

A User Study on Clarifying Comparative Questions

Alexander Bondarenko
alexander.bondarenko@informatik.uni-
halle.de
Martin-Luther-Universität
Halle-Wittenberg

Ekaterina Shirshakova
ekaterina.schirschakova@student.uni-
halle.de
Martin-Luther-Universität
Halle-Wittenberg

Matthias Hagen
matthias.hagen@informatik.uni-
halle.de
Martin-Luther-Universität
Halle-Wittenberg

ABSTRACT

Vague or ambiguous queries can make it difficult for a search engine to correctly interpret a user’s underlying information need. A relatively “simple” solution then is result diversification to cover different interpretations, while in more “conversational” search interfaces, the user can be prompted to clarify their original request. We study clarification in the scenario of comparative questions that ask to compare several options. In our experiment that reflects a conversational search interface with a clarification component, 70% of the study participants find clarifications useful to retrieve relevant results for questions with unclear comparison aspects (e.g., “Which is better, Bali or Phuket?”) or without explicit comparison objects and aspects (e.g., “What is the best antibiotic?”).

CCS CONCEPTS

• **Information systems** → **Query intent; Question answering; Web searching and information discovery.**

KEYWORDS

Comparative questions; Clarification; Comparison objects; Comparison aspects; User study; User experience

ACM Reference Format:

Alexander Bondarenko, Ekaterina Shirshakova, and Matthias Hagen. 2022. A User Study on Clarifying Comparative Questions. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*, March 14–18, 2022, Regensburg, Germany. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3498366.3505832>

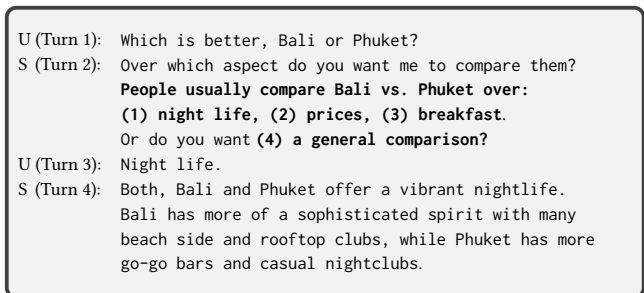
1 INTRODUCTION AND BACKGROUND

Since the first question answering systems were developed [23], brevity and ambiguity of human language are big challenges. To return personalized and more relevant results for vague requests, search engines usually use disambiguation techniques such as result diversification in the sense of including results for different potential intents [21] or query suggestions to let the user select a better query [19]. Though these techniques are rather common in current web search interfaces, their application on mobile devices or in voice search might be hard. To address this issue, recently,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '22, March 14–18, 2022, Regensburg, Germany

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9186-3/22/03...\$15.00
<https://doi.org/10.1145/3498366.3505832>



U (Turn 1): Which is better, Bali or Phuket?
S (Turn 2): Over which aspect do you want me to compare them?
People usually compare Bali vs. Phuket over:
(1) night life, (2) prices, (3) breakfast.
Or do you want (4) a general comparison?
U (Turn 3): Night life.
S (Turn 4): Both, Bali and Phuket offer a vibrant nightlife.
Bali has more of a sophisticated spirit with many beach side and rooftop clubs, while Phuket has more go-go bars and casual nightclubs.

Figure 1: Conceptual design of a search system (S) that interacts with a user (U) by suggesting clarification options.

new ideas have been proposed like query reformulation in a conversational context [11] or clarification [1, 12, 14, 26–29]. Several studies have already shown that in case of conversational search, users appreciate systems that ask for clarification [5, 8–10, 26, 28].

