HELPR: A Framework to Break the Barrier across Domains in Spoken Dialog Systems

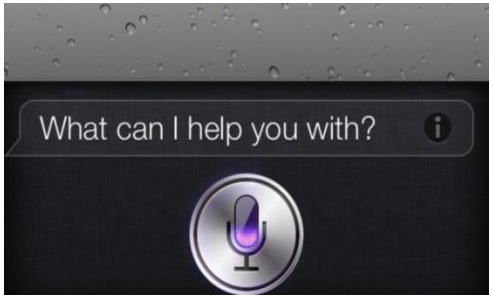
Ming Sun

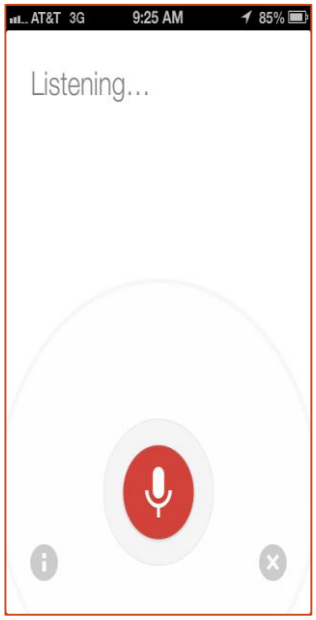
Yun-Nung Chen

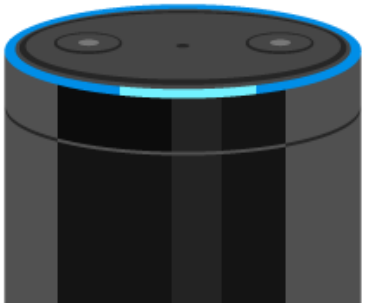
Alexander I. Rudnicky

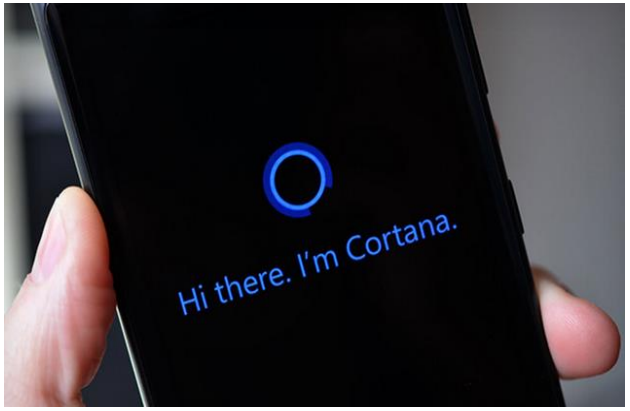
School of Computer Science

Carnegie Mellon University

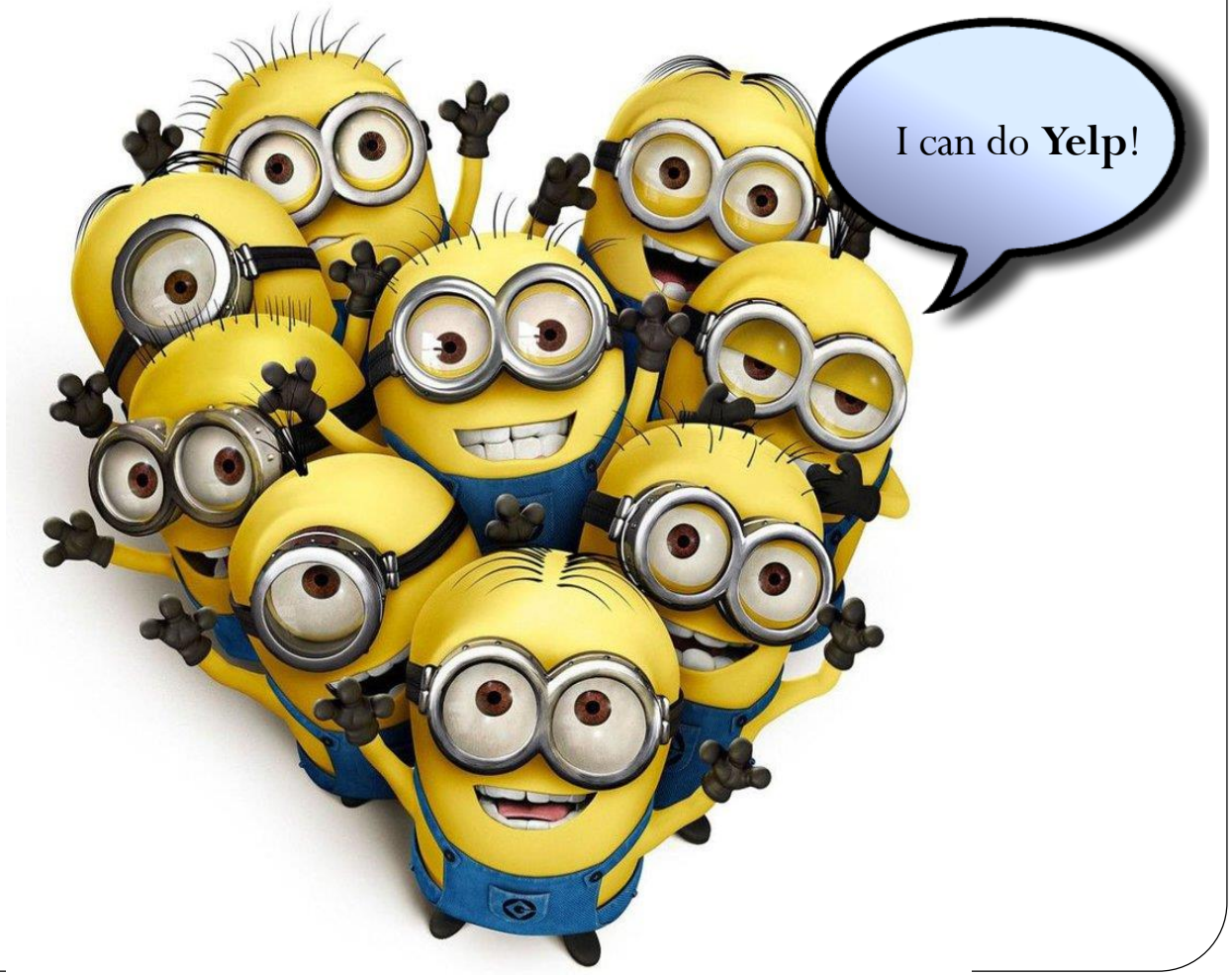

What can I help you with?

