Modelling Information Needs in Collaborative Search Conversations

Sosuke Shiga
Graduate School of Library, Information and Media Studies, University of Tsukuba
s1521619@u.tsukuba.ac.jp

Hideo Joho*
Research Center for Knowledge Communities, University of Tsukuba
hideo@slis.tsukuba.ac.jp

Roi Blanco
School of Computer Science and Information Technology, RMIT University
rblanco@udc.es

Johanne Trippas
School of Computer Science and Information Technology, RMIT University
johanne.trippas@rmit.edu.au

Mark Sanderson
School of Computer Science and Information Technology, RMIT University
mark.sanderson@rmit.edu.au

ABSTRACT

The increase of voice-based interaction has changed the way people seek information, making search more *conversational*. Development of effective conversational approaches to search requires better understanding of how people express information needs in dialogue. This paper describes the creation and examination of over 32K spoken utterances collected during 34 hours of collaborative search tasks. The contribution of this work is three-fold. First, we propose a model of conversational information needs (CINs) based on a synthesis of relevant theories in Information Seeking and Retrieval. Second, we show several behavioural patterns of CINs based on the proposed model. Third, we identify effective feature groups that may be useful for detecting CINs categories from conversations. This paper concludes with a discussion of how these findings can facilitate advance of conversational search applications.

CCS CONCEPTS

•**Information systems → Information retrieval;** *Query representation; Collaborative search;*

KEYWORDS

Conversation, Information Needs, Collaborative Search

ACM Reference format:
Sosuke Shiga, Hideo Joho, Roi Blanco, Johanne Trippas, and Mark Sanderson. 2017. Modelling Information Needs in Collaborative Search Conversations. In *Proceedings of SIGIR '17, Shinjuku, Tokyo, Japan, August 07-11, 2017,* 10 pages.
DOI: 10.1145/3077136.3080787

*Corresponding author. This work was carried out while the second author was visiting RMIT University.

SIGIR '17, Shinjuku, Tokyo, Japan

DOI: 10.1145/3077136.3080787

1 INTRODUCTION

Verbal conversation used to be an integral part of the searching process when users were required to explain their information needs to an intermediary (e.g. librarian) who operated a catalogue system to find relevant records in a database. When the intermediary was replaced by search engines, textual queries and clicks on ten blue links became the conversation between users and search applications. In the last few years, however, verbal communication between users and search applications has increased due to the advance of automatic speech recognition (ASR) technologies, and the availability of voice-based applications and devices such as Amazon Echo and Google Home.

The increase of voice-based interaction has changed the way people seek information. A major search engine reports that 20% of mobile queries in the US are now submitted using voice (Google [9]). A study also shows that voice queries have different attributes from conventional written queries (Guy [13]). These findings highlight the importance of research on speech oriented interaction in Information Retrieval (IR) and related fields. Kiseleva, et al. [23]'s study also demonstrates that ASR enables mining from spoken interaction data to expand the capability of search applications to offer intelligent assistance.

One promising direction suggested by these studies is that search interaction are becoming more *conversational*. We are no longer limited to short underspecified text strings to predict information needs behind queries. This also means that searchable content could potentially be any spoken word that is recorded: a resource much larger than the Web (Oard [35]). This paper aims to contribute to such a direction by providing insights into the behaviour of information needs expressed in conversations, which we call *conversational information needs* (CINs) in this paper.

This paper sets the following research questions to gain better understanding of how people express a broad range of information needs in conversations of collaborative search tasks.

- RQ1: How can we model information needs expressed in conversations of collaborative search tasks?
- RQ2: What are the characteristics of CINs?
- RQ3: What features are effective at detecting CINs?

The contributions of this paper are as follows. First, we propose a model of CINs based on a synthesis of relevant theories in Information Seeking and Retrieval (ISR). Second, we show several behavioural patterns of CINs using spoken dialogue data we built from approximately 34 hours of recordings of collaborative search tasks. Third, we identify effective feature groups that can be useful for detecting CIN categories from conversations. Finally, this paper discusses how these findings can facilitate the advancement of conversational search applications.

It should be noted that we do not consider the accuracy of ASR or dialogue analyses systems in this work (Jiang, et al. [21]). Instead, we used professionally transcribed data and manually annotated information to exclude the effect of system inaccuracies in our research. It should also be noted that all experiments were carried out in Japanese, and thus, the descriptions and findings in this paper are translations from the original language. We discuss the implication of this limitation in Section 6.

The rest of the paper is structured as follows. Section 2 reviews studies related to this work. Section 3 proposes a model of CINs. Section 4 describes the conversation dataset we built for this study. Section 5 presents the results of analyses that examined the behavioural patterns of CINs and their prediction. Section 6 discusses the major findings, implications on the design of conversational search applications, and limitations of our findings. Section 7 concludes the paper with future directions.

2 RELATED WORK

2.1 Conversations in Information Seeking & Retrieval Research

Research on conversations is not new in ISR. Well known models such as Taylor's taxonomy of information needs [39] and Kuhlthau's Information Searching Process (ISP) model [24] were based on analyses of conversations between librarians and users, or between students in group work. Qualitative analyses of conversations during ISR led researchers to develop conceptual models that help describe complex searching behaviours.

Other studies have focused on discourse aspects of conversations. For example, Belkin, et al. [3] proposed a coding scheme to annotate conversations between librarians and users to better understand the design of expert systems. Their scheme showed that one can extract a range of contextual information from dialogues, including the description, states, modes of problems at hand, user models, search strategies, and search interactions. Later, Belkin, et al. [4] introduced a concept of scripts that described functions of dialogues, and applied them to the design of interactive IR system. Yuan and Belkin [20] also applied a dialogue structure to design an interactive IR system, allowing them to study users' search tactics in detail. A recent study by Trippas, et al. [40] proposed an annotation scheme of conversational search tasks in a context of speech-only interactions.

The increase of interest in collaborative search since mid 2000 led to some quantitative analyses of conversations during search (Morris and Teevan [33] and Hansen, et al. [15]) . For example, Shah and González-Ibáñez [37] classified textual collaborative search chats into the six stages of the ISP model, and visualised the progression of ISP stages during collaborative search tasks. Imazu, et al. [16] also manually classified utterances during collaborative search into search-oriented and task-oriented utterances, suggesting that utterances can be a useful resource to mine contextual information about underlying tasks. Foster [12] performed discourse analysis on talks in group work to examine the relationship between the function of talk and information seeking activities.

These studies suggest that collaborative search conversations can be a promising source to mine contexts of search. However, existing studies remain either conceptual, small-scale, or use text chat data. This paper examines over 32K spoken utterances from 34 collaborative search sessions, which were manually annotated[1].

