Elaborative Subtopic Query Reformulation for Broad and Indirect Queries in Travel Destination Recommendation

Qianfeng Wen[1,†], Yifan Liu[1,†], Joshua Zhang[1], George Saad[1], Anton Korikov[1], Yury Sambale[2] and Scott Sanner[1,*]

[1]*University of Toronto, Toronto, Canada*

[2]*Air Canada, Montreal, Canada*

Abstract

In Query-driven Travel Recommender Systems (RSs), it is crucial to understand the user intent behind challenging natural language (NL) destination queries such as the broadly worded "*youth-friendly activities*" or the indirect description "*a high school graduation trip*". Such queries are challenging due to the wide scope and subtlety of potential user intents that confound the ability of retrieval methods to infer relevant destinations from available textual descriptions such as WikiVoyage. While query reformulation (QR) has proven effective in enhancing retrieval by addressing user intent, existing QR methods tend to focus *only* on expanding the range of potentially matching query subtopics (breadth) or elaborating on the potential meaning of a query (depth), but not both. In this paper, we introduce **E**laborative Subtopic **Q**uery **R**eformulation (**EQR**), a large language model-based QR method that combines *both* breadth and depth by generating potential query subtopics with information-rich elaborations. We also release **TravelDest**, a novel dataset for query-driven travel destination RSs. Experiments on **TravelDest** show that **EQR** achieves significant improvements in recall and precision over existing state-of-the-art QR methods.

Keywords

Recommender Systems, Query Reformulation, Large Language Models, Travel

1. Introduction

Air travel is a rapidly growing industry with revenues exceeding 800 billion USD in 2023 [3]. Providing personalized travel destination recommendations to users based on their natural language (NL) queries offers novel opportunities for the air travel industry to engage potential travelers and better serve their travel needs.

A Query-driven Travel Destination Recommender System (RS) can leverage a retrieval subsystem to *match* NL queries to relevant passages in destination descriptions (e.g., sourced from CC-licensed content such as WikiVoyage[1]) and *score* destinations (e.g., as later defined in Algorithm 1). However, such a retrieval approach is prone to fail in light of challenging NL queries that often occur in the Travel RS setting as evident in the following two query types:

1. **Broad Queries**: These are queries that specify a broad category and imply multiple potentially relevant subtopics (e.g., "*cities for youth-friendly activities*").
2. **Indirect Queries**: These are queries that do not directly reflect the user intent but require several further reasoning steps (e.g., "*cities for a high school graduation trip*").

To address such challenging NL queries, existing retrieval research has explored Query Reformulation (QR) to improve intent understanding [4, 5] with recent methods additionally leveraging Large Language Models (LLMs) [6, 7]. LLM QR methods focus on either (a) expanding queries w.r.t. diverse keywords using LLMs, which can be interpreted as a focus on adding subtopic *breadth* to the original query [1, 8] or (b) generating paraphrases or relevant answer passages, which can be interpreted as a focus on adding conceptual *depth* to the description [9, 1, 2, 10, 7].

We conjecture that *both* breadth and depth are key for connecting a user's NL queries to relevant recommendations in travel destination RSs. Further, we observe that LLMs' general reasoning abilities [11, 12] allow them to expand queries with *both* diverse (i.e., breadth) and nuanced (i.e. depth) content that facilitates more effective sparse and dense retrieval with broad and indirect queries.

In summary, we make the following contributions:

1. We propose a novel LLM-based **E**laborative Subtopic **Q**uery **R**eformulation (**EQR**) method [2] [3] that infers multiple subtopic intents covering an original query (i.e., breadth) while providing information-rich elaborations of each subtopic (i.e., depth). **EQR** leverages the abilities of LLMs to understand and reason about broad and indirect user intent on the query side that facilitates effective matching via sparse and dense retrieval. A comparison of two existing QR methodologies with **EQR** in Figure 1 provides clear evidence that **EQR** better addresses breadth and depth as we have conjectured.
2. We introduce and release **TravelDest**, a novel manually curated benchmark dataset for query-driven travel destination RSs with 50 *broad* and *indirect* NL queries and complete relevance labels for 774 destination cities to support our initial comparative evaluation of Travel RS QR methods.