In this paper, we address clarification specifically for comparative questions like “Which is better, Bali or Phuket?” that often represent a need to come to an informed decision about choosing one or another item. Taxonomies for question answering systems contain a respective category since 1990 [17], and recently Yang et al. [25] included questions asking to compare two objects in their HotpotQA question answering dataset. They also indicated that comparative questions represent an interesting but challenging task, since in many cases answering them would require multiple hops. A later study by Bondarenko et al. [4] found that comparative questions constitute at least 3% of the questions a search engine receives—clearly an amount that cannot be neglected. They also found that about 75% of the comparative questions do not specify an aspect on which the comparison should be based (e.g., “Which is better, Bali or Phuket?”) and that about 55% do not clearly state the to-be-compared objects (e.g., superlative questions like “What is the best antibiotic?”) [4]. In a follow-up study, Bondarenko et al. then proposed approaches to classify comparative questions as with or without comparison objects and aspects [3]. For such comparison scenarios where the to-be-compared objects or the comparison aspects initially were not specified, we now study whether clarification requests and suggestions from an interactive search system can improve the user satisfaction.

Figure 1 depicts an example for the underspecified question “Which is better, Bali or Phuket?”. In the clarification request, the system proposes aspects for the comparison or to search without the clarification (option ‘general comparison’ in Figure 1). Users seem to appreciate three clarification suggestions [9, 10], but Zamani et al. [28] found no correlation between Bing users’ engagement

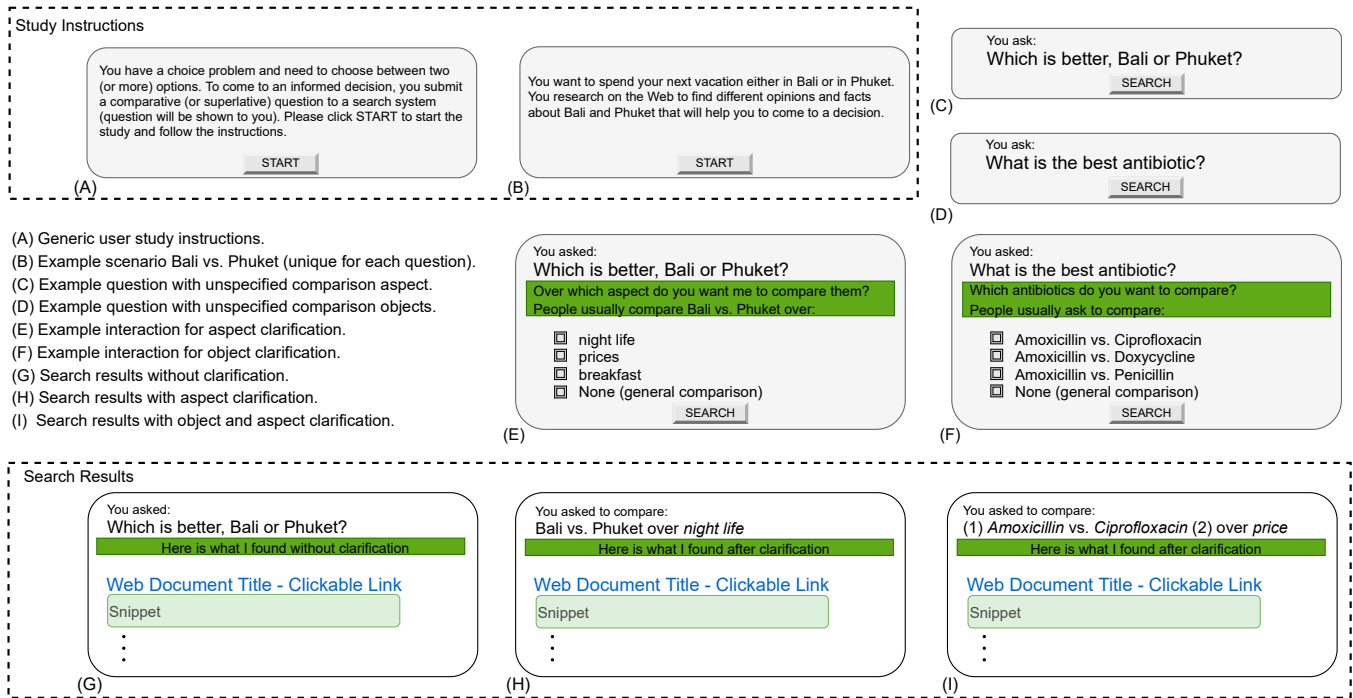


Figure 2: User study interface design.

rates and the number of clarification options. In the example, for the chosen aspect ‘night life’, the system then returns an answer.