2.2 Relevant Models of Information Needs

This section reviews theories, models, and findings that can help us investigate information needs expressed in conversations.

Information need has many definitions and conceptualisations in the literature (e.g., [2, 7, 8, 39]). In this paper, we adapt the one defined by Case [7][2]: "An *information need* is a recognition that your knowledge is inadequate to satisfy a goal that you have" (p. 5), since it serves as a good starting point in our discussion. As the definition implies, information needs emerge inside a user's head and are not immediately observable (Wilson [41]). A recent work by Moshfeghi, et al. [34] tackles this issue by analysing brain activities. Analysis of conversations during collaborative work has advantages since many information needs are naturally verbalised during the task, and they can be captured and examined without expensive equipment such as an fMRI.

Taylor's taxonomy of information needs [39] helps us conceptualise differences between perceived needs and queries. Taylor's model consists of four levels of information needs: *Visceral Needs*, *Conscious Needs*, *Formalised Needs*, and *Compromised Needs*. *Visceral Needs* are defined as a "vague sort of dissatisfaction" including nonverbal expressions. *Conscious Needs* are defined as an "ambiguous and rambling statement" of the needs, which may eventually evolve to *Formalised Needs*, which are a "qualified and rational" statement of the need. However, as query log analyses show (Jansen, et al. [19]), users often submit a short and underspecified query to search engines, which can be classified as *Compromised Needs* in Taylor's model. On the other hand, a Formalised Need can be seen as a question submitted to community-based QA sites or speech-oriented search applications.

Brystörm and Järvelin [6] state that the types of information searchers seek can vary over time in complex tasks. They include *problem information* (i.e., requirement of a task at hand), *domain information* (i.e., topical information about the task), and *problem solving information* (i.e., information regarding solutions). Furthermore, Hansen [14] and Ingwersen and Järvelin [17] emphasise the importance of distinguishing task levels such as *work task* (main task) and *search task* (sub tasks) to elicit contextual factors of a given information need. The significance of contextual factors to determine relevance of information objects is also recognised by Saracevic's stratified model of IR interaction [36].

[1] Although we are currently unable to release the dataset due to lack of explicit consent from participants, a project to develop a shareable English dataset is already in progress based on the methodology presented in this paper.

[2] Case [7] is a good place to learn more about various discussions on information needs.

Finally, McGareth's circumplex model of group tasks [11] informs that resolving group members' interest, preference, and opinion is a crucial step in successful collaborative work. This suggests that conversational information needs can be directed towards users themself, in addition to tasks at hand. In Section 3, we discuss how to operationalise these variables to investigate a broad range of information needs expressed in the conversations of a collaborative search task.

2.3 Spoken Dialogue Systems and Dialogue Acts

Finally, we briefly discuss the research on Spoken Dialogue Systems (SDS), which allows for interaction with computer-based applications such as expert systems or databases by spoken natural language (McTear [28]). For example, some SDS were designed to provide travel information or allow users to book trains or flights [27]. However, many of these SDS were based on a finite state-based or frame-based dialogue control which guides the user through a specified dialogue. This suggests that we might need a different approach to conversational search applications, supporting diverse domains, modes, scenarios, and levels of information needs expressed by search engine users.

Nevertheless, studies of broad dialogues offer promising features to better model information needs in conversations. One such example are Dialogue Acts. Dialogue Acts are communicative functions of dialogue segments [5], such as *request*, *inform*, *question*, *suggestion*, and *offer*. The scope and taxonomy of Dialogue Acts are wide and complex, with many different markup schemes available. In this work, we examine a relevant part of the Dialogue Acts' taxonomy defined by ISO 24617-2 [18] since it is the outcome of synthesising major annotation schemes. Another relevant concept, which is rarely examined in IR, is *turn-taking*. Dialogues can be divided into segments called turns. In one turn a single speaker has control over the dialogue and can produce several spoken segments (Khouzaimi, et al. [22]). Analysis of turn-taking structures allows researchers to examine, for example, how smooth the switches between two speakers were. In this work, we incorporate a simplified statistic of turn-taking phenomena as a feature to detect information needs from dialogue.

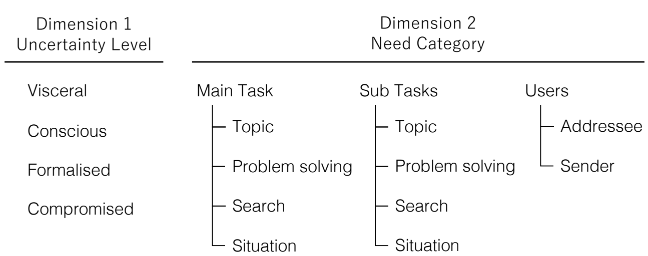


Figure 1: Proposed model of conversational information needs (CINs), consisting of two dimensions: Uncertainty level and need category.

3 PROPOSED MODEL OF CONVERSATIONAL INFORMATION NEEDS (CINS)

Having surveyed the relevant ISR models in literature, we propose to use two dimensions to describe conversational information needs: uncertainty level and need category (See Figure 1). The rest of this section discusses these two dimensions in detail.

3.1 Dimension 1: Uncertainty Level

The first dimension of our proposed model is about the level of uncertainty expressed by collaborative work conversations. Conversations allow us to mine information needs with multiple levels of uncertainty, which is difficult with conventional resources. The conceptual definitions of the uncertainty level are based on the taxonomy proposed by Taylor [39] (see Section 2.2). What we need are operational definitions of these levels. Based on the implications from Kuhlthau's ISP model [24], Belkin, et al.'s ASK [2], and Dialogue Act ISO [18], we devise a set of quantifiable operational definitions to the uncertainty levels as follows.

Regarding *Visceral Needs* (*L*1), we propose to explore negative expressions in conversations such as negation, rejection, disagreement, frustration, and hesitation. Affective states described in the ISP model seem to be appropriate to be considered at this high level of uncertainty in information needs. As for *Conscious Needs* (*L*2), we propose to focus on expressions related to lack of knowledge as informed by ASK and other models. For *Formalised Needs* (*L*3), we propose to take advantage of questions in conversations, since it offered clear and explicit expressions of information needs. This was also informed by the concept of information needs used in the Dialogue Act taxonomy. Finally, for *Compromised Needs* (*L*4), we propose to use actual queries submitted to search applications.

It should be noted that in this work, we focus on Conscious Needs (*L*2) and Formalised Needs (*L*3) since their potential impact on determining people's information needs from conversation is considered to be large. Investigation on Visceral Needs (*L*1) and cross examination with Compromised Needs (*L*4) are equally interesting, but are left to future work.