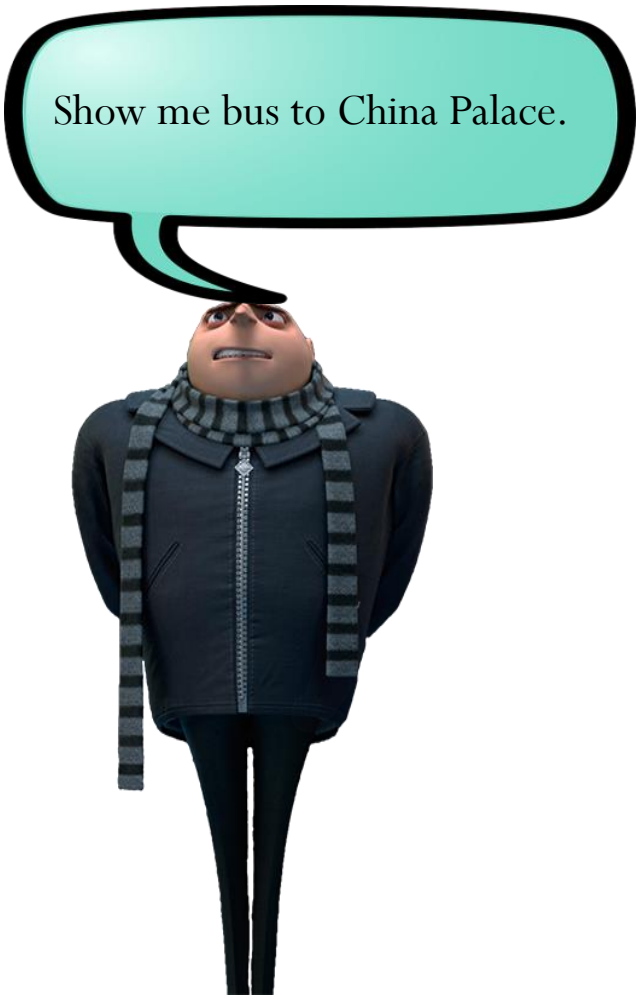

AT&T 3G
9:25 AM
85%
Listening...

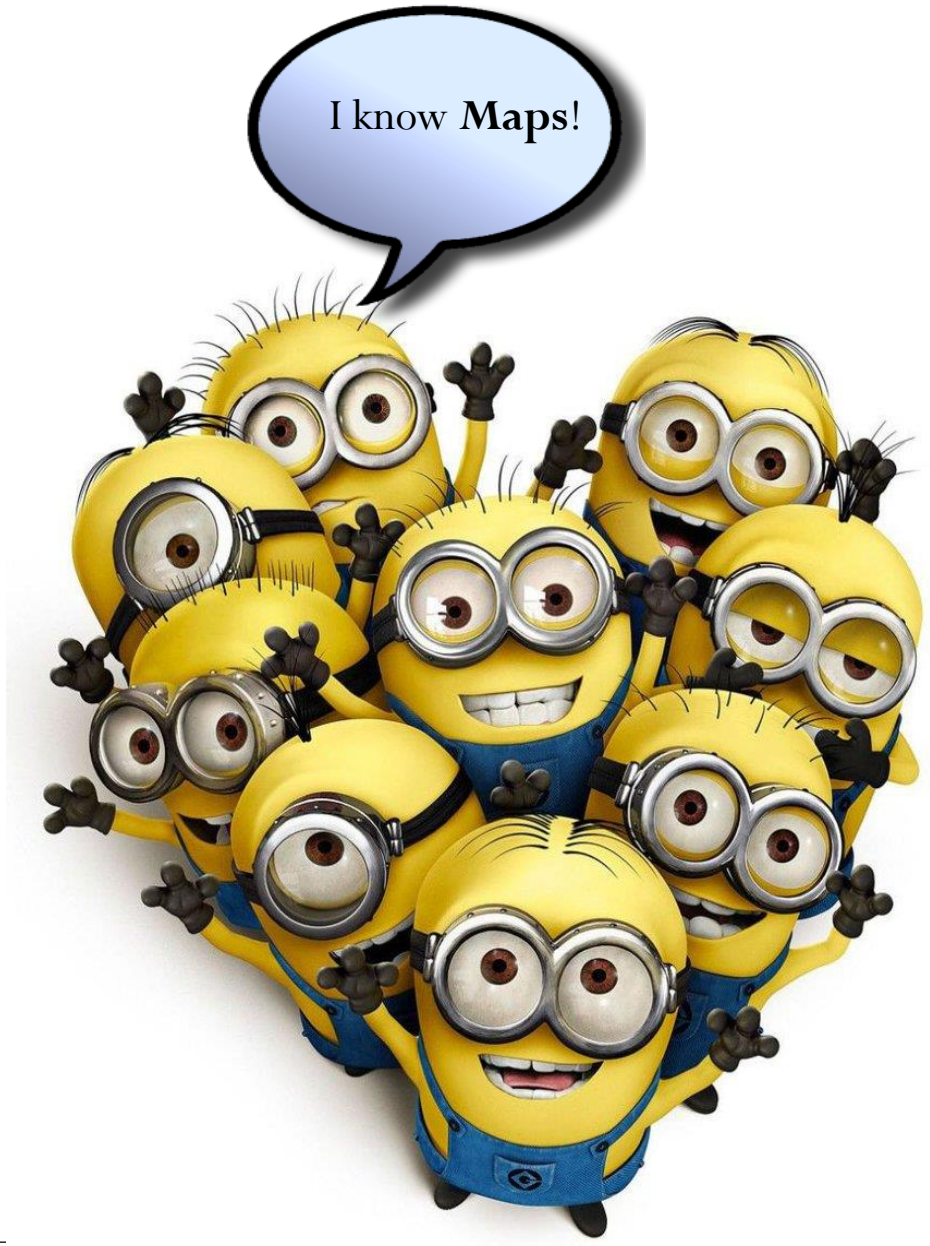



Hi there. I'm Cortana.


Book a restaurant!