We conclude with a comparative evaluation of QR methods on **TravelDest** and find that **EQR** outperforms existing methods in terms of recall and precision metrics, thus improving the ability of Travel RSs to address challenging broad and indirect NL queries.

[1]https://www.wikivoyage.org/

[2]code: https://github.com/YifanLiu2/ROEGEN-RecSys-24-EQR.git
[3]video: https://youtu.be/WoU77Nw_Z8o

Figure 1: Examples of broad (*Cities for youth-friendly activities*, right) and indirect (*Cities for a high school graduation trip*, left) queries. The comparisons are made among different QR methods discussed in Section 3.2: Q2E [1] (top), Query2Doc [2] (middle), and EQR (bottom). The top 5 recommendations are shown, with irrelevant results marked in red with an *X*. One can clearly identify that **Q2E** covers subtopic breadth, but lacks depth of description, while **Query2Doc** provides in-depth concrete suggestions without covering general subtopics. **EQR** covers both general subtopic breadth with deep subtopic elaboration.

2. Related Work

We review Travel Recommender Systems followed by Query Reformulation that is the methodological focus of our contributions.

2.1. Travel Recommender Systems

Traditional travel RSs have primarily focused on query-free recommendations using collaborative filtering (CF) that leverages the collective travel patterns of users [13, 14], or content-based filtering (CBF) that recommends items similar to a user's past preferences using item features such as location, cost, and reviews [15, 16, 17].

These methods provide coarse insights into user information needs through their past non-textual interactions, such as clicks, purchases, and likes. However, NL queries provide a more nuanced form of user interaction that encapsulates more explicit and personalized user information needs [18, 19]. This highlights a research need for exploration of query-driven travel RSs as we do in this work.

2.2. Query Reformulation

The literature on LLM-based QR methods can be broadly taxonomized into *keyword*, *paraphrase*, and *relevant answer passage* methods. While such QR techniques have been studied for decades — with well-known examples being latent semantic analysis (LSA) [20, 21] and pseudo relevance feedback (PRF) [22, 23, 24] — modern LLMs introduce the ability to use internalized NL knowledge for query reformulation in highly versatile ways as we discuss below.

Keyword-based methods focus on expanding the original query to include broader coverage of subtopics or relevant terms [1, 8, 25]. In a recent line of work, Q2E[1] experimented with different setups for keyword-level LLM-

based expansion, including zero-shot vs. few-shot prompting, chain-of-thought prompting, and the incorporation of PRF. ProQE [26] utilized LLMs in a PRF setting to assess the relevance of retrieved pseudo-documents and extract keywords from them.

Paraphrase-based methods [10] and answer passage-based methods [1, 2, 9] focus on enhancing the depth of the original query by either rephrasing the query to better reflect user intent (e.g., GenQR [7]) or enriching it with information-rich relevant answers (e.g., GQR [10]). Additionally, several studies propose generating relevant answer passages directly using LLMs and demonstrate effectiveness in both sparse and dense retrieval, e.g., Query2Doc [2] in few-shot settings by concatenating the original query with the answer passage and Q2D [1] using various prompting setups and a similar concatenation approach.

Current QR methods appear to focus exclusively on either expanding the breadth or depth of queries, but not simultaneously, which we previously conjectured in Section 1 is a limitation of QR methods for broad and indirect queries. Further, while non-LLM-based QR methods have been explored in RSs [27, 28], there is a research need to investigate the use of LLM-based methods in query-driven RSs, which is important as the information needs and intent behind NL queries in RSs can differ from standard retrieval [15, 29].

3. Methodology

3.1. Query-driven Travel Recommendation

Let q be an NL travel destination query, and let $\mathcal{D}$ be the set of all travel destinations. Each destination $d_i \in \mathcal{D}$ is associated with a document $\mathcal{C}_{d_i} = \{c_1, c_2, \dots, c_m\}$ describing the destination d_i, where each c_j represents a paragraph within

the document.