An existing search system that helps with comparative information needs is CAM (comparative argumentative machine) by Schildwächter et al. [22]. It accepts two to-be-compared objects and optional comparison aspects in separate input boxes. The search results vary depending on the specified comparison aspects. The system also offers some potential further aspects but does not proactively clarify unspecified aspects—Schildwächter et al. [22] did not study clarification at all. We close this gap by addressing the following research question: *How do clarification interactions improve user satisfaction in comparative search scenarios?*

To answer this question, we conduct two user studies: a searcher interacts with a “conversational” system that actively tries to clarify unspecified comparison objects and aspects (cf. Figure 2). With our focus on the specific use case of comparisons, we complement previous more general clarification studies. These studies, for example, found that search engine users find clarifications useful (functional and emotional benefits) [26], are less dissatisfied when interacting with clarifications [28], and that clarification interactions between users at Stack Exchange usually are helpful to get better answers to the original question [24].

Our results on clarifications in comparative search scenarios are similar. The participants of our study use one of the three suggested clarification options in at least 70% of the cases. More than 85% of the participants enjoyed their experience with the system and indicated that the clarification options were helpful to find satisfactory answers for at least 75% of their assigned tasks.

2 DATA FOR THE USER STUDY

The realistic comparative scenarios for our user study were selected as follows. Using an ALBERT-based classifier from our previous study on identifying comparative questions [3] (ALBERT [16] fine-tuned on 31,000 questions annotated as comparative or not), we found a total of 64,000 probably comparative questions in the MS MARCO [20] dataset (Bing questions), the Google Natural Questions dataset [15], and in a Stack Exchange archive.¹ Focusing on questions that might need clarification, we then ran a RoBERTa-based classifier from our previous study [3] (RoBERTa [18] fine-tuned on comparative questions manually labeled as with or without comparison aspects or objects) and found 22,500 questions that mention comparison objects but have an unclear aspect (e.g., “Which is better, Bali or Phuket?”) and 20,000 questions without comparison objects and aspects (e.g., “What is the best antibiotic?”). We randomly sampled 15 questions for the aspect clarification and 10 questions for the object and aspect clarification. Each query was manually checked and replaced in case of misclassification until we had found 15 with unclear aspects and 10 without objects and aspects. In the selection, we also manually ensured that the comparative questions cover diverse topics like cars, food, electronics, travel, sports, health, arts, and occupation.

To select object clarification options for the 10 queries with missing comparison objects (e.g., “What is the best occupation?” or “What is the best antibiotic?”), we scraped entities from ‘list of’ Wikipedia articles² (e.g., list of occupations) and searched for Wikidata entries via ‘instance of’ (P31) queries against the Wikidata

¹<https://archive.org/details/stackexchange>

²https://en.wikipedia.org/wiki/List_of_lists_of_lists

Table 1: Results of the user study on clarifying comparative questions without aspects (15 questions, 7 participants).

<i>Search result quality</i>		<i>Aspect clarification</i>		<i>Overall</i>	
I found an answer:	(%)	Clarification helpful:	(%)	Pleasant to use:	(%)
Yes	76	Yes	41	Yes	15
More or less	23	More or less	28	More or less	85
No	1	No	21	No	0
Don't know	0	Don't know	0	Don't know	0
		Clarification not used	10		
	$\alpha=0.42$		$\alpha=0.32$		

query service³ (e.g., instance of antibiotics (Q12187)). From the obtained entities, we selected the pairs with the highest sentence-wise co-occurrence frequencies in the Common Crawl snapshot 2014-15⁴ (e.g., ‘drummer’ and ‘guitarist’ for occupation or ‘amoxicillin’ and ‘ciprofloxacin’ for antibiotics).