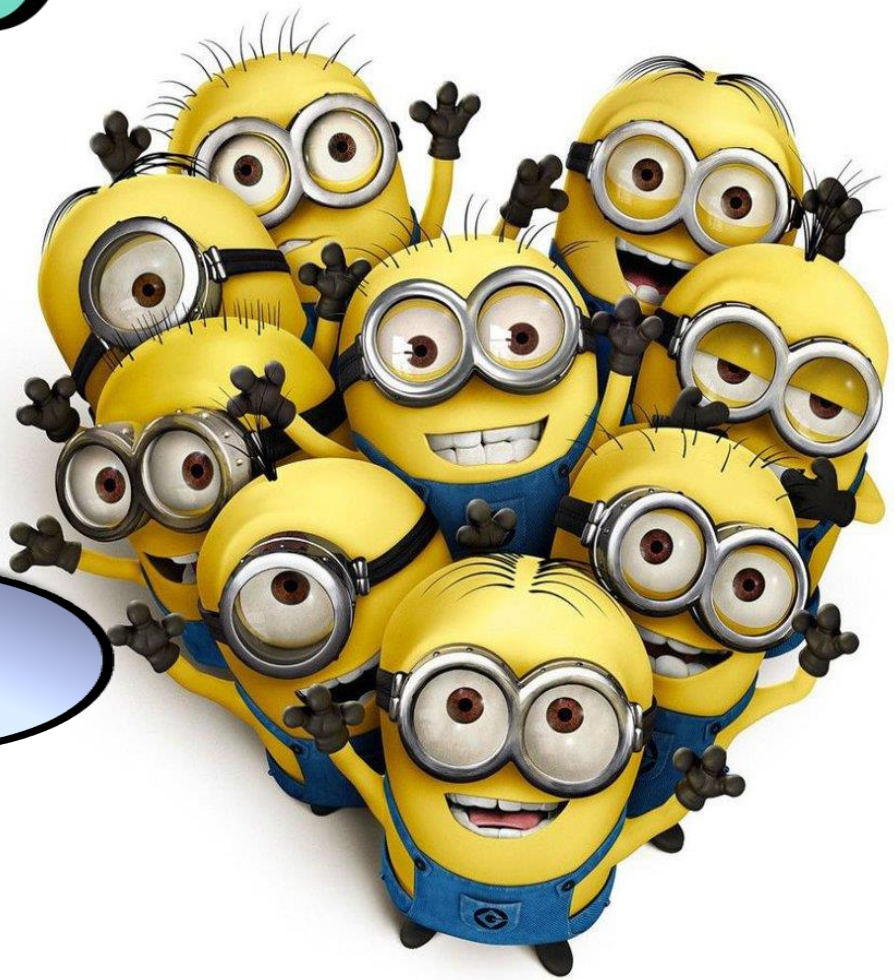

I can do **Yelp**!


Show me bus to China Palace.


I know **Maps**!

Tell kids we will have Chinese tonight.

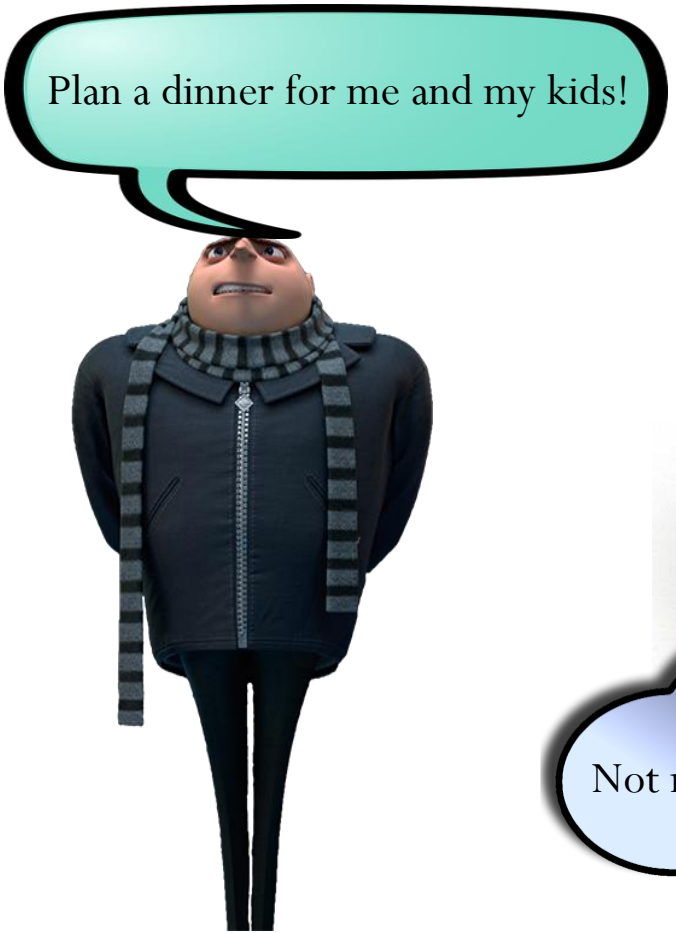
I will **Text**!

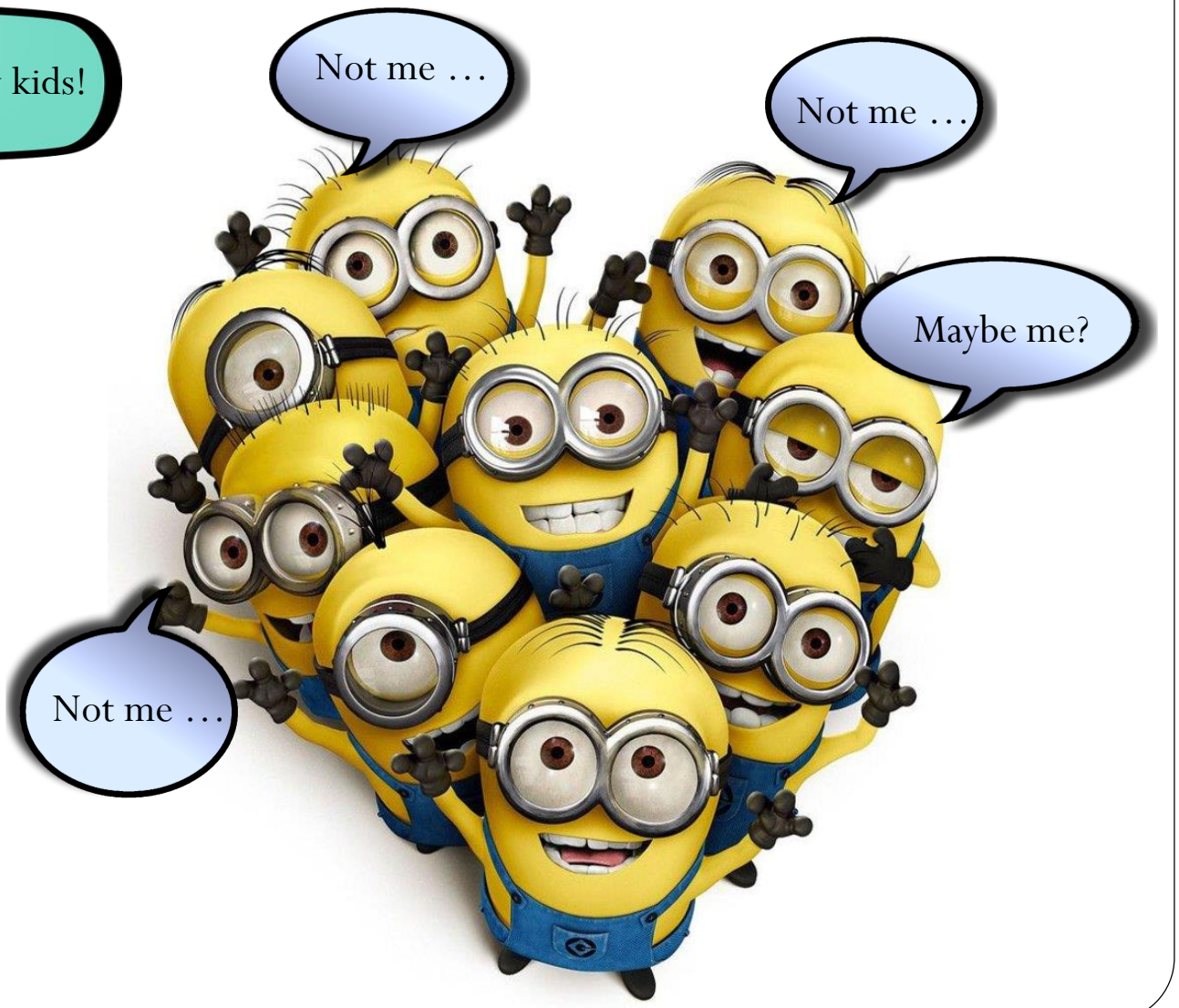


Problems (potential improvement)