The objective of the travel RS is to generate a ranked list $\mathcal{S}$ that orders each $d_i \in \mathcal{D}$ based on a scoring function. A straightforward and effective destination scoring algorithm is defined as follows:

Algorithm 1 Destination Scoring Algorithm

1: $q' \leftarrow \text{Reformulate}(q)$ {See Section 3.2}
2: **for** each destination $d_i \in \mathcal{D}$ **do**
3:    $\mathbf{q}' \leftarrow \text{Encode}(q')$
4:    **for** each paragraph $c_j \in \mathcal{C}_{d_i}$ **do**
5:       $\mathbf{c}_j \leftarrow \text{Encode}(c_j)$
6:       $\text{score}(q',c_j) \leftarrow \cos(q',c_j)$ {dense} or $\text{BM25}(q',c_j)$ {sparse}
7:    **end for**
8:    $\mathcal{C}_{q'} \leftarrow$ top-n paragraphs $\{c_1, c_2, \ldots, c_n\}$ using $\text{score}(q', c_j)$
9:    $\text{score}(d_i) \leftarrow \frac{1}{n}\sum_{c_j \in \mathcal{C}_{q'}} \text{score}(q', c_j)$ {Avg of top-n scores}
10: **end for**
11: $\mathcal{S} \leftarrow$ Sort destinations d_i in descending order of $\text{score}(d_i)$

3.2. Query Reformulation

In this work, we fix the structure of the Query-driven Travel Recommender as in Algorithm 1 while experimenting with the impact of different QR methods to implement Line 1, defined as follows:

No QR : $q' = q$, which means no QR is applied.

Q2E [1]: $q' = q + \text{LLM}(q, \text{Q2E-prompt})$, which expands the original query by adding multiple keywords using the LLM. See Figure 2 (top-left) for a detailed prompt.

Query2Doc [2]: $q' = q + \text{LLM}(q, \text{Query2Doc-prompt})$, which generates relevant answer passages from the query using the LLM and concatenates them with the original query. See Figure 2 (top-right) for a detailed prompt.

GenQR [7]: $q' = q + \text{LLM}(q, \text{GenQR-prompt})$, which paraphrases the original query using the LLM. See Figure 2 (bottom-left) for a detailed prompt.

EQR : $q' = q + \text{LLM}(q, \text{EQR-prompt})$, which generates k subtopic elaboration paragraphs from the query using the LLM. See Figure 2 (bottom-right) for a detailed prompt and Section 3.3 for a detailed discussion on **EQR**.

3.3. EQR: Elaborative Subtopic Query Reformulation

The general idea behind our novel contribution of **EQR** as motivated in Section 1 is to infer multiple subtopics from an original query (i.e., breadth) while elaborating each with information-rich content using the LLM's general reasoning abilities (i.e., depth).

Specifically, **EQR** begins with generating k distinct subtopics from a given NL query q, which adds a breadth aspect to capture a wider range of relevant or latent subtopics
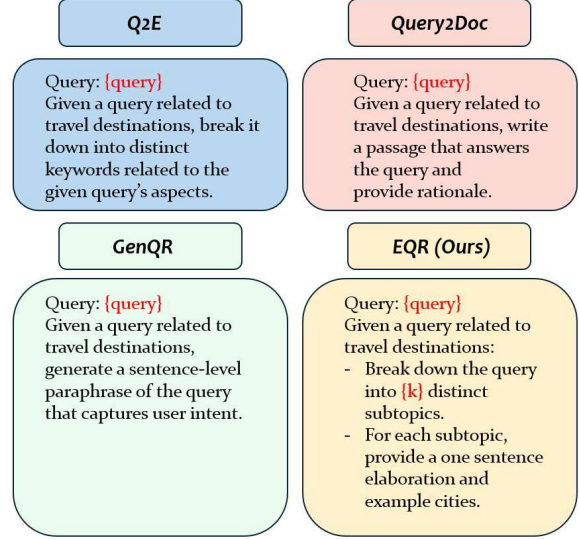
Figure 2: LLM prompts for various QR methods discussed in Section 3.2, with LLM output shown in Figure 1 using two broad and indirect query examples. Q2E [1] (top-left), Query2Doc [2] (top-right), GenQR [10] (bottom-left), and EQR (bottom-right).

compared to answer-based and paraphrase-based methods [9, 2, 1, 10]. Following this, **EQR** produces an information-rich elaboration for each subtopic, denoted $e_1, e_2, \cdots, e_k$, which adds depth via more detailed descriptions and logically entailed connections between the queries and inferred subtopics compared to keyword-based methods [1, 8].

