

## 1 PARTICIPANT ANNOTATION

We design a series of quite detailed annotation items covering turn level, dialogue level and pagelog level (Table 1 and Table 2). After the dialogue, user and agent are asked to conduct dialogue-related annotations.

Table 1 and Table 2 show the specific annotation items as well as their annotation rules in user-side and agent-side, respectively. We adopt the five-level Likert scale to design the annotation rules for most of our annotation items. For *preference* item in user-side, we employ the weaker version of preference judgment [1] to evaluate the preference between conversational search and traditional Query-SERP search by user. In addition, *key* field is the key where the annotation item is stored in the *json* file named *class*.

## REFERENCES

- [1] Ben Carterette, Paul N Bennett, David Maxwell Chickering, and Susan T Dumais. 2008. Here or there. In *European Conference on Information Retrieval*. Springer, 16–27.

**Table 1: User-side annotation items and their rules**

Key	Level	Description	Annotation form and rules
satisfaction-turn	turn	The degree to which the user feel satisfied with the agent response in the current turn	Five grade levels ranging from 0 to 4 with increasing satisfaction.
understand-turn	turn	The degree to which the user feels that the agent has understood his or her search intent	Five grade levels ranging from 0 to 4 with increasing understand.
topic_user	dialogue	Topic category to which the dialogue belongs	Eight nominal categories, including politics, sports, technology, cuisine, history, health, fashion and others.
intent_user	dialogue	The actual background and content of user’s information need	Text format, no word limit.
satisfaction_user	dialogue	The degree to which the user feels satisfied with agent responses of the entire conversation	Five grade levels ranging from 0 to 4 with increasing satisfaction.
effort_user	dialogue	The degree of effort the user puts into driving the conversation to complete	Five grade levels ranging from 0 to 4 with increasing effort.
preference_user	dialogue	The preference between conversational search and traditional query-SERP search with the information need of this dialogue	Five grade levels ranging from -2 to 2 with increasing preference for conversational search. Grade 0 means neutral.
adoption_user	dialogue	The degree to which you apply the information or suggestions provided by the agent after dialogue	Five grade levels ranging from 0 to 4 with increasing adoption.
improvement_user	dialogue	Improvement suggestions towards agent responses	Text format, no word limit.

**Table 2: Agent-side annotation items and their rules**

Key	Level	Description	Annotation form and rules
clarity-turn	turn	The degree of clarity in expression of user information need in the current turn	Five grade levels ranging from 0 to 4 with increasing clarity.
difficulty-turn	turn	The difficulty of satisfying user’s information need in the current turn	Five grade levels ranging from 0 to 4 with increasing difficulty.
helpfulness-pagelog	pagelog	The degree to which this webpage is helpful in generating agent response	Five grade levels ranging from 0 to 4 with increasing helpfulness.
understand_agent	dialogue	The degree to which the agent understand user’s information need of the whole dialogue from the agent perspective	Five grade levels ranging from 0 to 4 with increasing understand.
satisfaction_agent	dialogue	The degree of agent-perceived user’s satisfaction towards agent responses of the entire conversation	Five grade levels ranging from 0 to 4 with increasing satisfaction.
difficulty_agent	dialogue	The difficulty of meeting user’s information need in the whole dialogue	Five grade levels ranging from 0 to 4 with increasing difficulty.