3.2 Dimension 2: Need Category

While the first dimension of our proposed CIN model was designed to quantify the level of uncertainty expressed in the conversations, the second dimension is designed to shape the types of information needs expressed by users. We devise ten categories stemming from the implications of studies such as Saracevic's relevance model [36], Belkin, et al.'s discourse analyses [3], Brystörm and Järvelin's task complexity [6], and McGrath's model of group tasks [p.14, 11].

The definition and sample utterances of the ten information need categories are shown in Table 1. The top level is divided into three groups: *Main task*, *Sub tasks*, and *Users*. In this work, the main task was travel planning, while sub tasks included finding interesting places to visit at a destination, comparing flight schedules, making a decision on hotels to stay in, and other smaller tasks that needed to be completed to achieve the main task. The main task and sub tasks categories have the common second level of needs on *Topic*, *Problem solving*, *Search*, and *Situation*. The user category has the second level of needs on *Addressee* (partner) and *Sender* (speaker).

The next section describes how we operationalised these variables using conversation data and their annotations.

4 CONVERSATION DATA

This section describes the development and annotation process for our spoken conversation corpus.

4.1 Data Collection

Conversations of collaborative work were collected by a laboratory-based user study[3]. The detail of the study is as follows.

4.1.1 Participants. A total of 34 pairs of participants was recruited for the study at University of Tsukuba. A call for participation was distributed via selected institution mailing lists, lab websites, and subsequent snowballing methods. People were asked to apply as a pair with someone who would be comfortable to perform collaborative work. Suitable applicants were recruited on a first-come first-serve manner. Of the 68 participants, 30 were female and 38 were male. All participants were in the age group of 18 to 24 years old. The academic background of participants varied from Computer Science and Information Science to Engineering, Social Science, and Humanities and Arts.

4.1.2 Collaborative Task. Each pair was asked to perform a travel planning task, one of the most common collaborative tasks performed with search (Morris, [31]). It is a highly information intensive task with many decisions to make based on outcomes of searching. It is also an engaging and familiar task to participants, which was essential to collect natural rich conversations in this work. Furthermore, a travel planning task has successfully been used by collaborative search studies (e.g., Morris and Horvitz [32], Aizenbud-Reshef, et al. [1], Arif, et al. [30]).

In our setting, participant pairs were given 60 minutes to make a travel plan for a 3-4 day trip in Japan with mutual friends. The destination of the trip was decided by participants. Before starting, a budget was set to a reasonable price to ensure realistic decisions were made during the task. We imposed two restrictions in the experiment. First, participants were not allowed to select a pre-defined package tour for the whole schedule. Second, all decisions had to be discussed and agreed between partners. In addition, participants were asked to decide initial roles of operating a shared PC (see below for experimental apparatus), and of writing down a travel plan. They were allowed to switch the roles any time.

4.1.3 Experimental Apparatus. Participants were given a PC, an sheet of A3 paper for writing a travel plan, and pens. Conversations were recorded during the task by a microphone on the table. The PC logged the transaction of operations, however, the focus of this paper is on the conversation between participants. Consequently, close examination of search logs is left to future work.

4.1.4 Protocol. Each experiment was carried out as follows. 1) Participants were given an information sheet to explain the aim and design of the experiment. 2) After a consent form was signed, an entry questionnaire was administered to gather demographic and background information about collaborative work and travel planning. 3) Next, the task of travel planning was introduced with a sample travel plan, and restrictions (see Section 4.1.2) were explained. 4) After a question answering session regarding the task design, participants started to perform the task and conversations and PC operations were recorded. 5) After the task was completed (or 60 minutes were gone), participants were asked to fill in an exit questionnaire to capture general feedback on the task and experiment. An experimenter was in the lab to address any technical issues.

[3]Ethics approval was obtained from University of Tsukuba.

4.2 Segmentation and Annotations

4.2.1 Segmentation. Approximately 34 hours of recordings of collaborative task conversations were first transcribed by a professional service. Transcription included a speaker ID, timestamp, and filler annotations. Then, transcribed conversations were manually segmented into utterances based on guidelines provided by Bunt, et al. [5]. In particular, we adapted the concept of *functional segments*, which is defined as "functionally relevant minimal stretches of communicative behaviour" [p.2549, 5], since they gave us a fine granularity to investigate the characteristics of information needs expressed in conversations. After the initial segmentation by the first author, a random sample of segmentations was examined by another author to ensure consistency in the process. As a result, 32,950 utterances were identified in the dataset, which formed the basis of our analyses.

4.2.2 Annotations. We used a crowdsourcing service[4] to annotate the segmented utterances with ten categories of conversational information needs as described in Section 3. The procedure was as follows. Each utterance was presented to a crowd worker who was also shown the previous and subsequent utterance as context. The annotation interface also provided, as instructions, the information shown in Table 1: the label, definition, and sample of ten information need categories. Crowd workers were then asked to select one of the ten need labels or choose the option "None of them".

We obtained three independent annotations for every dialogue in our dataset, and took a majority vote to assign the final label. When three votes disagreed on a category, we assigned them to "Others", along with those judged as "None of them". All crowd workers had to pass a labelling screening test with a score of > 80%, before they annotated our dataset. We also manually validated the credibility of annotations and removed those that were given by poor quality workers (approx. 20%).

Dialogue act categories were also labelled in the corpus using Dialogue Act definitions provided by ISO [18]. The ISO definition was written in a formal manner: e.g. the *question* act is defined as "A dialogue act performed by Sender, *S*, in order to obtain the information, described by the semantic content, which *S* assumes the addressee, *A*, posses; *S* puts pressure on *A* to provide this information". We rephrased the definitions slightly to ensure that crowd workers understood the scope of each act category. Since we focused on a subset of dialogue acts, only one label was selected per utterance based on a majority voting manner. The resulting distribution of dialogue acts is shown in Table 2. As can be seen, nearly half the dialogues in our dataset were categorised as the *Inform* act (45.0%), followed by *Question* (13.4%) and *Suggestion* (13.2%).

[4]https://lancers.jp/ Accessed: 24/01/2017.

Table 1: Definition and sample utterances of the proposed information need category. $L2$ and $L3$ indicates *Conscious Needs* (expressing lack of knowledge) and *Formalised Needs* (expressing explicit questions), respectively.