As for the clarification options for missing comparison aspects, we manually identified the compared objects and then used the following two strategies. (1) We queried the API of CAM [22]⁵ with the object pair (e.g., ‘Bali’ vs. ‘Phuket’) and collect the returned aspect suggestions (CAM finds them in comparative sentences using patterns like “Object 1 is better than Object 2 for Aspect”). (2) We searched for manually annotated comparison aspects in the existing corpora of comparative sentences [2, 6, 7]. For all aspects found by these two strategies, we manually checked their validity until we had found three options per query.

The search result pages that should be shown to the study participants were created before the actual study since the possible clarification options were also pre-computed as explained above. We manually submitted the original comparative question and versions with included clarification options to Google, stored the HTML files of the search results pages, and extracted the document titles, the URLs, and the snippets to show web search-like results but leaving out the ads displayed by Google, etc.

3 USER STUDY

To address the question whether clarifications improve the user “satisfaction” in comparative search scenarios, we conduct a user study for the cases: (1) comparative questions with unspecified comparison aspects (e.g., “Which is better, Bali or Phuket?”) and (2) questions without explicitly specified to-be-compared objects and without aspects (e.g., “What is the best antibiotic?”). In particular, we study whether clarifications are helpful to find satisfactory answers to comparative questions and whether a search interface with comparison aspect and object clarifications overall is pleasant to use. The user interface for the study (inspired by the studies of Zamani et al. [26, 28]) reflects an “interactive” way of questions and answers that allows the system to express uncertainty about a specific part of the question (e.g., comparison aspects) and to suggest some clarification options (cf. Figure 2).

³<https://query.wikidata.org/>

⁴<http://commoncrawl.org/2014/07/april-2014-crawl-data-available/>

⁵<http://ltdemos.informatik.uni-hamburg.de/cam/api-info>

3.1 Study Participants

For the user study, we recruited seven volunteers: five male and two female between 20 and 39 years old. Two of them had a Bachelor’s degree, three held a Master’s degree, and two had no completed high school degree. For all participants, English is a second language—two participants stated to have an intermediate level of English, one stated upper-intermediate, and four stated to have an advanced level of English. The participants had diverse occupational and educational backgrounds: bioinformatics, computer science, construction works, service industry, and web development. All the participants originated from or lived in Europe and Asia.

3.2 Study Setup

We developed the study interface in Python using the graphical user interface package tkinter.⁶ At the beginning of the study, the participants were notified that the study is voluntarily, that they can refuse participating or continuing at any point without providing a reason, that their names or email addresses are not collected (their identity cannot be determined), and that the collected data is used solely for research purposes. After accepting these conditions, each participant saw the general description of the study scenario (cf. Figure 2 (A)): they will need to assume that they are facing a choice problem and want to make an informed decision based on submitting a comparative question to a search engine and that the actual question will be predefined. When clicking on ‘Start’, the actual study began by showing a description of a random scenario from our set of 15 questions without aspects followed by the 10 questions without objects and aspects (each study participant worked on each question; order randomized in the two question groups). The brief scenario descriptions (cf. Figure 2 (B) for an example) had been manually created. After starting a topic, the respective initial question was displayed: either one with unclear aspects (cf. Figure 2 (C)) or one with unclear objects and aspects (cf. Figure 2 (D)).

3.3 Clarifying Comparison Aspects

After reading the short search scenario description, the participants were shown the respective comparative question (with an unspecified comparison aspect in this part of our study; example in Figure 2 (C)). After clicking on the search button, the participants were shown the results for that underspecified question (cf. Figure 2 (G);

⁶<https://docs.python.org/3/library/tkinter.html>

Table 2: Results of the user study on clarifying comparative questions without objects and aspects (10 questions, 7 participants).