EQR uses the LLM prompt in Figure 2 (bottom-right). The reformulated query q' is a keyword merge $q' = \text{concat}(q, e1, e2, \cdots, e_k)$ for **sparse retrieval via BM25** [30], or a [SEP]-delimited text concatenation $q' = \text{concat}(q, [\text{SEP}], e_1, \cdots, [\text{SEP}], e_k)$ for LLM encoding and **dense retrieval via cosine similarity** (cf. Line 6 in Alg 1).

4. TravelDest: Benchmark Dataset for Travel Destination Recommender Systems

To facilitate our empirical comparison of QR methods for Query-driven Travel RS, we introduce **TravelDest**[4] designed specifically for travel RS with a focus on broad and indirect travel queries. **TravelDest** consists of 50 broad and indirect NL travel queries, spanning various categories, including cultural, adventure, nature, entertainment, and culinary. The dataset contains full per-query labels for 774 target destination cities accessible by major airlines, each with a `WikiVoyage`[1] CC-licensed[5] detailed text description.

The ground truth for each query was established by asking three human labellers to manually assess all 774 target cities per query and assign relevance scores on a scale from 1 to 5. Cities achieving an average score of at least 3 were selected and verified by two additional independent travel experts to ensure relevance.

Table 1

Comparative performance of QR methods using different retrievers on the TravelDest dataset, averaged over 50 broad and indirect queries. An asterisk (*) denotes a statistically significant improvement (paired t-test, $p < 0.01$) compared to the best-performing baseline. Error bars provide 95% confidence intervals on the reported mean metrics.

	QR Methods	MAP@30	MAP@50	Recall@30	Recall@50	R-Precision
Dense - TAS-B	**No QR**	0.513±0.068	0.477±0.064	0.175±0.031	0.267±0.043	0.343±0.044
	GenQR	0.510±0.065	0.471±0.059	0.175±0.031	0.27±0.042	0.351±0.042
	Q2E	0.601±0.057	0.558±0.053	0.209±0.030	0.306±0.036	0.396±0.035
	Query2Doc	0.699±0.058	0.641±0.056	0.209±0.034	0.335±0.043	0.414±0.040
	EQR (Ours)	**0.704**±0.055	**0.661**±0.052 *	**0.260**±0.036 *	**0.366**±0.044 *	**0.455**±0.040 *
Dense - MiniLM	**No QR**	0.528±0.067	0.496±0.062	0.193±0.033	0.287±0.043	0.369±0.044
	GenQR	0.554±0.061	0.517±0.059	0.197±0.033	0.286±0.044	0.369±0.044
	Q2E	0.582±0.055	0.545±0.051	0.206±0.029	0.293±0.039	0.381±0.039
	Query2Doc	0.648±0.053	0.595±0.049	0.211±0.033	0.299±0.038	0.376±0.034
	EQR (Ours)	**0.674**±0.049	**0.636**±0.047 *	**0.248**±0.034 *	**0.351**±0.043 *	**0.437**±0.036 *
Sparse - BM25	**No QR**	0.435±0.069	0.405±0.064	0.146±0.031	0.208±0.038	0.278±0.040
	GenQR	0.431±0.054	0.390±0.050	0.134±0.025	0.204±0.034	0.268±0.037
	Q2E	0.434±0.056	0.402±0.051	0.146±0.023	0.215±0.030	0.281±0.038
	Query2Doc	0.413±0.068	0.387±0.061	0.135±0.029	0.209±0.039	0.276±0.041
	EQR (Ours)	**0.447**±0.061	**0.411**±0.056	**0.147**±0.031	**0.225**±0.041	**0.297**±0.044

5. Experiments

We comparatively evaluate the query-driven RS methodology of Algorithm 1 using each of the QR methods defined in Section 3.2 on the **TravelDest** benchmark dataset. We aim to address the following research questions:

RQ1 : How does **EQR** perform compared to other baseline QR methods?