Label	Definition	Sample utterance(s)
Main Task		
Topic	Needs of topical knowledge of a main task	*How much time do we have for our trip?* ($L3$)
Problem Solving	Needs of knowledge about solving a problem of a main task	*I don't know how I should describe this in the plan sheet.* ($L2$)
Search	Needs of knowledge about search of a main task	*I can't bookmark this page.* ($L2$)
Situation	Needs of knowledge about a current situation of a main task	*How much have we spent so far?* ($L3$)
Sub Tasks		
Topic	Needs of topical knowledge of sub tasks	*I have no idea about interesting places in [location].* ($L2$)
Problem Solving	Needs of knowledge about solving a problem of sub tasks	*What's the best way to visit [location]?* ($L3$)
Search	Needs of knowledge about search of sub tasks	*Hmm. I can't see access information on this page.* ($L2$)
Situation	Needs of knowledge about a current situation of sub tasks	*Have we checked the price of the entry fee?* ($L3$)
Users		
Addressee	Needs about a partner's opinions, preferences, and knowledge	*Have you been to [location]?* ($L3$)
Sender	Needs about a speaker's opinions, preferences, and knowledge	*Ah, the spelling doesn't come out.* ($L2$)

Table 2: Distribution of dialogue acts (DA).

	N	%
Inform	14,832	45.0
Question	4,412	13.4
Suggestion	4,362	13.2
Offer	548	1.7
Request	471	1.4
Others	8,325	25.3
Total	32,950	100.0

Table 3: Features used to detect CINs.

Category	Features
Temporal	Utterance Sequential ID (Sequential ID)
Dialogue	Dialogue Acts (except Question), Speaker IDs, Turn Ratio
Semantic	Word Embeddings (200 dimensions)
Linguistic	Entities, Filler, Interjection, Mark, Adjective, Post-positions, Auxiliary, Conjunction, Prefix, Verb, Adverb, Noun, Abdominal, Punctuation
Statistic	Characters, Words, Inverse Term Frequency (ITF)

Close examination of temporal patterns of dialogue acts in our dataset shows that the frequency of Dialogue Acts expressed by a pair of participants varies during the collaborative task. Some parts have frequent turns while other parts have a single user dominating the conversations. These observations suggest that mining dialogue acts, speaker IDs, and temporal aspects will be important signals to understand conversational information needs, which will be discussed in the next section in more detail.

Utterances were then categorised into *Concious Needs* or *Formalised Needs* (Uncertainty level dimension in the proposed CIN model) based on a simple rule as follows. If an utterance had one of the ten need category labels and the *Question* dialogue act label, then the utterance was categorised to a Formalised Need, since a question regarding one of the ten needs was explicitly asked. If an utterance had one of the ten need category labels and one of the dialogue act labels except *Question* and *Non Act*, then the utterance was categorised as a Conscious Need, since it expressed a lack of knowledge on an aspect of the task but the expression was not quite as explicit.

In addition to manual annotations, we used the Cloud Natural Language API[5] offered by Google to annotate Part of Speech (POS) and other linguistic tags and entities in utterances.

[5]https://cloud.google.com/natural-language/ Accessed: 14/01/2017.

4.3 Features

Finally, we developed and organised a set of features that were used to train and predict conversational information need categories. We used a Random Forest classifier, which is able to detect interdependencies between covariate features. We were able to estimate the importance of each feature based on its overall contribution to the final predictive model, dependent on every other feature used as a covariate. The features used in our study are summarised in Table 3. To better understand the effectiveness of features, we categorised them into five groups:

Temporal group represents temporal aspects of dialogues such as their sequential IDs;

Dialogue group includes features representing dialogue acts, speaker IDs, and turn ratio. The turn ratio was the number of utterance switches that occurred between two users divided by the sliding window size of N (N was arbitrarily set to 5 in our study). This was designed to quantify the frequency of exchanges of ideas between users;

Semantic group has word embeddings of non-functional words in utterances. We used the embedding vectors[6] (200 dimensions) trained by the entire Japanese Wikipedia corpus (as of 01/11/2016) using `word2vec`[7] [29];

[6]http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ Accessed: 14/01/2017.
[7]https://code.google.com/archive/p/word2vec/ Accessed: 14/01/2017.

Table 4: Distribution of CINs ($L2$ and $L3$).

	N	%
Conscious Needs ($L2$)	2,410	7.3
Formalised Needs ($L3$)	3,072	9.3
Others	27,468	83.3
Total	32,950	100.0

Table 5: Distribution of Conscious Needs ($L2$) and Formalised Needs ($L3$) Categories.

		$L2$		$L3$	
		N	%	N	%
Main Task		179	7.4	233	7.6
	Topic	80	3.3	113	3.7
	Problem Solving	62	2.6	48	1.6
	Search	19	0.8	7	0.2
	Situation	18	0.7	65	2.1
Sub Tasks		1,224	50.8	1,988	64.7
	Topic	723	30.0	1,556	50.7
	Problem Solving	147	6.1	188	6.1
	Search	340	14.1	212	6.9
	Situation	14	0.6	32	1.0
Users		1,007	41.8	851	27.7
	Addressee	797	33.1	790	25.7
	Sender	210	8.7	61	2.0
Total		2,410	100.0	3,072	100.0

Linguistic group consists of POS tags and entity annotations. They are represented as a proportion of their frequency of occurrence in the total number of words in an utterance; and

Statistical group includes length of utterances both in characters and words, and inverse term frequency (ITF), which is calculated as $ITF = \log \frac{n_i}{N_i}$, where i represents an utterance sequence ID, n_i is a frequency of occurrence of term n at i, and N_i is a total number of frequency of occurrence of all terms at i. ITF of an utterance was the average of all terms in the utterance. The score will be high when an utterance has more new or infrequent terms, while the score will be low when an utterance consists of only highly repeated terms.

5 DATA ANALYSIS

We present the results of data analysis performed on the annotated conversation corpus described in Section 4. We begin by looking at some descriptive data of the corpus, followed by temporal analysis, state transition analysis, and finally, feature analysis for prediction.

5.1 Descriptive Analysis

The distribution of CINs based on the two dimensions of the proposed model is shown in Tables 4 and 5, respectively. These tables inform us about several aspects regarding the characteristics of conversational information needs. Table 4 shows that just under 17% of utterances in collaborative work includes expressions of information needs: 7.3% as Conscious Needs ($L2$) and 9.3% as Formalised Needs ($L3$).

The breakdown of need related utterances in Table 5 shows that the largest proportion of such utterances comes from conversations on sub tasks, both in Conscious Needs ($L2$) and Formalised Needs ($L3$). Similarly, many utterances were devoted to express needs of users at both levels. However, Conscious Needs ($L2$) include more dialogues on users while Formalised Needs ($L3$) include more dialogues on sub tasks. Under the main task and sub tasks, utterances on needs related to *Topic* were commonly the largest proportion at both levels.

