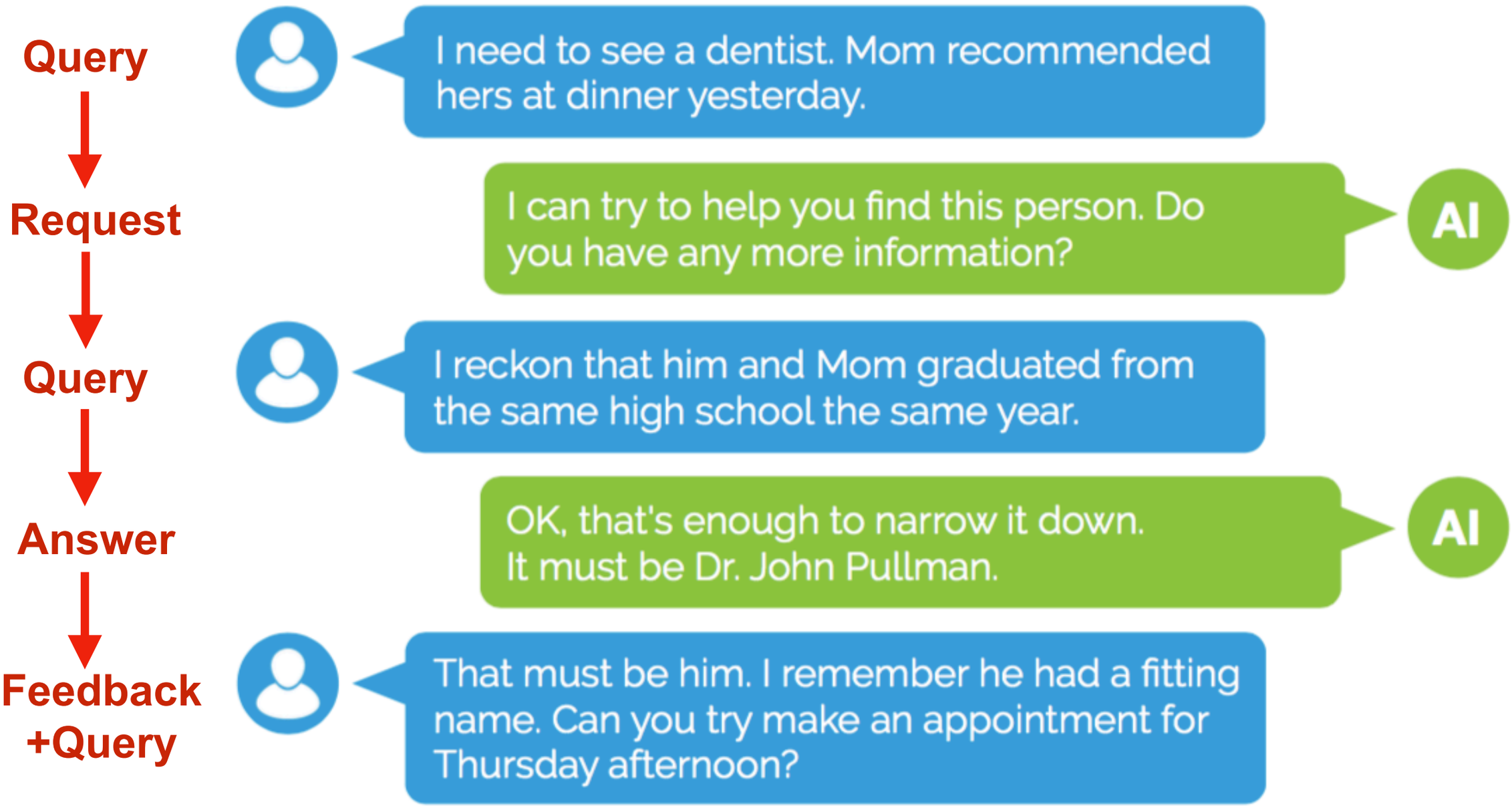
		Proactive	
User	**Q**uery	Information	
		Prompt	
Agent	**R**equest	Offer	
		Understand	

QRFA Annotation Schema

	Proactive		Reactive	
User	**Q**uery	Information	**F**eedback	Positive
		Prompt		Negative
Agent	**R**equest	Offer	**A**nswer	Results
		Understand		Backchannel
				Empty