<i>Search result quality</i>		<i>Object clarification</i>		<i>Aspect clarification</i>		<i>Overall</i>	
I found an answer:	(%)	Clarification helpful:	(%)	Clarification helpful:	(%)	Pleasant to use:	(%)
Yes	43	Yes	37	Yes	34	Yes	14
More or less	41	More or less	33	More or less	20	More or less	72
No	3	No	1	No	6	No	14
Don't know	13	Don't know	0	Don't know	0	Don't know	0
		Clarification not used	29	Clarification not used	40		
	$\alpha=0.49$		$\alpha=0.45$		$\alpha=0.27$		

similar to standard web search results pages: ten results with snippets and clickable document titles linking to the original web page) along with a clarification prompt that asks “Over which aspect do you want me to compare them?” suggesting the three predefined aspect clarification options (cf. Figure 2 (E)). The participants could explore the original results and decide whether an aspect clarification could be useful. In case of choosing a clarification option, another result page for the question with the clarified aspect was shown (cf. Figure 2 (H)). Afterwards, the participants were asked to provide their feedback. During answering the survey questions, the result page(s)—before and, if chosen, after clarification—were available for comparison or further inspection. We asked the participants to answer the following questions on each scenario: (1) whether they found a satisfactory answer to their question (‘Yes, I found the answer to my question’, ‘More or less: I found something useful, but might search further’, ‘No, I did not find anything useful at all’, and ‘I don’t know’); (2) how useful / helpful clarification was in case they selected one of the clarification options (‘Yes, I found the answer to my question using clarification’, ‘More or less: Results after clarification gave me some useful additional information’, ‘No, results after clarification did not provide any useful additional information’). Finally, after completion of the 15 questions without aspects, we asked the participants to rate the overall experience using the system (whether the system was pleasant to use with the options ‘yes’, ‘more or less’, or ‘no’). After this exit question, the ten questions without objects and aspects followed (cf. Section 3.4).

Results. The results of our user study on clarifying comparative questions without comparison aspects are shown in Table 1. In 76% of the 105 cases, the participants stated that they were able to find satisfactory answers to their questions; 23% found only partial answers so that they would want to search for more information. The initial vague questions were refined with a suggested comparison aspect in 90% of the cases. For a majority of the cases with a used clarification option, the participants found the clarification helpful to obtain good results. All the participants enjoyed using the system, however, only 15% were entirely satisfied. The actual agreement between the participants’ votes per question is rather low (cf. the Krippendorff’s α [13] values in Table 1) indicating that assessing the clarification results and the overall clarification usefulness is a rather subjective task. Still, the votes on whether a satisfactory answer was found have a slightly higher agreement than the ones on aspect clarification helpfulness.

3.4 Clarifying Comparison Objects and Aspects

To evaluate the usefulness of clarifications for comparative questions that do not explicitly mention the to-be-compared objects and that have no aspects (e.g., “What is the best antibiotic?”), we ran a second part of the user study with the same seven participants. Similar to the first part, the participants started by “submitting” the original query (cf. Figure 2 (D)) but this time the results were complemented by three suggestions for object clarification (e.g., Amoxicillin vs. Ciprofloxacin, Figure 2 (F)). If a participant selected an object clarification option, the system then showed results for the adjusted question and suggested clarification options for the comparison aspect similar to the first part of the study (cf. Figure 2 (E)). If a participant then also had selected a clarification for the aspect, the respectively adjusted final query was submitted to show results that match both clarifications (cf. Figure 2 (I)). For the final assessment, all search result pages that a participant had used were available (without clarification, with object clarification if selected, and with object + aspect clarification if selected). We asked the participants to evaluate whether they found a satisfactory answer, whether the object and aspect clarifications were helpful (if used), and about their overall satisfaction with the system (same answer options as in Section 3.3).