5.2 Temporal Analysis

Figure 2(a) shows the progression of information needs over the collaborative task. For every participant pair, all utterances were divided into 25 bins, where Bin 1 indicates the beginning of the task and Bin 25 the end. For each bin, the proportion of utterances that were annotated as Conscious Needs ($L2$), Formalised Needs ($L3$), and others was calculated; the figure shows the average of 34 pairs.

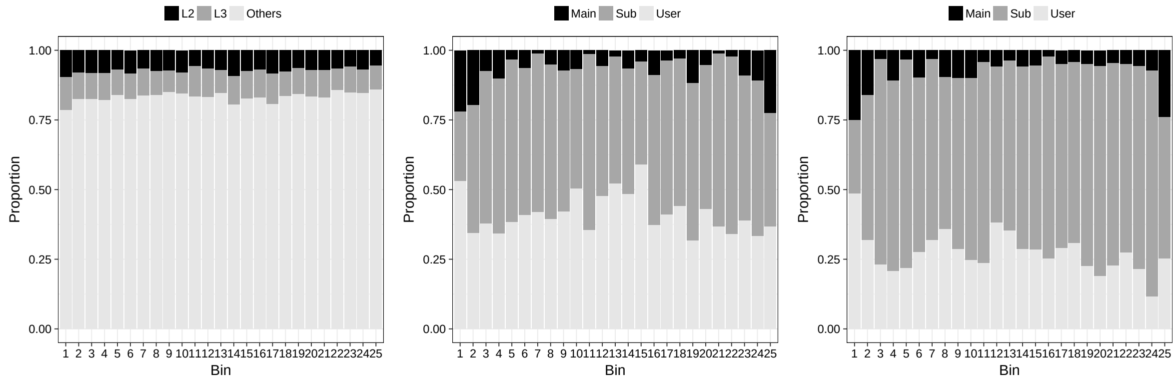
Figure 2(a) shows that the proportion of utterances that express information needs ($L2 + L3$) is approximately 15-20% across the bins. We can observe a slight decrease in need related utterances towards the end of the bins. Pearson's coefficient shows that the proportion of Conscious Needs ($L2$) has a significant large negative correlation ($r = -.54, p \leq .001$) with the Bin IDs, suggesting that expressions of lack of knowledge slowly degrade as the collaborative task progresses. Conversely, the proportion of other dialogues has a significant medium level positive correlation ($r = .45, p \leq .02$) with the Bin IDs.

Figure 2(b) shows the progression of utterances that were annotated as Conscious Needs ($L2$). The figure shows that the proportion of needs related to the main task was larger at the beginning (e.g., confirming task requirements) and end of the task (e.g., validating the travel plan created) than the middle parts. On the other hand, the proportion of needs related to sub tasks and users remains at a similar level although they vary across the bins. No significant correlation was observed between the Conscious Needs categories and Bin IDs.

Figure 2(c) shows the same data as Figure 2(b) but for Formalised Needs ($L3$). The figure shows that Formalised Needs have a similar pattern to Conscious Needs on the main task related utterances. However, the proportion of sub task related information needs is larger than the other two categories of dialogues, which is different from Conscious Needs. Pearson's coefficient shows that the proportion of information needs related to users has a significant medium level of negative correlation ($r = -.46, p \leq .02$) with Bin IDs, suggesting that participants' explicit enquiries on each other slowly degrades towards the end of task.

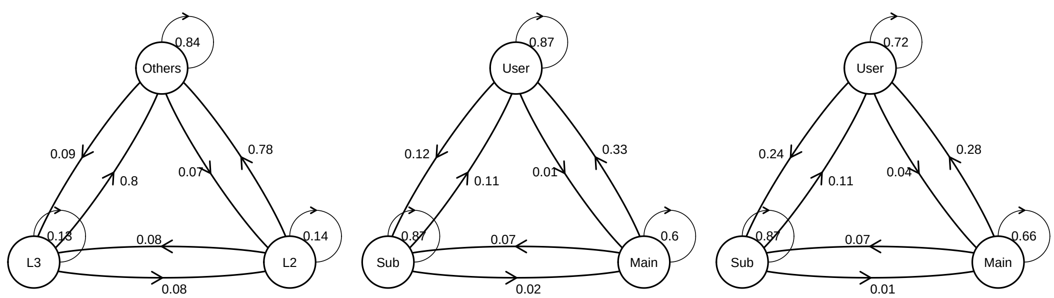
5.3 State Transition Analysis

Figure 3 shows state transfer diagrams of utterances and information needs. Figure 3(a) depicts transitions between Conscious Needs utterances, Formalised Needs utterances, and others. A large proportion of utterances was transferred from need categories to



(a) All Dialogues (b) Conscious Needs ($L2$) (c) Formalised Needs ($L3$)

Figure 2: Progression of information needs over the collaborative task. Figure 2(a) shows that the proportion of utterances expressing information needs is approximately 15-20% across the bins. Figure 2(b) shows that the proportion of Conscious Needs (*L*2) related to the main task is larger at the beginning and end of the task. The proportion of needs related to sub tasks and users varies over the bins but remains similar size. Figure 2(c) shows that the proportion of Formalised Needs (*L*3) related to the main task is larger at the beginning and end of the task. Compared to the Conscious Needs, the proportion of sub task related needs is larger than the other two categories of utterances.



(a) All Dialogues (b) Conscious needs ($L2$) (c) Formalised needs ($L3$)

Figure 3: State transfer diagram of dialogues. Figure 3(a) shows that conversations between users do not necessarily progress from Conscious Needs to Formalised Needs directly, but via other dialogues. Figure 3(b) and 3(c) show that dialogues between users do not necessarily progress from main-task needs to sub task needs directly, but via needs regarding users.

the *Other* category, rather than directly transferring between the two levels of needs.

Figure 3(b) and 3(c) show the transition of the main task related needs, sub task related needs, and user related needs into Conscious Needs (*L*2) and Formalised Needs (*L*3), respectively. Note that the *Others* category were not included in these diagrams. The figures show that there exists a relatively high probability of self-state transition in all need categories, suggesting that utterances on a similar group of needs can continue in conversations. This is common in Conscious Needs (*L*2) and Formalised Needs (*L*3).

Another finding is that we observed little direct transition between *Main Task* and *Sub Tasks*. Instead, a common route of transition was via needs on *Users*. This suggests that it was common to express needs to a participants' partner or to themselves before moving on to a different subtask. This trend is stronger in Conscious Needs (*L*2) than Formalised Needs (*L*3).