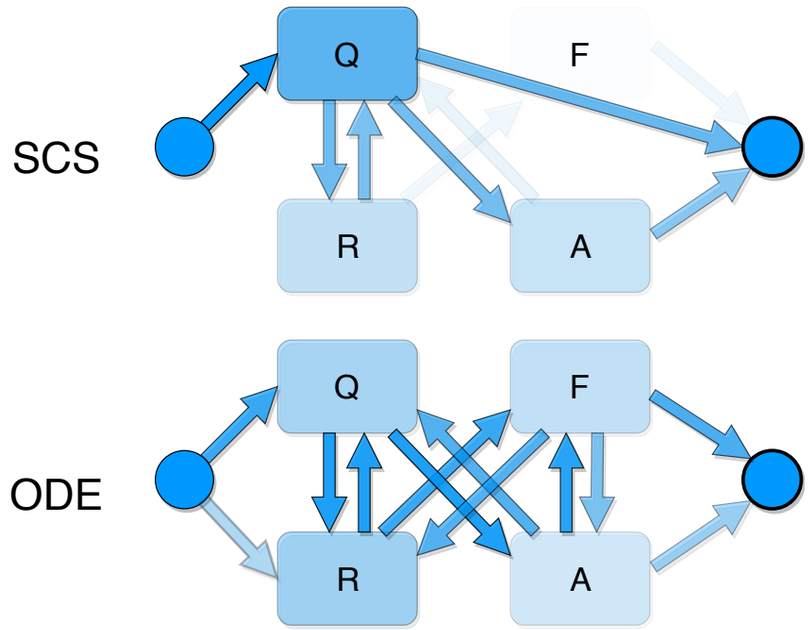
Dialogue Annotation

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS



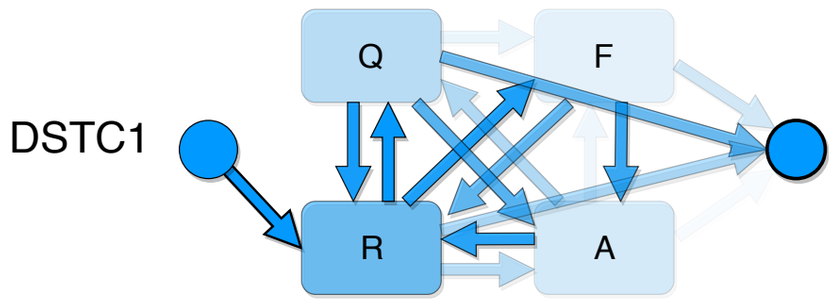
QRFA Model

- **SCS** 39 web search dialogues

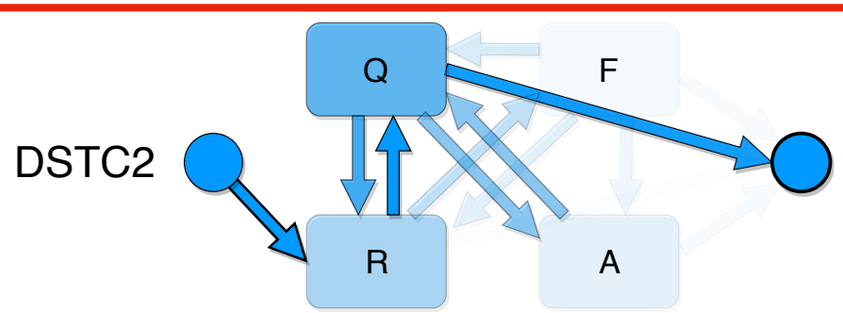


- **ODE** 26 Open Data dialogues

- **DSTC1** 15,866 bus schedule dialogues



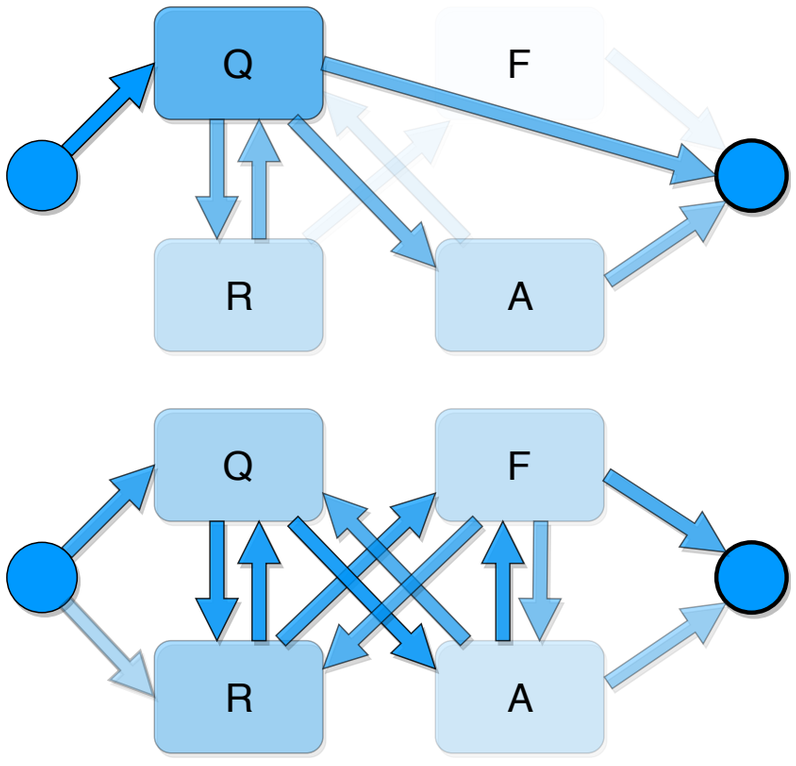
- **DSTC2** 2,118 restaurants dialogues



QRFA Model

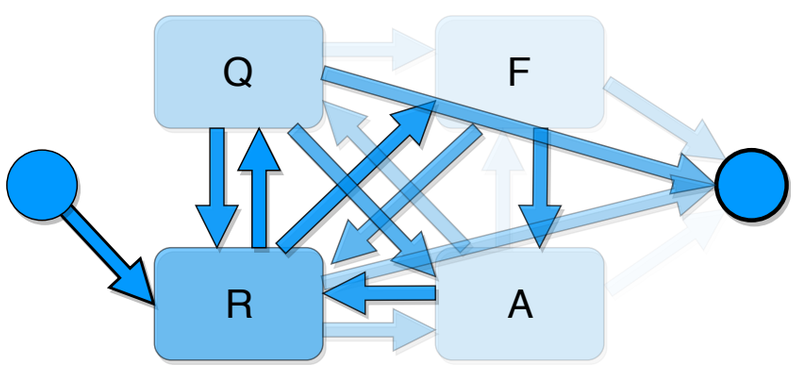
WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