RQ2 : How do dense and sparse retrieval methods affect the performance of each QR method in query-driven travel RSs?

5.1. Setup

We tested **dense retrieval via cosine similarity** using the TAS-B [31] and MiniLM [32] encoders, which are both popular BERT-based sentence transformer models [33] based on a HuggingFace implementation [34]. We tested **sparse retrieval via BM25** [30] using the Python Rank-BM25 library implementation [35]. For each query, the system retrieves the top-n most relevant paragraphs from each WikiVoyage destination document. We refer to the Appendix for full details and analysis of hyperparameter tuning for n.

We compared **EQR** against various LLM-based QR methods (cf. Section 3.2), with both sparse and dense retrievers as outlined above. Across all QR methods, we utilized GPT-4o [36] as the commonly shared gold standard LLM for query reformulation.

5.2. Metrics

We primarily focus on Recall metrics to ensure the minimal number of relevant items are missed (critical for geographical fairness) and report Recall@30 and Recall@50. Additionally, we report R-Precision to verify system precision relative to the actual ground truth size and MAP@30 and MAP@50 to assess system ability to rank more relevant destinations earlier in the results list.

5.3. Results

Table 1 presents the full comparative results for different QR methods on 50 broad and indirect queries using the **TravelDest** dataset, with various recall and precision evaluation metrics reported.

RQ1: **EQR** outperforms other LLM-based QR methods across all evaluation metrics and retriever types (and with high statistical significance in many cases). **EQR** enhances coverage of retrieved destinations compared to depth-focused answer-passage-based methods, as reflected by its notable improvement at higher k-levels (i.e., @50) over **Query2Doc**, and also appears to provide better intent reasoning compared to breadth-focused expansion-based methods, as indicated by its significant improvement compared to **Q2E**.

RQ2: All QR methods achieved greater scores in dense retrieval than in sparse retrieval, as well as greater improvements in scores when compared to **No QR**. This suggests that for query-driven RSs in the travel domain (with broad and indirect queries), dense retrieval may be more effective than sparse retrieval, likely because keyword matching in sparse retrieval cannot fully capture the nuances embedded in both original and reformulated queries. Interestingly, **EQR** still performs the best among all baselines in sparse retrieval, indicating that the idea of combining breadth and depth in query reformulation applies to both sparse and dense retrieval.

6. Conclusion

We introduced Elaborative Subtopic Query Reformulation (**EQR**), an LLM-based QR method that adds both depth and breadth by generating multiple, information-rich subtopic elaborations to a broad or indirect query. We also introduced the **TravelDest** benchmark dataset to evaluate query-reformation in travel RSs, with **EQR** demonstrating state-of-the-art performance across evaluation metrics and retriever types. Future work includes extending **EQR** to *conversational* RSs (e.g., [37, 38, 39, 40]) and diverse travel data types such as destination reviews (e.g., [41, 42, 43]).

References

[1] R. Jagerman, H. Zhuang, Z. Qin, X. Wang, M. Bendersky, Query expansion by prompting large language models, arXiv preprint arXiv:2305.03653 (2023).

[2] L. Wang, N. Yang, F. Wei, Query2doc: Query expansion with large language models, arXiv preprint arXiv:2303.07678 (2023).

[3] O. Drivers, Airline profitability outlook strengthens (2023).

[4] F. Radlinski, M. Szummer, N. Craswell, Inferring query intent from reformulations and clicks, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, Association for Computing Machinery, New York, NY, USA, 2010, p. 1171–1172. URL: https://doi.org/10.1145/1772690.1772859. doi:10.1145/1772690.1772859.

[5] C. Carpineto, G. Romano, A survey of automatic query expansion in information retrieval, ACM Comput. Surv. 44 (2012). URL: https://doi.org/10.1145/2071389.2071390. doi:10.1145/2071389.2071390.

[6] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.

[7] X. Wang, S. MacAvaney, C. Macdonald, I. Ounis, Generative query reformulation for effective adhoc search, arXiv preprint arXiv:2308.00415 (2023).