5.4 Effective Features to Predict CINs

The final part of the analysis aimed to identify effective features to predict the ten information need categories.

5.4.1 Setup. We tested several algorithms, including Naïve Bayes, SVM, and Random Forest. However, we only report the outcomes of Random Forest, since it provided an overall best performance among them without requiring a large amount of training data. In addition, Mean Decrease Gini (MDG) values were obtained to rank features based on its importance to categorisation, and thus, identify effective features to predict needs related utterances. We used R's *randomForest* package[8] (Liaw and Wiener [25]).

The results of Random Forest and three best features are shown in Table 6, where we assume that the prediction of need categories is performed in a tree structure. In other words, the top level

[8] https://cran.r-project.org/web/packages/randomForest/ Accessed: 14/01/2017.

Table 6: Prediction accuracy of information needs and top 3 effective features determined by Random Forest's Mean Decrease Gini (MDG) values. f1 ... f200 indicate Word Embeddings feature dimensions. N/A indicates cases where the sample size was considered too small to examine ($N \leq 50$). Numbers in parentheses are the actual size of positive samples used in the experiment due to down-sampling requirement. Accuracy is a mean value of 20 iterations of experiments with 95% confidence interval. One sample t-test shows that all results are statistically significance ($p \leq .001$).

		Conscious Needs (*L*2)			Formalised Needs (*L*3)		
		N	Accuracy	Top 3 Features	*N*	Accuracy	Top 3 Features
All		2,410	.73±.01	f151, f17, f97	3,072	.82±.01	Punctuation, f34, f187
Main Task		179	.67±.03	Sequential ID, ITF, f117	233	.72±.02	f3, Sequential ID, f117
	Topic	80	.78±.03	ITF, Sequential ID, f130	113	.81±.03	f130, ITF, Sequential ID
	Problem Solving	62	.72±.05	f70, f50, f124	48	N/A	N/A
	Search	19	N/A	N/A	7	N/A	N/A
	Situation	18	N/A	N/A	65	.78±.04	Sequential ID, ITF, f185
Sub Tasks		1,224 (1,186)	.75±.01	DA, f149, f19	1,988 (1,084)	.74±.01	f52, f149, ITF
	Topic	723 (501)	.73±.02	f111, f191, f13	1,556 (432)	.73±.01	f111, f81, f129
	Problem Solving	147	.60±.02	f81, f86, f28	188	.77±.02	f81, f129, f111
	Search	340	.71±.02	f123, f197, f101	212	.73±.02	f101, f197, f85
	Situation	14	N/A	N/A	32	N/A	N/A
Users		1007	.75±.01	DA, f149, f11	851	.75±.01	f11, f149, f97
	Addressee	797 (210)	.84±.02	Punctuation, DA, f11	790 (61)	.74±.04	f11, f156, f196
	Sender	210	.85±.02	Punctuation, DA, f196	61	.73±.04	f11, f156, f196

classification of *Conscious Needs*, *Formalised Needs*, and *Others* is first performed. Then, within the *Conscious Needs* data, we classify *Main Task*, *Sub Tasks*, and *Users*, individually. Finally, within the *Main Task*, we further classify *Topic*, *Problem Solving*, *Search*, and *Situation* in a similar manner.

In all cases, the prediction was a binary judgement so that we can determine the prediction power of our features for each need category. For a given number of positive utterance samples of a particular need category (e.g., *Topic* of *Main Task*), we sampled the same number of negative samples. Then, a training set and testing set were divided into a 4:1 ratio. The training set was further divided for repeated sampling to determine optimal parameters of Random Forests. In the case where a sufficient number of either positive or negative samples was not available, we performed down-sampling. The numbers in parentheses in Table 6 are the actual size of positive samples used in the experiment. We also removed those categories with fewer than 50 positive samples from the analysis, since the sample size was considered to be too small to examine.

Since the division of training and testing sets involves random sampling, we repeated all experiments 20 times. Therefore, the accuracy in the tables were a mean of the 20 iterations, and the top 3 features were those most frequently selected features in 20 iterations. We ran one sample t-test on the 20 iterations, and all results were found to be statistically significant at $p \leq .001$.

5.4.2 Results. First, the classifier was able to indentify most categories with over 70% accuracy, suggesting that Random Forests were able to model the characteristics of needs related dialogues using our feature sets. Second, some need categories were easier than others to predict. For example, *Addressee* and *Sender* of *Users* in Conscious Needs (*L*2), and *Topic* of *Main Task* in Formalised Needs (*L*3) achieved over 80% of accuracy.

We gained insights into effective features too. For example, word embeddings (denoted as f1 ... f200 in the tables) were generally found to be useful for prediction, suggesting that the semantic features are effective at identifying a range of need categories in conversations. Formalised Needs (*L*3) seem to benefit from the semantic features more than Conscious Needs (*L*2). Dialogue features such as dialogue acts was also found to be useful for many categories of needs. In particular, user related utterances in Conscious Needs benefited from dialogue act information, along with linguistic features such as punctuation. Temporal features such as Dialogue sequence ID and statical features such as ITF were also found in several categories to be useful. For example, *Topic* and *Problem Solving* of *Main Task* (*L*2), and *Situation* of *Sub Tasks* (*L*3) were well categorised by these feature groups.

6 DISCUSSION

This work explored ways to model information needs that are expressed in collaborative search conversations. A spoken dialogue corpus was created and annotated to facilitate the study. Compared to past work, the corpus had several novel aspects. First, it is a transcription of spoken dialogues rather than text messages, which were the focus of past work [37]. Second, the annotations of information needs and a subset of dialogue acts were manually added to all utterances in the corpus. Third, the scale of the corpus was larger than most (if not all) dialogue-based studies in ISR. This section discusses the insight obtained from the study.

6.1 Major Findings

We had three research questions to address in this paper. The following discusses our findings on the three research questions, respectively.

6.1.1 RQ1: Modelling Conversational Information Needs (CINs). In the modelling of CINs, we proposed to use a two-dimensional structure which consisted of uncertainty levels and need categories, based on a synthesis of information seeking behaviour models. Although our operationalisation was limited to Conscious Needs and Formalised Needs, the proposed model was found to be useful for shaping behaviour patterns of CINs, and to determine effective features to predict CINs.

6.1.2 RQ2: Behavioural patterns of CINs. Analyses of 32K annotated utterances based on the proposed CIN model allowed us to obtain the following insights into the behaviour of conversational information needs in our dataset.