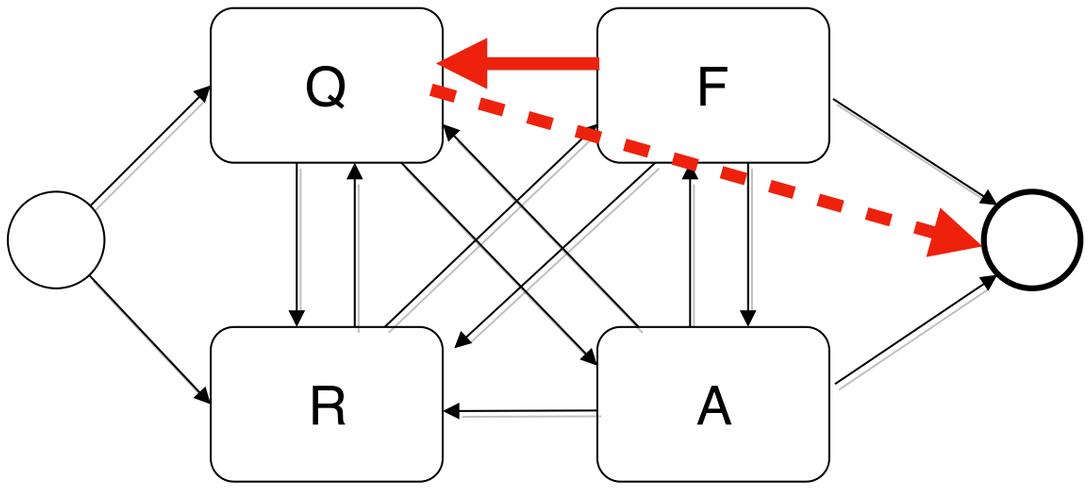
SCS



ODE

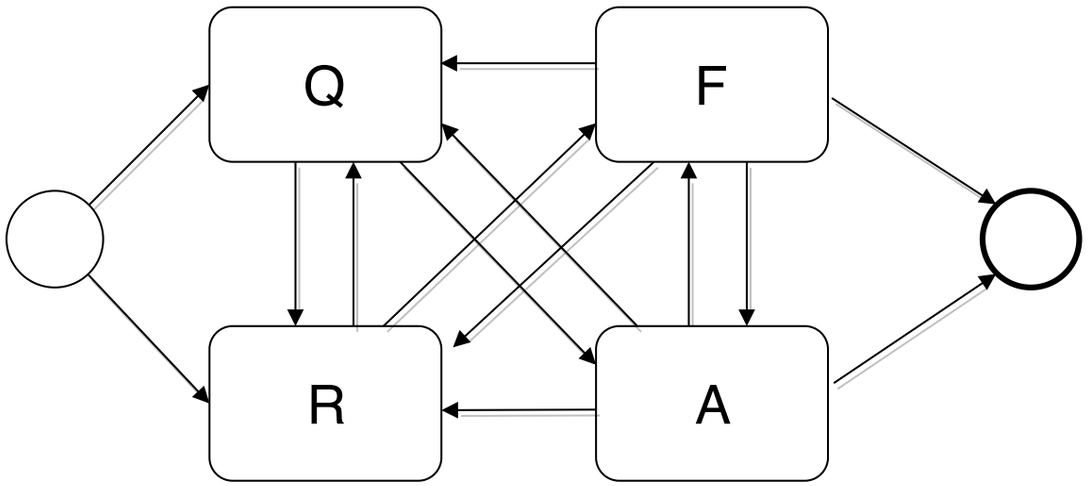
DSTC1



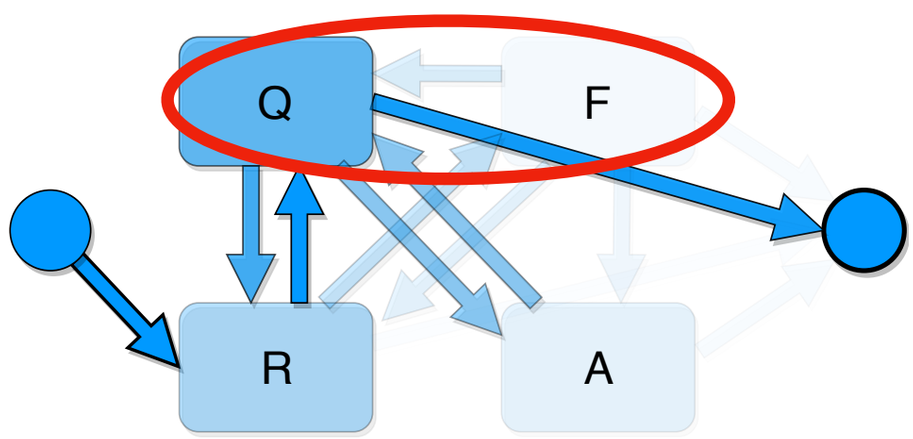


Model Evaluation