[8] K. D. Dhole, E. Agichtein, Genqrensemble: Zero-shot llm ensemble prompting for generative query reformulation, in: European Conference on Information Retrieval, Springer, 2024, pp. 326–335.

[9] L. Gao, X. Ma, J. Lin, J. Callan, Precise zero-shot dense retrieval without relevance labels, arXiv preprint arXiv:2212.10496 (2022).

[10] M. A. K. Ayoub, Z. Su, Q. Li, A case study of enhancing sparse retrieval using llms, in: Companion Proceedings of the ACM on Web Conference 2024, 2024, pp. 1609–1615.

[11] O. Tafjord, B. D. Mishra, P. Clark, Proofwriter: Generating implications, proofs, and abductive statements over natural language, arXiv preprint arXiv:2012.13048 (2020).

[12] S. Yao, D. Yu, J. Zhao, I. Shafran, T. Griffiths, Y. Cao, K. Narasimhan, Tree of thoughts: Deliberate problem solving with large language models, Advances in Neural Information Processing Systems 36 (2024).

[13] C. C. Aggarwal, et al., Recommender systems, volume 1, Springer, 2016.

[14] J. L. Sarkar, A. Majumder, C. R. Panigrahi, S. Roy, B. Pati, Tourism recommendation system: A survey and future research directions, Multimedia tools and applications 82 (2023) 8983–9027.

[15] P. Lops, M. De Gemmis, G. Semeraro, Content-based recommender systems: State of the art and trends, Recommender systems handbook (2011) 73–105.

[16] E.-y. Kang, H. Kim, J. Cho, Personalization method for tourist point of interest (poi) recommendation, in: B. Gabrys, R. J. Howlett, L. C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 392–400.

[17] S. Stabb, H. Werther, F. Ricci, A. Zipf, U. Gretzel, D. Fesenmaier, C. Paris, C. Knoblock, Intelligent systems for tourism, IEEE Intelligent Systems 17 (2002) 53–66. doi:10.1109/MIS.2002.1134362.

[18] Y. Deldjoo, Z. He, J. McAuley, A. Korikov, S. Sanner, A. Ramisa, R. Vidal, M. Sathiamoorthy, A. Kasrizadeh, S. Milano, et al., Recommendation with generative models, arXiv preprint arXiv:2409.15173 (2024).

[19] Y. Deldjoo, Z. He, J. McAuley, A. Korikov, S. Sanner, A. Ramisa, R. Vidal, M. Sathiamoorthy, A. Kasirzadeh, S. Milano, A review of modern recommender systems using generative models (gen-recsys), in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024, pp. 6448–6458.

[20] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman, Indexing by latent semantic analysis, Journal of the American society for information science 41 (1990) 391–407.

[21] S. T. Dumais, G. W. Furnas, T. K. Landauer, S. Deerwester, R. Harshman, Using latent semantic analysis to improve access to textual information, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 1988, pp. 281–285.

[22] J. Rocchio, Relevance feedback information retrieval, The Smart retrieval system-experiments in automatic document processing (1971) 313–323.

[23] S. E. Robertson, On term selection for query expansion, Journal of documentation 46 (1990) 359–364.

[24] G. Amati, C. J. Van Rijsbergen, Probabilistic models of information retrieval based on measuring the divergence from randomness, ACM Trans. Inf. Syst. 20 (2002) 357–389. URL: https://doi.org/10.1145/582415.582416. doi:10.1145/582415.582416.

[25] A. Korikov, G. Saad, E. Baron, M. Khan, M. Shah, S. Sanner, Multi-aspect reviewed-item retrieval via LLM query decomposition and aspect fusion, in: Proceedings of the 1st SIGIR'24 Workshop on Information Retrieval's Role in RAG Systems, 2024.

[26] M. S. Rashid, J. A. Meem, Y. Dong, V. Hristidis, Progressive query expansion for retrieval over cost-constrained data sources, arXiv preprint arXiv:2406.07136 (2024).

[27] S. Hirsch, I. Guy, A. Nus, A. Dagan, O. Kurland, Query reformulation in e-commerce search, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1319–1328. URL: https://doi.org/10.1145/3397271.3401065. doi:10.1145/3397271.3401065.