First, approximately 17% of utterances were found to express Conscious Needs (7.3%) or Formalised Needs (9.3%) in a collaborative search task. This means that one can increase the opportunity to gain information regarding users' needs by 80% if we extend our scope of information needs from Formalised Needs, which are explicit questions, to Conscious Needs, which are the expression of a lack of knowledge. Whether the additional 80% of signals can extend the range of user needs or not is yet to be determined. Nevertheless, it would seem that conversations in collaborative search can be a promising source to mine the information needs of searchers, which might not be expressed by conventional querying behaviour (e.g., Jansen, et al. [19], Guy [13]).

We also obtained some basic findings of such CINs: 1) The proportion of CINs in utterances slowly degrades as the task progresses but not in a drastic manner; 2) A large proportion of CINs consists of needs related to sub tasks and users, searchers discuss sub task related needs more than user related needs at Formalised Needs level; 3) The proportion of CINs on the main task (travel planning in this study) increases at the beginning and end of the task. The findings suggest that properties such as task development, uncertainty level, need category, and distinction of main task and sub task all play an important role in the modelling of CINs.

Second, the transition between utterances and CINs provided the following insights. 1) Direct transition between Conscious Needs and Formalised Needs is rare in conversations, and often takes place via other types of utterances. This suggests that further investigation on non-CIN utterances is still important to fully leverage conversations for mining searchers' needs. 2) Direct transition between the main task and sub tasks is also rare in conversations. Needs regarding users are often expressed to bridge the transition between the main task and sub tasks.

6.1.3 RQ3: Effective Features to Predict CINs. Here, we examined effective features to learn and predict CINs in conversations. We applied a Random Forests classifier to investigate the performance of five groups of features such as temporal, dialogue, semantic, linguistic, and statistical features.

Our results show that the semantic features based on word embeddings, partly due to their large dimensions, were the most highly ranked features to predict a range of information need categories both in Conscious Needs and Formalised Needs. Dialogue features such as Dialogue Acts, statistical features such as inverse term frequency, and temporal features such as dialogue sequence ID were also useful for several categories of information needs. On the other hand, our implementation of turn-taking phenomena was found to be too simplistic to be effective.

6.2 Implications on the Design of Conversational Search Applications

We demonstrated the effectiveness of word embeddings in predicting a range of CIN categories. This is promising since, unlike Formalized Needs, expressions of Conscious Needs (i.e., lack of knowledge) are unlikely to have clear question forms. Exploring more sophisticated representations of semantic features (e.g. domain specific word embeddings [10]) is likely to further improve CIN category prediction. Techniques such as question mining (e.g., Margolis and Ostendorf [26]) should improve the performance of detection in Formalised Needs.

Dialogue Acts were one of the more expensive features to obtain as they required manual annotations to ensure high accuracy. However, Dialogue Acts were often found to be effective features to detect CINs, therefore, advances in automatic detection of such Acts [38] will significantly impact the development of conversational search applications. Punctuations were artefacts of the transcription process in our investigation. However, given that they were effective features to identify some categories of CINs, developing a method to label punctuations on speech data might be worthwhile.

Finally, our study suggests that careful monitoring of task stages and users' progress through those stages is important for the accurate detection of information need categories from conversations. The proportion of categories varies over time and the stage of tasks. Features that capture temporal aspects were also found to be useful for detecting CINs. This echoes the findings from literature on single-person searches (e.g., Byström and Järvelin [6]), but in the context of conversational collaborative search.

6.3 Limitations

We examined one type of main collaborative task. Other tasks, which require different types of information or decision making should be studied to gain a more comprehensive view of CINs. In particular, the effectiveness of word embeddings for other CIN types is yet to be determined. In addition, our participants and conversation data were based on a particular age group in one organisation in one country. Cultural effects on our findings are left to future work.

7 CONCLUSION AND FUTURE WORK

Advances in ASR accuracy are resulting in significant opportunities and challenges in IR to make search more *conversational*. A deeper understanding of how people express a broad range of information needs in conversations can facilitate the development of conversational search applications. This paper examined over 32K spoken utterances collected during approximately 34 hours of a collaborative search task, based on a proposed model of CINs. The model consisted of two dimensions: uncertainty level and need category.

Our analyses elicited some key behavioural patterns of CINs such as the ratio of CINs in conversations, changes of CINs frequency over task development, and frequent transition to user related needs. Analyses with Random Forest classifiers also identified a range of features – such as semantic features, dialogue features, and temporal features – that are useful for detecting utterances that contain CINs in our dataset. The implications of our findings on the design of conversational search applications include the elicitation of effective features and opportunities for the development of new technologies to determine stages and progress of complex collaborative search tasks.

As for future work, close examination of query log data will allow us to better understand the impact of spoken dialogues on information searching behaviour. One might want to predict or formulate queries from conversations, or develop click models using conversation data. Expanding the scope of analysis to Visceral Needs (*L*1) using other speech features is also planned future work. More sophisticated techniques such as localised word embeddings [10] and advanced turn-taking phenomena [22] may be a promising direction to improve the detection of CINs. Finally, one could investigate the impact of languages and cultures on the behaviour of CINs.

ACKNOWLEDGMENTS

Authors thank to anonymous reviewers for their constructive comments to improve this paper. The work was supported in part by Microsoft Research Asia under MSR CORE-12 Project, Australian Research Council Projects (LP130100563 and LP150100252), Real Thing Entertainment Pty Ltd, and finally, SEEK Ltd. Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsors.

REFERENCES

[1] Netta Aizenbud-Reshef, Artem Barger, Ido Guy, Yael Dubinsky, and Shiri Kremer-Davidson. 2012. Bon Voyage: Social Travel Planning in the Enterprise. In *Proceedings of CSCW 2012*. 819–828. DOI:http://dx.doi.org/10.1145/2145204.2145326

[2] N.J. Belkin, R.N. Oddy, and H.M. Brooks. 1982. Ask for Information Retrieval: Part I. Background and Theory. *Journal of Documentation* 38, 2 (1982), 61–71. DOI:http://dx.doi.org/10.1108/eb026722

[3] N. J. Belkin, H. M. Brooks, and P. J. Daniels. 1987. Knowledge Elicitation Using Discourse Analysis. *International Journal of Man-Machine Studies* 27, 2 (1987), 127–144. DOI:http://dx.doi.org/10.1016/S0020-7373(87)80047-0

[4] Nicholas J. Belkin, Colleen Cool, Adelheit Stein, and Ulrich Thiel. 1995. Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems. *Expert Systems with Applications* 9 (1995), 379–395.