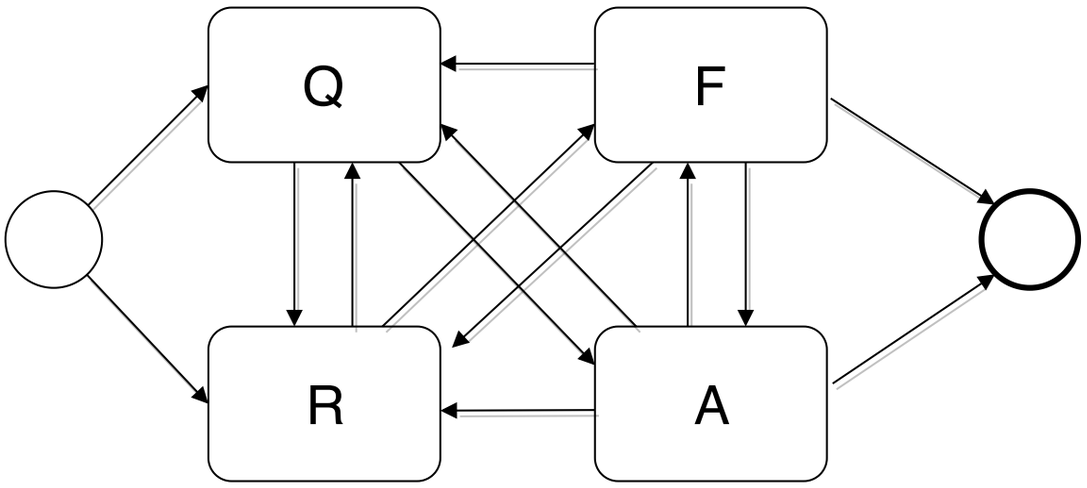
- **Fitness**/case: Min 0.65 Max 1 Average 0.96 SD 0.04
- Cases with value 1: 0.61
- **Break-down detection:** P 0.92 R 0.55

