Analyzing and Characterizing User Intent in Information-seeking Conversations

Chen Qu[1], Liu Yang[1], W. Bruce Croft[1], Johanne R. Trippas[2], Yongfeng Zhang[3], Minghui Qiu[4]
[1] University of Massachusetts, Amherst, [2] RMIT University, [3] Rutgers University, [4] Alibaba Group

Abstract

Research Problem: understanding and characterizing how people interact in **information-seeking** conversations

MSDialog:

- **QA** interactions between information seekers and providers
- **Multi-turn technical support** dialogs on Microsoft products
- More than 2,000 QA dialogs with 10,000 utterances
- Annotated with **user intent** on the utterance level using **crowdsourcing**
- Freely available to encourage exploration of information-seeking conversation models `https://ciir.cs.umass.edu/downloads/msdialog/`

Data Characterization: **user intent distribution, co-occurrence, and flow patterns**.

Key Properties and Comparisons

Dataset	Multi-turn	Human-human	Information-seeking	User intent label
DSTC 1-3	✓			
DSTC 4-5	✓	✓		
Switchboard	✓	✓		
Twitter Corpus	✓	✓		
DSTC 6 (2nd Track)	✓	✓	✓	
Ubuntu Dialog Corpus	✓	✓	✓	
MSDialog	✓	✓	✓	✓

Data Collection and Filtering

Data Collection:

- Crawled over 35,000 dialogs from **Microsoft Community**
- The forum is **well-moderated** and contains **user-generated questions** with **high-quality answers**
- The answers are provided by **Microsoft staff**, community moderators, article authors, and other experienced users including Microsoft Most Valuable Professionals.

Data Selection Criteria: To ensure **quality** and **consistency**, we use the following criteria:

- With 3 to 10 turns
- With 2 to 4 participants
- With at least one correct answer selected by the community
- Falls into one of the categories of Windows, Office, Bing, and Skype, which are the major categories of Microsoft products

Data Statistics

Items	Min	Max	Mean	Median
# Turns Per Dialog	3	10	4.56	4
# Participants Per Dialog	2	4	2.79	3
Dialog Length (Words)	27	1,467	296.90	241
Utterance Length (Words)	1	939	65.16	47