- **Cannot handle complex intention**
- **Passively support cross-domain dialog**
- **No shared context**

Problem 1 --- Complex Intention





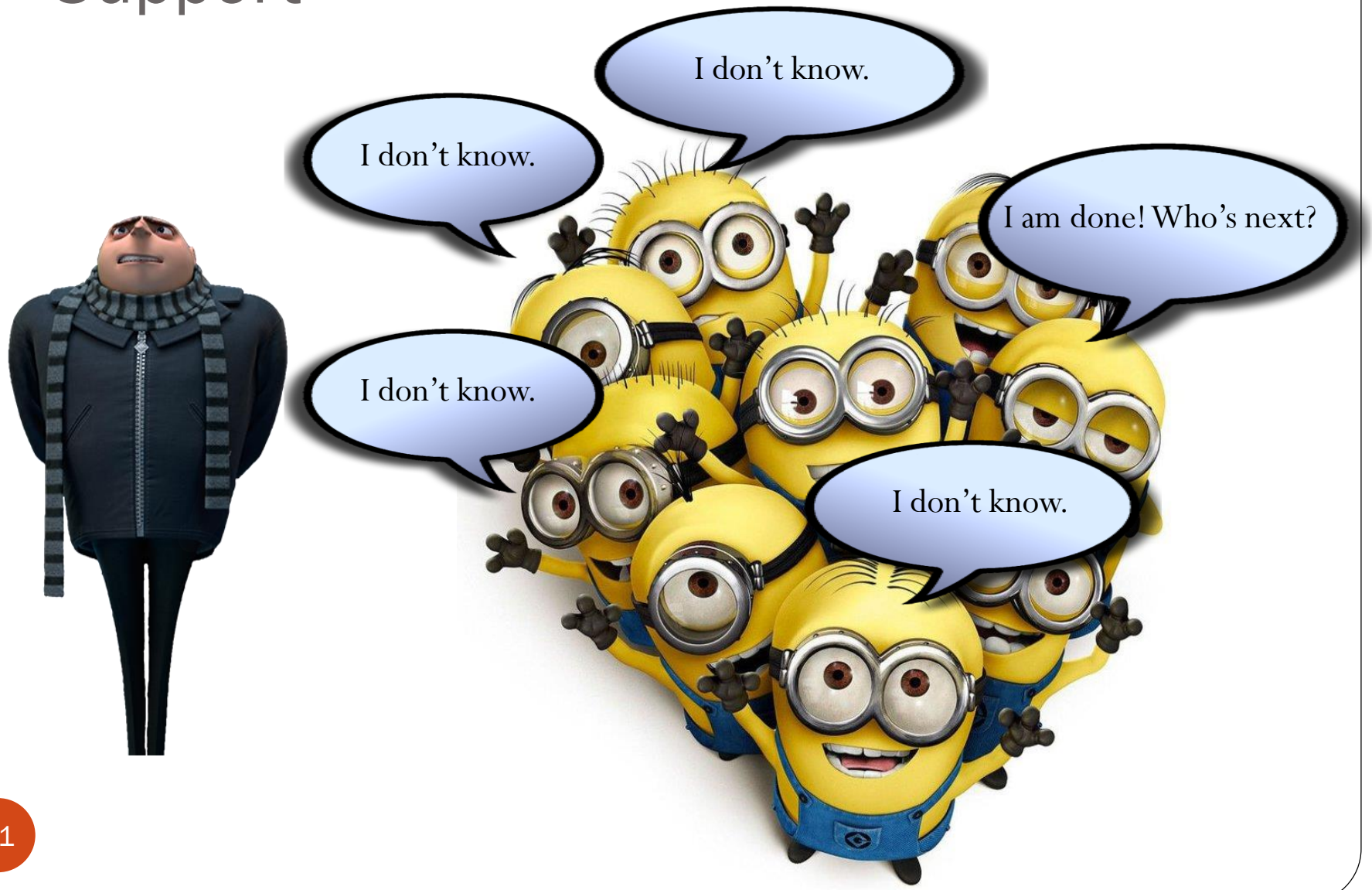
Problem 1 --- Complex Intention

- Given a high-level/complex input, **ONE** closest domain is selected

Problem 2 --- Passive Cross-Domain Support

Problem 2 --- Passive Cross-Domain Support

Problem 2 --- Passive Cross-Domain Support



Problem 2 --- Passive Cross-Domain Support

- Given the current context, not aware of the next step/domain.

Problem 3 --- No Shared Context

Problem 3 --- No Shared Context

Problem 3 --- No Shared Context

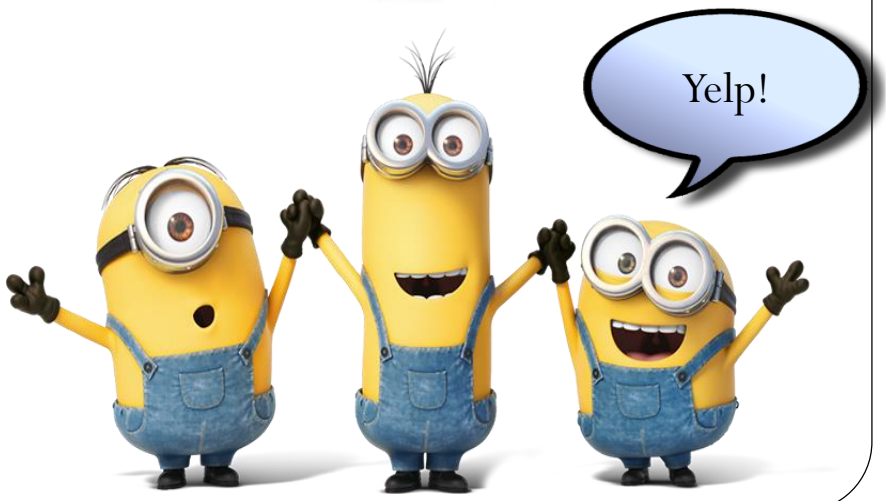
- Not aware of the relationship with previous domains.