[5] Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010. Towards an ISO Standard for Dialogue Act Annotation. In *Proceedings of LREC 2010*.

[6] Katriina Byström and Kalervo Järvelin. 1995. Task complexity affects information seeking and use. *Information Processing and Management* 31, 2 (1995), 191 – 213.

[7] Donald Case. 2012. *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior* (3rd ed.). Emerald Group Publishing. 491 pages.

[8] Brenda Dervin. 1983. Information as a user construct: The relevance of perceived information needs to synthesis and interpretation. In *Knowledge Structure and Use: Implications for Synthesis and Interpretation*. 155–183.

[9] Google Developers. 2016. Google I/O 2016 - Keynote. (2016). https://www.youtube.com/watch?v=862r3XS2YB0 Online. Accessed: 23/01/2017.

[10] Fernando Diaz, Bhaskar Mitra, and Nick Craswell. 2016. Query Expansion with Locally-Trained Word Embeddings. In *Proceedings of ACL 2016*. 367–377. http://aclweb.org/anthology/P/P16/P16-1035.pdf

[11] Donelson R. Forsyth. 2005. *Group Dynamics* (international student ed.). Thomson Wadsworth. 640 pages.

[12] Jonathan Foster. 2009. Understanding interaction in information seeking and use as a discourse: A dialogic approach. *Journal of Documentation* 65, 1 (2009), 83–105. DOI:http://dx.doi.org/10.1108/00220410910926130

[13] Ido Guy. 2016. Searching by Talking: Analysis of Voice Queries on Mobile Web Search. In *Proceedings of SIGIR 2016*. 35–44. DOI:http://dx.doi.org/10.1145/2911451.2911525

[14] Preben Hansen. 1999. User interface design for IR interaction. a task-oriented approach. In *Proceedings of CoLIS 3*. 191–205.

[15] P. Hansen, C. Shah, and C.P. Klas. 2015. *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*. Springer International Publishing.

[16] Marika Imazu, Shin'ichi Nakayama, and Hideo Joho. 2011. Effect of Explicit Roles on Collaborative Search in Travel Planning Task. In *Proceedings of AIRS 2011*. 205–214. DOI:http://dx.doi.org/10.1007/978-3-642-25631-8_19

[17] Peter Ingwersen and Kalervo Järvelin. 2006. *The turn: Integration of information seeking and retrieval in context*. Springer. 448 pages.

[18] ISO 24617-2:2012 2012. *Language resource management – Semantic annotation framework (SemAF) – Part 2: Dialogue acts*. Standard. International Organization for Standardization, Geneva, CH. 104 pages.

[19] Bernard J. Jansen, Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management* 36, 2 (2000), 207–227.

[20] Xiaojun (Jenny) Yuan and Nicholas J. Belkin. 2014. Applying an information-seeking dialogue model in an interactive information retrieval system. *Journal of Documentation* 70, 5 (2014), 829–855. DOI:http://dx.doi.org/10.1108/jd-06-2013-0079

[21] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How do users respond to voice input errors?: lexical and phonetic query reformulation in voice search. In *Proceedings of SIGIR 2013*. 143–152.

[22] Hatim Khouzaimi, Romain Laroche, and Fabrice Lefevre. 2015. Turn-taking phenomena in incremental dialogue systems. In *Proceedings of EMNLP 2015*. 1890–1895. DOI:http://dx.doi.org/10.18653/v1/D15-1216

[23] Julia Kiseleva, Kyle Williams, Ahmed Hassan Awadallah, Aidan C. Crook, Imed Zitouni, and Tasos Anastasakos. 2016. Predicting User Satisfaction with Intelligent Assistants. In *Proceedings of SIGIR 2016*. 45–54. DOI:http://dx.doi.org/10.1145/2911451.2911521

[24] Carol C. Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science* 42, 5 (1991), 361–371.

[25] Andy Liaw and Matthew Wiener. 2002. Classification and Regression by randomForest. *R News* 2, 3 (2002), 18–22. http://CRAN.R-project.org/doc/Rnews/

[26] Anna Margolis and Mari Ostendorf. 2011. Question detection in spoken conversations using textual conversations. In *Proceedings of ACL 2011*. 118–124.

[27] Michael McTear, Zoraida Callejas, and David Griol. 2016. *The conversational interface*. Springer.

[28] Michael F McTear. 2002. Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)* 34, 1 (2002), 90–169.

[29] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of NIPS 2013*. 3111–3119.

[30] Abu Shamim Mohammad Arif, Jia Tina Du, and Ivan Lee. 2015. Understanding tourists' collaborative information retrieval behavior to inform design. *Journal of the Association for Information Science and Technology* 66, 11 (2015), 2285–2303. DOI:http://dx.doi.org/10.1002/asi.23319

[31] Meredith Ringel Morris. 2013. Collaborative Search Revisited. In *Proceedings of CSCW 2013*. 1181–1192. DOI:http://dx.doi.org/10.1145/2441776.2441910

[32] Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: An Interface for Collaborative Web Search. In *Proceedings of UIST 2007*. 3–12. DOI:http://dx.doi.org/10.1145/1294211.1294215

[33] Meredith Ringel Morris and Jaime Teevan. 2009. Collaborative web search: Who, what, where, when, and why. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, 1 (2009), 1–99.

[34] Yashar Moshfeghi, Peter Triantafillou, and Frank E. Pollick. 2016. Understanding Information Need: An fMRI Study. In *Proceedings of SIGIR 2016*. 335–344. DOI:http://dx.doi.org/10.1145/2911451.2911534

[35] Douglas W. Oard. 2008. Unlocking the Potential of the Spoken Word. *Science* 321, 5897 (2008), 1787–1788. DOI:http://dx.doi.org/10.1126/science.1157353

[36] Tefko. Saracevic. 1997. The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the American Society for Information Science*, Vol. 34. 313–327.

[37] Chirag Shah and Roberto González-Ibáñez. 2010. Exploring Information Seeking Processes in Collaborative Search Tasks. In *Proceedings of ASIS&T 2010*. 60:1–60:10.

[38] Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics* 26, 3 (2000), 339–373. DOI:http://dx.doi.org/10.1162/089120100561737

[39] Robert S. Taylor. 1962. The process of asking questions. *American Documentation* 13, 4 (1962), 391–396. DOI:http://dx.doi.org/10.1002/asi.5090130405

[40] Johanne R Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis. In *Proceedings of CHIIR 2017*. ACM, 325–328.

[41] Tom D. Wilson. 1981. On User Studies and Information Needs. *Journal of Documentation* 37, 1 (Jan 1981), 3–15. DOI:http://dx.doi.org/10.1108/eb026702