User Intent Distribution and Co-occurrence

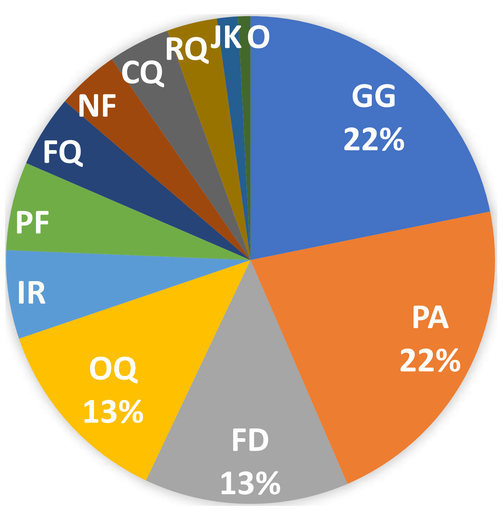

Figure: Distribution

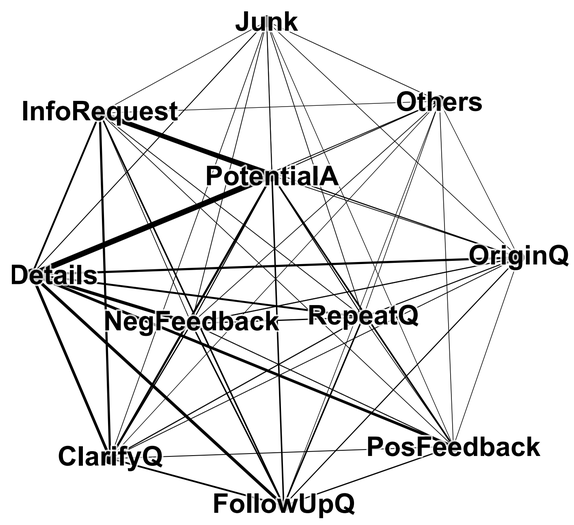

Figure: Co-occurrence

User Intent Flow Pattern

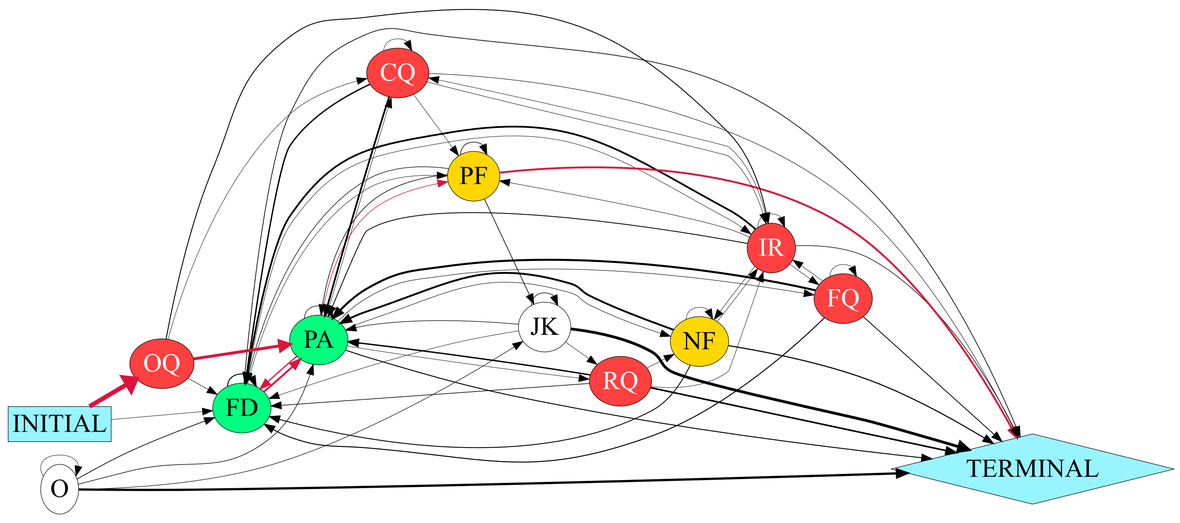

Figure: Flow pattern with a Markov model. Node colors: red (questions), green (answer related), yellow (feedback). Edges are directed and weighted by transition probability.

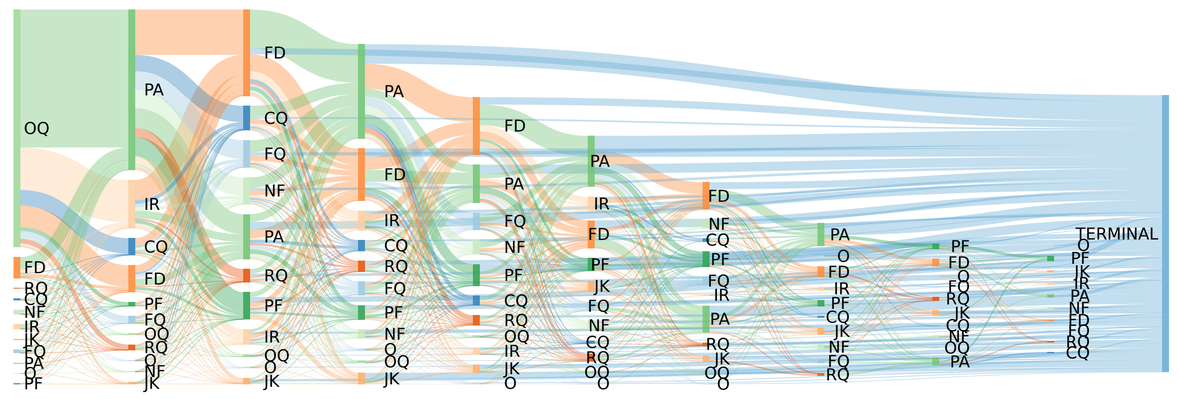

Figure: Flow pattern with a Sankey diagram. Edges are weighted by transition count.

Data Taxonomy

Code	Label	Description	Example
OQ	Original Question	The first question by a user that initiates the QA dialog.	If a computer is purchased with win 10 can it be downgraded to win 7?
RQ	Repeat Question	Posters other than the user repeat a previous question.	I am experiencing the same problem ...
CQ	Clarifying Question	Users or agents ask for clarification to get more details.	Your advice is not detailed enough. I'm not sure what you mean by ...
FD	Further Details	Users or agents provide more details.	Hi. Sorry for taking so long to reply. The information you need is ...
FQ	Follow Up Question	Users ask follow up questions about relevant issues.	Thanks. I really have one simple question – if I ...
IR	Information Request	Agents ask for information of users.	What is the make and model of the computer? Have you tried installing ...
PA	Potential Answer	A potential answer or solution provided by agents.	Hi. To change your PIN in Windows 10, you may follow the steps below: ...
PF	Positive Feedback	Users provide positive feedback for working solutions.	Hi. That was exactly the right fix. All set now. Tx!
NF	Negative Feedback	Users provide negative feedback for useless solutions.	Thank you for your help, but the steps below did not resolve the problem ...
GG	Greetings/Gratitude	Users or agents greet each others or express gratitude.	Thank you all for your responses to my question ...
JK	Junk	There is no useful information in the post.	Emojis. Sigh .... Thread closed by moderator ...
O	Others	Posts that cannot be categorized using other classes.	N/A

Acknowledgement

This work was supported in part by the CIIR and in part by NSF grant #IIS-1419693 and NSF grant #IIS-1715095. Travel was sponsored in part by CIIR and in part by SIGIR Student Travel Grant.

Chen Qu, Liu Yang, W. Bruce Croft, Johanne R. Trippas, Yongfeng Zhang, Minghui Qiu. Analyzing and Characterizing User Intent in Information-seeking Conversations, In Proceedings of SIGIR 2018, Ann Arbor, Michigan, U.S.A. July 8-12, 2018.