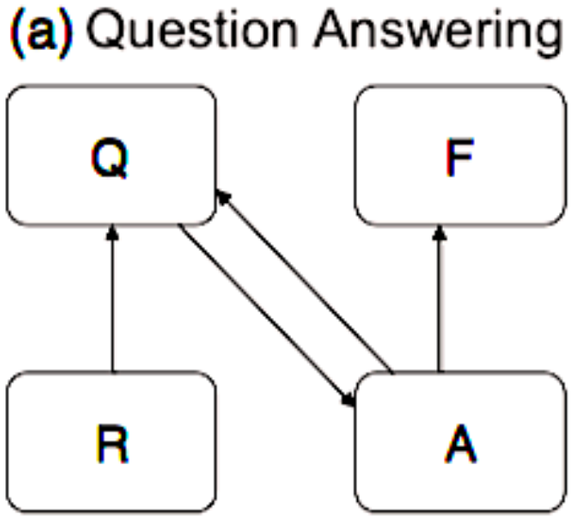


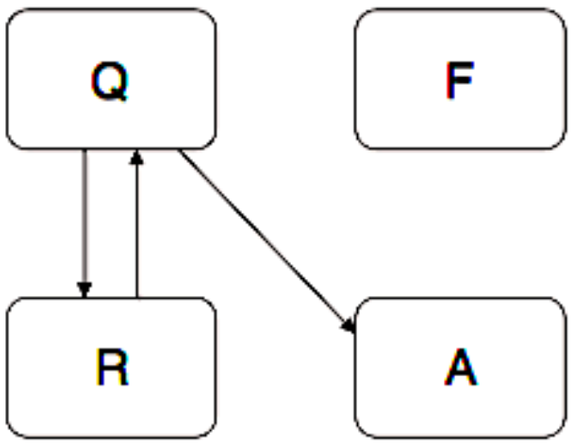
DSTC2

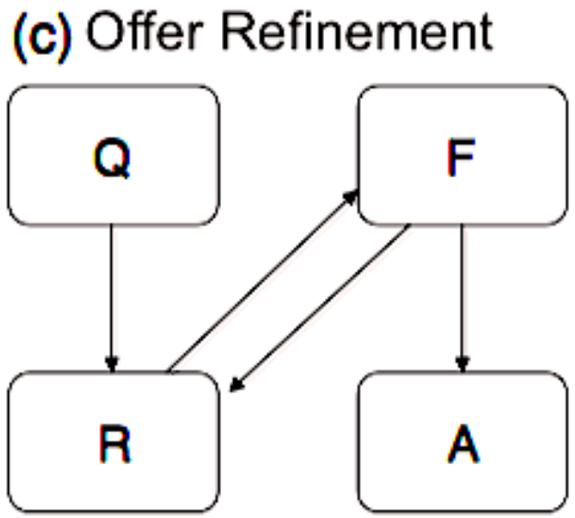


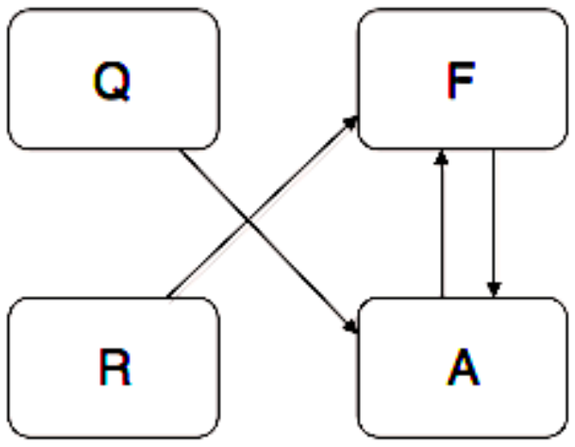
Loops as Interaction Modes











Conclusions and Future Work

- We proposed **conversation mining** for analysing transcripts
- and used it to extract **QRFA** model of information-seeking dialogues
- Future work: **complex** models integrating more labels & datasets

Data + Code: https://github.com/svakulenk0/conversation_mining

@svakulenk0 @katerevoredo @mdr

Acknowledgements

THIS WORK WAS SUPPORTED BY THE FOLLOWING PROJECTS:

EU H2020 PROGRAMME UNDER THE MSCA-RISE AGREEMENT 645751 (RISE_BPM)

PROJECT OPEN DATA FOR LOCAL COMMUNITIES FUNDED BY THE AUSTRIAN FEDERAL MINISTRY OF TRANSPORT, INNOVATION AND TECHNOLOGY (BMVIT) UNDER THE PROGRAM "ICT OF THE FUTURE“, BETWEEN NOVEMBER 2016 AND APRIL 2019.

MORE INFORMATION HTTPS://IKTDERZUKUNFT.AT/EN/

Austrian Ministry
for Transport,
Innovation and Technology