Solutions

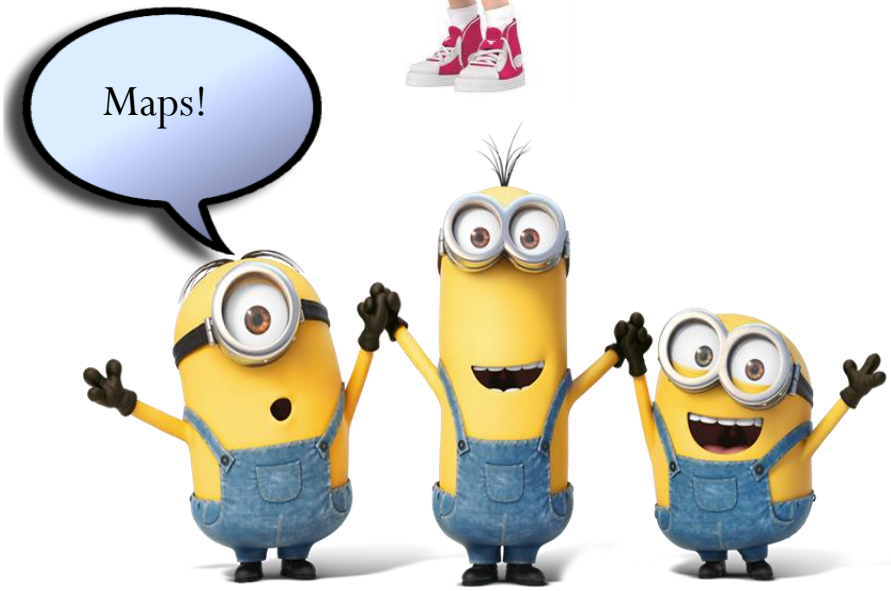
- Cannot handle complex intention
  - **Sol**: Intention Recognition and Understanding (IWSDS'16, IUI'16)
- Passively support cross-domain dialog
  - **Sol**: Context-aware domain prediction (NIPS-SLU'15)
- No shared context
  - **Sol**: co-reference resolution ('there' -> 'China Palace')

Intention Understanding

Intention Understanding



Intention Understanding



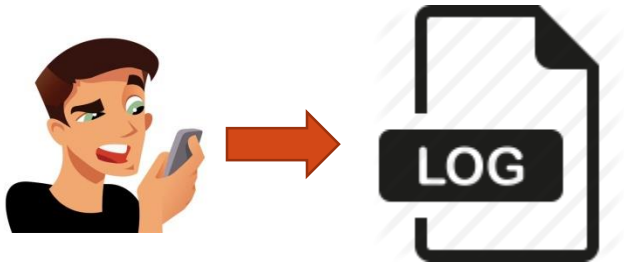
Intention Understanding

Intention Understanding

Approaches

- Step 1: Observe human user perform multi-domain tasks
  - Data collection
- Step 2: Learn to assist at task level
  - Map an input command to a set of domains
  - Talk at task level

Data Collection 1 – Smart Phone



- Log app invocation + time/date/location
- Separate log into episodes if there is 3 minute inactivity