Results. The results in Table 2 show that 71% of the participants decided to use one of the suggested object clarification options. The lower ratio compared to the 90% in the aspect-only clarification of the first part might be explained by the observation that the search results for superlative questions (i.e., “What is/are the best ...?”) often contain a single “best” option or a list of “best options”. Some participants simply found that to be sufficient. Only the participants who had selected an object clarification then received aspect clarification options; 60% of the participants decided to use both, an object and an aspect clarification. About 86% of the participants enjoyed the system in this second part of our study (vs. 100% for the first part) and 84% of the participants stated that they had found a satisfactory answer (vs. 99% in the first part). Again, the agreement of the participants’ votes per query is rather low (cf. the Krippendorff’s α values in Table 2) with slightly higher rates for the satisfaction with the answers and the object clarifications.

3.5 Study Limitations

Even though we have used “real” Google search results and had participants with diverse backgrounds (gender, education, occupation,

etc.), our current small study (7 participants, 25 questions each) should be viewed as a pilot experiment with interesting initial results that justify a larger and deeper exploration. In such a broader follow-up study, the manual intervention to select the object and aspect clarification options could even be replaced by an actual system’s choices. Still, our selection of the most frequent clarification options (as per their frequency in the respective corpora) was meant to reflect a straightforward baseline approach.

4 CONCLUSIONS AND FUTURE WORK

We have conducted a user study on clarifications in the scenario of vague comparative searches (i.e., without comparison objects or aspects). In our study, we have mimicked an interactive interface of a search engine that pro-actively suggests comparison aspects or objects for comparative questions without aspects or objects. Our study results are similar to previous more general clarification studies—in at least 70% of the cases, the participants decided to use clarifications to refine the initial search results. The majority of the participants also enjoyed their experience with the system’s clarification component and found clarifications to be helpful for finding satisfactory answers. Since, the general feedback about the clarification helpfulness was positive, a natural next step for future work is to develop the actual approaches that generate clarifying suggestions for comparison aspects and objects and to then repeat our study with more participants for such a “real” system.

ACKNOWLEDGMENTS

This work has been partially supported by the DFG through the projects “ACQuA” and “ACQuA 2.0” (Answering Comparative Questions with Arguments; grants HA 5851/2-1 and HA 5851/2-2) as part of the priority program “RATIO: Robust Argumentation Machines” (SPP 1999). We are especially grateful to our study participants.