Schedule a visit to CMU Lab

Wednesday
17:08 – 17:14
CMU

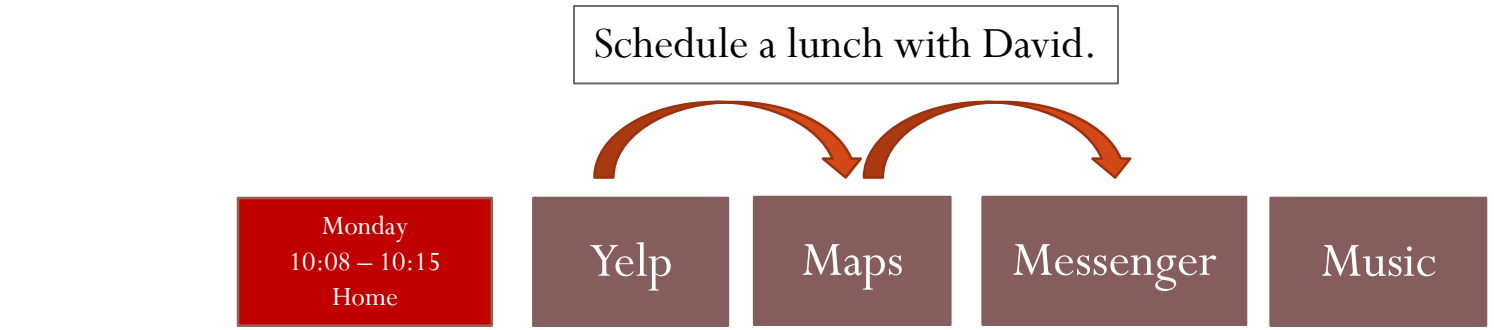
Messenger

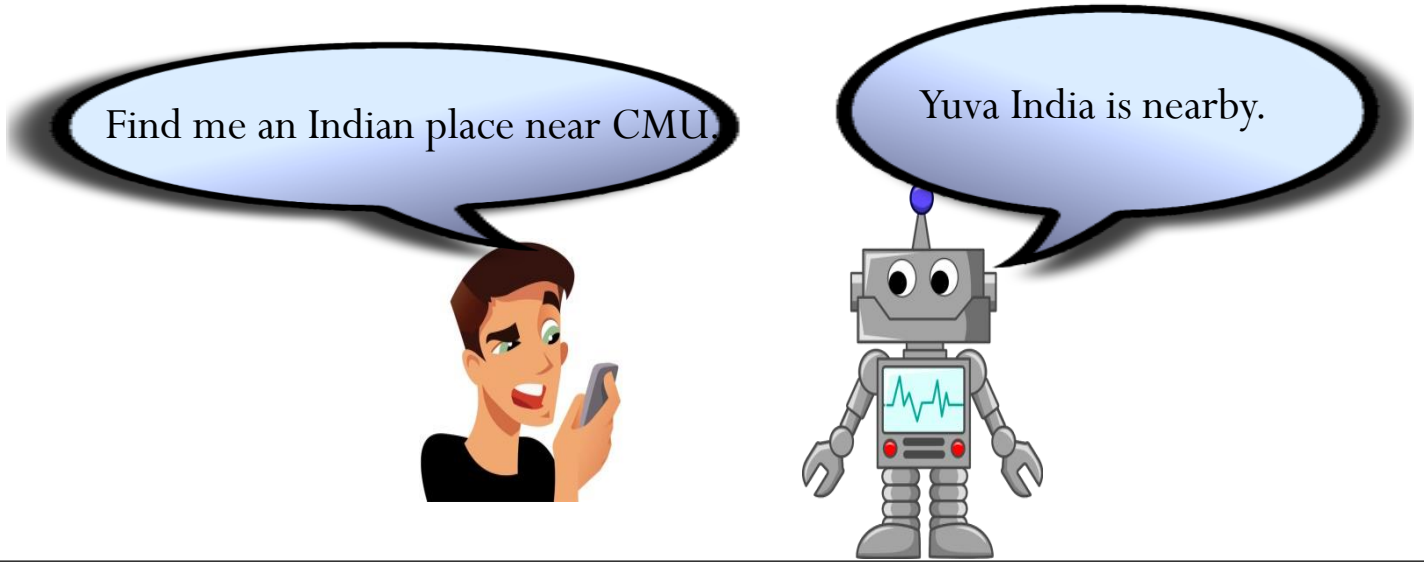
Gmail

Browser

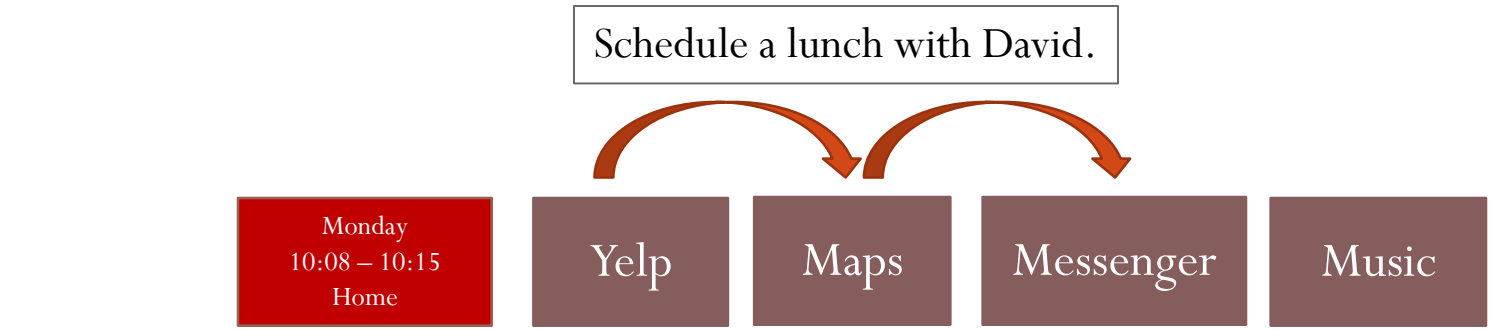
Calendar

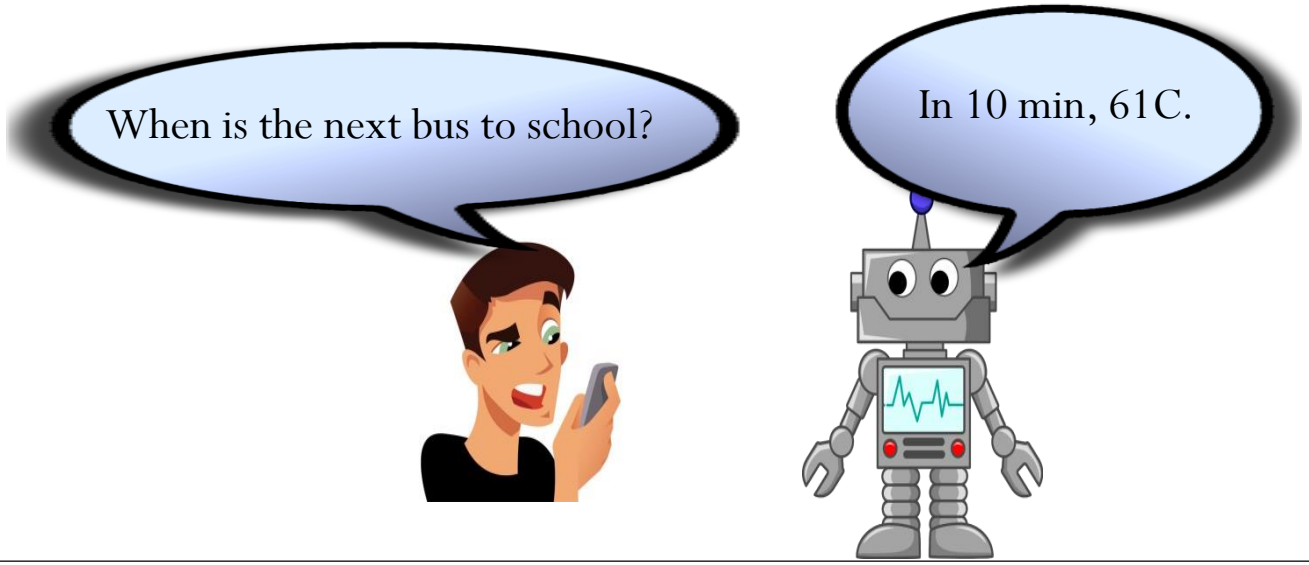
Data Collection 2 – Wizard-of-Oz



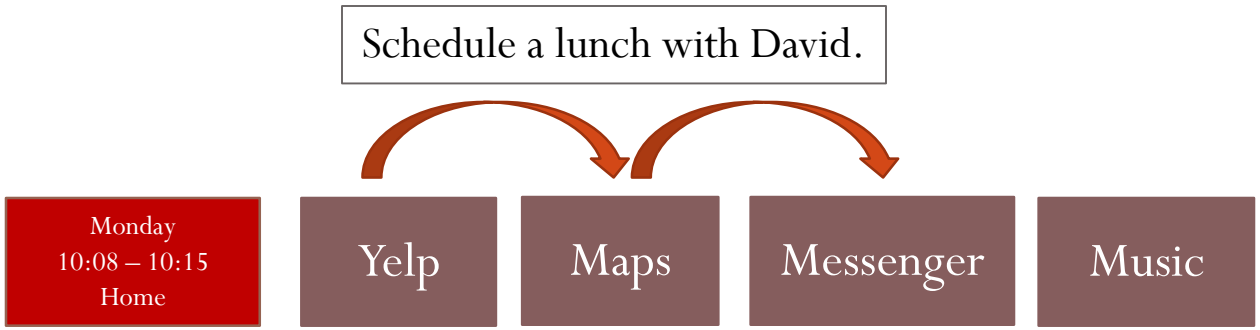


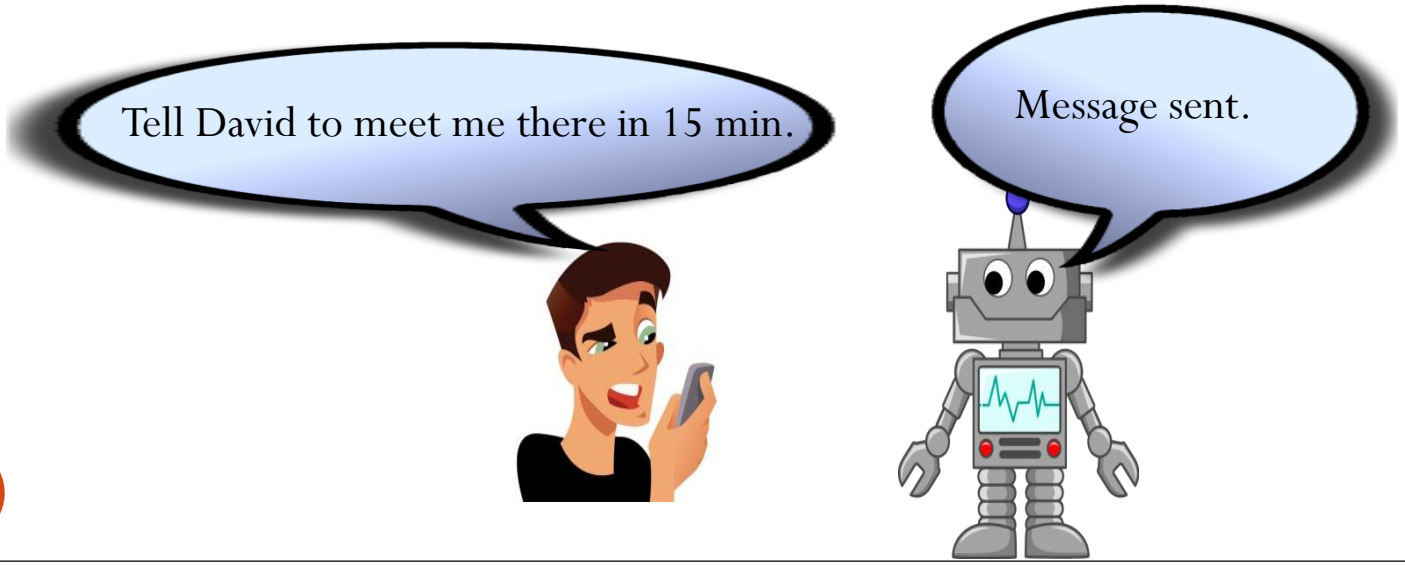
Data Collection 2 – Wizard-of-Oz





Data Collection 2 – Wizard-of-Oz

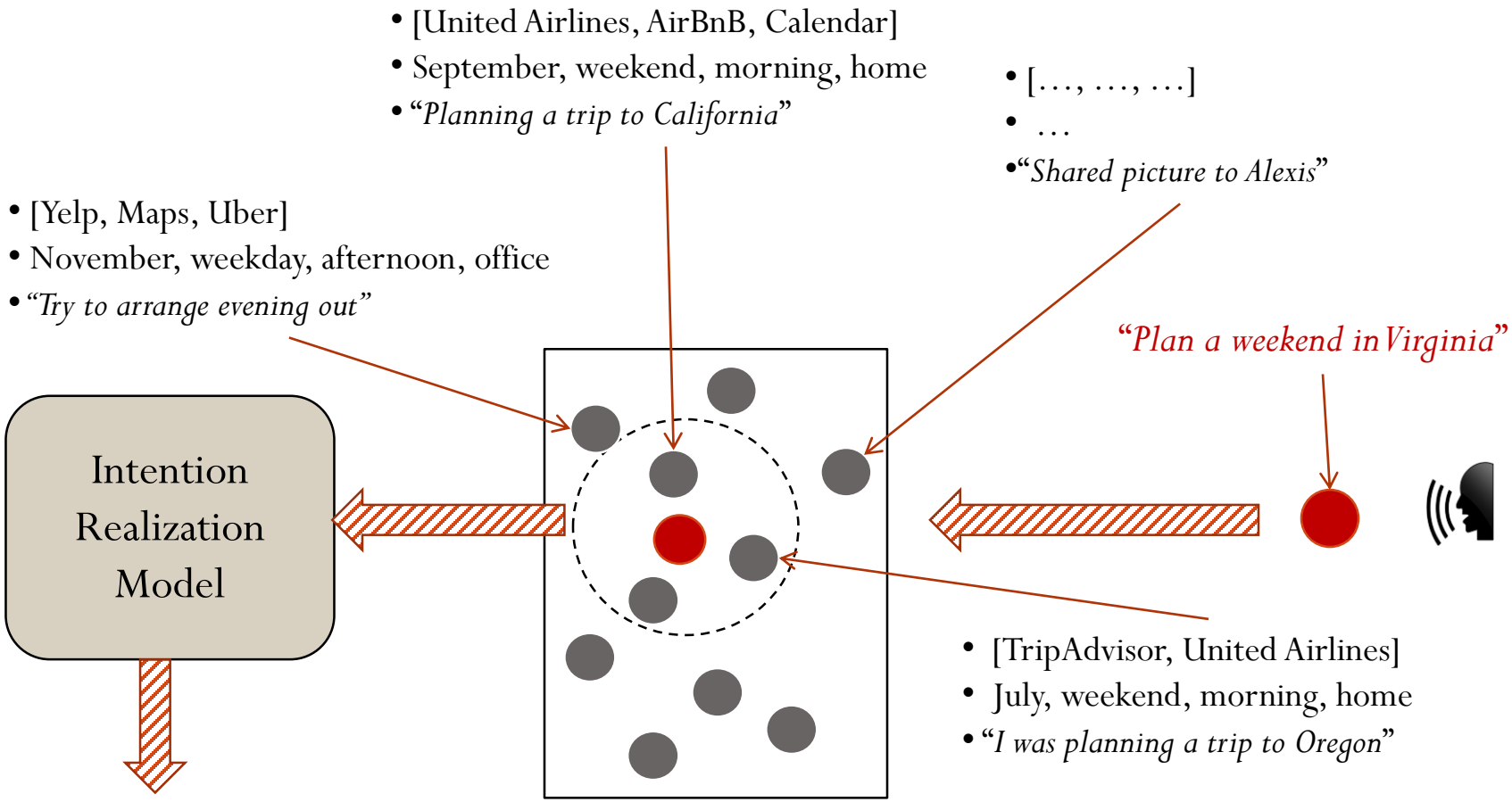




Data Collection Summary

- 533 real-life multi-domain interactions from 14 real users
  - 12 native English speakers
  - 4 males & 10 females
  - Mean age: 31
  - Total # unique apps: 130 (Mean = 19/user)

Resources	Examples	Usage
App sequences	Yelp->Maps->Messenger	structure/arrangement
Task descriptions	"Schedule a lunch with David"	nature of the intention, language reference
User utterances	"Find me an Indian place near CMU."	language reference
Meta data	Monday, 10:08 – 10:15, Home	contexts of the tasks

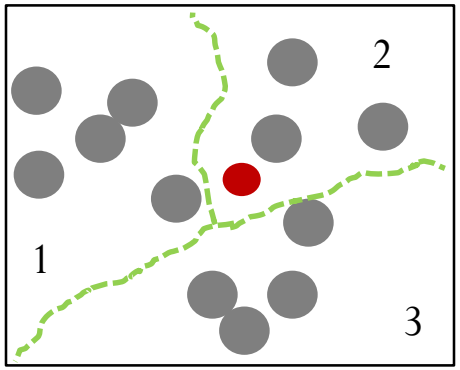


Infer:

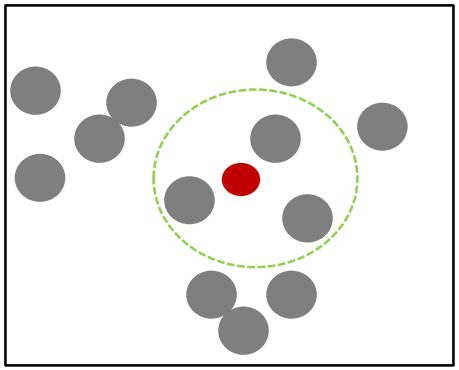
1) Supportive Domains: *United Airlines*, *TripAdvisor*, *AirBnB*
2) Summarization: "*plan trip*"

Find similar past experience

- Cluster-based:
  - K-means clustering on user generated language
- Neighbor-based:
  - KNN



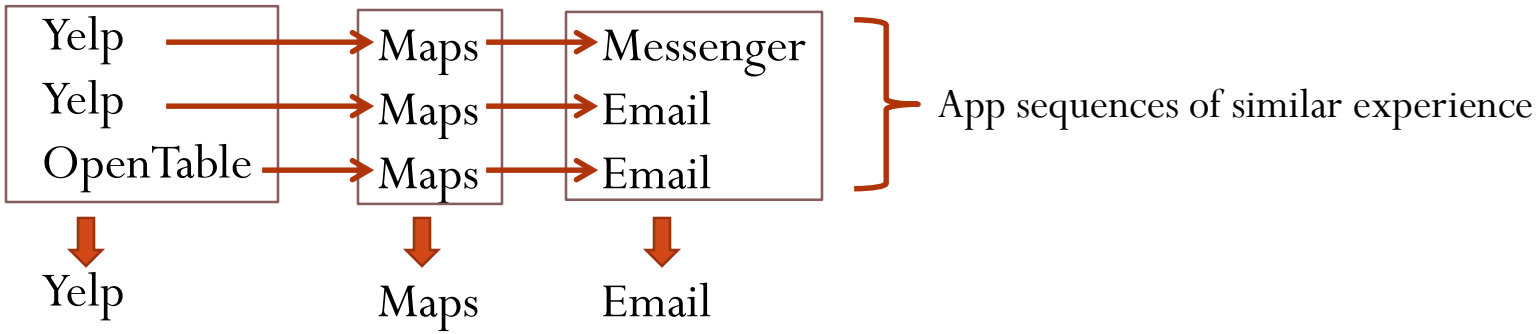
Cluster-based



Neighbor-based

Realize domains from past experience

- Representative Sequence



- Multi-label Classification

Obstacles

- Language-mismatch
  - **Sol**: Query Enrichment (QryEnr)
  - [“shoot”, “photo”] -> [“shoot”, “take”, “photo”, “picture”]

- App-mismatch
  - **Sol**: Ideally, use functionality (extracted from app descriptions) to compute distance between two apps

Realize domains from language

	Personalized			Generic		
	Precision	Recall	F	Precision	Recall	F
Cluster-based	**71.4**	47.4	55.7	48.4	16.5	23.8
+QryEnr	63.5	**51.6**	**56.3**	31.1	27.6	27.9
+AppSim	**71.4**	47.4	55.7	**56.5**	19.1	27.8
+QryEnr+AppSim	63.5	**51.6**	**56.3**	40.4	**35.1**	**36.1**
Neighbor-based	**72.2**	41.6	51.3	45.8	12.3	19.1
+QryEnr	68.1	**49.9**	**57.0**	35.2	17.4	22.9
+AppSim	**72.2**	41.6	51.3	**55.6**	14.6	22.7
+QryEnr+AppSim	68.1	**49.9**	**57.0**	42.2	**21.5**	**28.0**

Realize domains from language

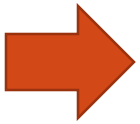
	Personalized			Generic		
	Precision	Recall	F	Precision	Recall	F
Cluster-based	**71.4**	47.4	55.7	48.4	16.5	23.8
+QryEnr	-	+	+	-	+	+
+AppSim	**N/A**	N/A	N/A	+	+	+
+QryEnr+AppSim	-	+	+	-	+	+
Neighbor-based	**72.2**	41.6	51.3	45.8	12.3	19.1
+QryEnr	-	+	+	-	+	+
+AppSim	**N/A**	N/A	N/A	+	+	+
+QryEnr+AppSim	-	+	+	-	+	+

Talk at task level

- Techniques:
  - (Extractive/abstractive) summarization
  - Key phrase extraction

- User study:
  - Key phrase extraction + user generated language
  - Ranked list of key phrases + user's binary judgment

Looking up math problems. (Desc)
Now open a browser. (Utt)
Go to slader.com. (Utt)
Doing physics homework. (Desc)
…
Go to my Google drive. (Utt)
Look up kinematic equations. (Utt)
Now open my calculator so I can plug in numbers. (Utt)

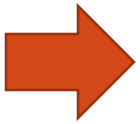


1. solutions online
2. project file
3. Google drive
4. math problems
5. physics homework
6. answers online
…

Talk at task level

- Metrics
  - Mean Reciprocal Rank (MRR)
- Result:
  - MRR > 0.6
    - understandable verbal reference among the top 2 of the list

Looking up math problems. (Desc)
Now open a browser. (Utt)
Go to slader.com. (Utt)
Doing physics homework. (Desc)
…
Go to my Google drive. (Utt)
Look up kinematic equations. (Utt)
Now open my calculator so I can plug in numbers. (Utt)



1. solutions online
2. project file

3. Google drive
4. math problems
5. physics homework
6. answers online
…

Conclusion

- Collected real-life cross-domain interactions from real users
- HELPR framework to learn to assist at task level
  - Suggest a set of supportive domains to accomplish the task
    - Personalized model > Generic model
    - The gap can be reduced by QryEnr + AppSim
  - Generate language reference to communicate verbally at task level

Future Work

- From multiple domains to a **unified** dialog
  - Dynamically and automatically configure the dialog flow
  - Actively maintain a shared context
    - Reduce redundancy
    - Improve naturalness

Thanks!

- References:

  - M. Sun, YN. Chen and A. I. Rudnicky, “An Intelligent Assistant for High-Level Task Understanding”. (to appear) ACM Conference on Intelligent User Interfaces (IUI), 2016.
  - M. Sun, YN. Chen and A. I. Rudnicky, "HELPR: A Framework to Break the Barrier across Domains in Spoken Dialog Systems". (to appear) International Workshop on Spoken Dialogue Systems (IWSDS) 2016.
  - M. Sun, YN. Chen and A. I. Rudnicky, "Understanding User's Cross-Domain Intentions in Spoken Dialog Systems". NIPS Workshop on Machine Learning for SLU and Interaction 2015.