REFERENCES

- [1] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019)*. ACM, 475–484.
- [2] Jatin Arora, Sumit Agrawal, Pawan Goyal, and Sayan Pathak. Extracting Entities of Interest from Comparative Product Reviews. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM 2017)*. ACM, 1975–1978.
- [3] Alexander Bondarenko, Yamen Ajjour, Valentin Dittmar, Niklas Homann, Pavel Braslavski, and Matthias Hagen. Towards Understanding and Answering Comparative Questions. In *Proceedings of the 15th ACM International Conference on Web Search and Data Mining (WSDM 2022)*. ACM.
- [4] Alexander Bondarenko, Pavel Braslavski, Michael Völske, Rami Aly, Maik Fröbe, Alexander Panchenko, Chris Biemann, Benno Stein, and Matthias Hagen. Comparative Web Search Questions. In *Proceedings of the 13th ACM International Conference on Web Search and Data Mining (WSDM 2020)*. ACM, 52–60.
- [5] Pavel Braslavski, Denis Savenkov, Eugene Agichtein, and Alina Dubatovka. What Do You Mean Exactly?: Analyzing Clarification Questions in CQA. In *Proceedings of the 2017 Conference on Human Information Interaction and Retrieval (CHIIR 2017)*. ACM, 345–348.
- [6] Nitin Jindal and Bing Liu. Identifying Comparative Sentences in Text Documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006)*. ACM, 244–251.
- [7] Nitin Jindal and Bing Liu. Mining Comparative Sentences and Relations. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference (AAAI 2006)*. AAAI Press, 1331–1336.
- [8] Makoto P. Kato, Ryan W. White, Jaime Teevan, and Susan T. Dumais. Clarifications and Question Specificity in Synchronous Social Q&A. In *Proceedings of the 2013 ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)*. ACM, 913–918.
- [9] Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. Toward Voice Query Clarification. In *Proceedings of the 41st International ACM Conference on Research and Development in Information Retrieval (SIGIR 2018)*. ACM, 1257–1260.
- [10] Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. Clarifying False Memories in Voice-based Search. In *Proceedings of the 2019 Conference on Human Information Interaction & Retrieval (CHIIR 2019)*. ACM, 331–335.
- [11] Johannes Kiesel, Xiaoni Cai, Roxanne El Baff, Benno Stein, and Matthias Hagen. Toward Conversational Query Reformulation. In *Proceedings of the 2nd International Conference on Design of Experimental Search & Information Retrieval Systems (DESIRE 2021) (CEUR Workshop Proceedings)*. 10 pages.
- [12] Antonios Minas Krasakis, Mohammad Aliannejadi, Nikos Voskarides, and Evangelos Kanoulas. Analysing the Effect of Clarifying Questions on Document Ranking in Conversational Search. In *Proceedings of the 2020 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR 2020)*. ACM, 129–132.
- [13] Klaus Krippendorff. 1980. *Content Analysis: An Introduction to its Methodology*. Sage Publications.
- [14] Vaibhav Kumar and Alan W. Black. ClarQ: A Large-Scale and Diverse Dataset for Clarification Question Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*. ACL, 7296–7301.
- [15] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur P. Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. Natural Questions: A Benchmark for Question Answering Research. *Trans. Assoc. Comput. Linguistics* 7 (2019), 452–466.
- [16] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In *Proceedings of the 8th International Conference on Learning Representations (ICLR 2020)*. OpenReview.net.
- [17] Thomas W Lauer and Eileen Peacock. An Analysis of Comparison Questions in the Context of Auditing. *Discourse Processes* 13, 3 (1990), 349–361.
- [18] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR abs/1907.11692* (2019).
- [19] Hao Ma, Michael R. Lyu, and Irwin King. Diversifying Query Suggestion Results. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010)*. AAAI Press.
- [20] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. MS MARCO: A Human Generated Machine Reading Comprehension Dataset. In *Proceedings of the Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016) (CEUR Workshop Proceedings, Vol. 1773)*. CEUR-WS.org.
- [21] Rodrygo L. T. Santos, Craig MacDonald, and Iadh Ounis. Search Result Diversification. *Found. Trends Inf. Retr.* 9, 1 (2015), 1–90.
- [22] Matthias Schildwächter, Alexander Bondarenko, Julian Zenker, Matthias Hagen, Chris Biemann, and Alexander Panchenko. Answering Comparative Questions: Better than Ten-Blue-Links?. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR 2019)*. ACM, 361–365.
- [23] R. F. Simmons. Natural Language Question-Answering Systems. *Commun. ACM* 13, 1 (1970), 15–30.
- [24] Leila Tavakoli, Hamed Zamani, Falk Scholer, William Bruce Croft, and Mark Sanderson. Analyzing Clarification in Asynchronous Information-Seeking Conversations. *Journal of the Association for Information Science and Technology* (2021).
- [25] Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*. ACL, 2369–2380.
- [26] Hamed Zamani, Susan T. Dumais, Nick Craswell, Paul N. Bennett, and Gord Lueck. Generating Clarifying Questions for Information Retrieval. In *Proceedings of the Web Conference (WWW 2020)*. ACM / IW3C2, 418–428.
- [27] Hamed Zamani, Gord Lueck, Everest Chen, Rodolfo Quispe, Flint Luu, and Nick Craswell. MIMICS: A Large-Scale Data Collection for Search Clarification. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM 2020)*. ACM, 3189–3196.
- [28] Hamed Zamani, Bhaskar Mitra, Everest Chen, Gord Lueck, Fernando Diaz, Paul N. Bennett, Nick Craswell, and Susan T. Dumais. Analyzing and Learning from User Interactions for Search Clarification. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020)*. ACM, 1181–1190.
- [29] Zhiling Zhang and Kenny Q. Zhu. Diverse and Specific Clarification Question Generation with Keywords. In *Proceedings of the Web Conference (WWW 2021)*. ACM / IW3C2, 3501–3511.