[28] M. Bhandari, M. Wang, O. Poliannikov, K. Shimizu, Recqr: Using recommendation systems for query reformulation to correct unseen errors in spoken dialog systems, in: Proceedings of the 17th ACM Conference on Recommender Systems, 2023, pp. 1019–1022.

[29] B. Bhattacharya, I. Burhanuddin, A. Sancheti, K. Satya, Intent-aware contextual recommendation system, in: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017, pp. 1–8. doi:10.1109/ICDMW.2017.8.

[30] S. E. Robertson, K. S. Jones, Simple, Proven Approaches to Text Retrieval, Technical Report, University of Cambridge, Computer Laboratory, 1994.

[31] S. Hofstätter, S.-C. Lin, J.-H. Yang, J. Lin, A. Hanbury, Efficiently teaching an effective dense retriever

with balanced topic aware sampling, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 113–122.

[32] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, M. Zhou, Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers, Advances in Neural Information Processing Systems 33 (2020) 5776–5788.

[33] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, arXiv preprint arXiv:1908.10084 (2019).

[34] H. Face, Sentence transformers, https://huggingface.co/sentence-transformers, 2024. Retrieved August 3, 2024.

[35] D. Brown, Rank-BM25: A Collection of BM25 Algorithms in Python, 2020. URL: https://doi.org/10.5281/zenodo.4520057. doi:10.5281/zenodo.4520057.

[36] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., Gpt-4 technical report, arXiv preprint arXiv:2303.08774 (2023).

[37] D. E. Austin, A. Korikov, A. Toroghi, S. Sanner, Bayesian optimization with llm-based acquisition functions for natural language preference elicitation, in: 18th ACM Conference on Recommender Systems (RecSys '24), ACM, 2024, pp. 1–19. doi:10.1145/3640457.3688142.

[38] L. Friedman, S. Ahuja, D. Allen, T. Tan, H. Sidahmed, C. Long, J. Xie, G. Schubiner, A. Patel, H. Lara, et al., Leveraging large language models in conversational recommender systems, arXiv preprint arXiv:2305.07961 (2023).

[39] X. Wang, K. Zhou, J.-R. Wen, W. X. Zhao, Towards unified conversational recommender systems via knowledge-enhanced prompt learning, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1929–1937.

[40] R. Li, S. E. Kahou, H. Schulz, V. Michalski, L. Charlin, C. Pal, Towards deep conversational recommendations, in: Advances in Neural Information Processing Systems 31 (NIPS 2018), 2018.

[41] S. Kemper, J. Cui, K. Dicarlantonio, K. Lin, D. Tang, A. Korikov, S. Sanner, Retrieval-augmented conversational recommendation with prompt-based semi-structured natural language state tracking, in: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024, pp. 2786–2790.

[42] M. M. Abdollah Pour, P. Farinneya, A. Toroghi, A. Korikov, A. Pesaranghader, T. Sajed, M. Bharadwaj, B. Mavrin, S. Sanner, Self-supervised contrastive bert fine-tuning for fusion-based reviewed-item retrieval, in: European Conference on Information Retrieval, Springer, 2023, pp. 3–17.

[43] L. Li, Y. Zhang, L. Chen, Personalized prompt learning for explainable recommendation, ACM Transactions on Information Systems 41 (2023) 1–26.

A. Analysis of Hyperparameters

A.1. Effect of Top-n Paragraphs

In this section, we discuss how the number of top paragraphs aggregated by the retrievers affects the performance of travel RSs in order to determine the optimal n for various retrievers. We examine the scenario where no QR method is applied. According to Figure 3, we observed a performance increase across different evaluation metrics for all types of retrievers w.r.t. the number of top paragraphs, which was then followed by a plateau. We report the optimal values for n and adopt them in Section 5.1:

- **Dense – TAS-B**: $n = 31$
- **Dense – MiniLM**: $n = 18$
- **Sparse – BM25**: $n = 13$

A.2. Effect of Number of Subtopics

In this section, we discuss the influence of the number of subtopics k generated in **EQR** methods by testing k values from 5 to 20. As shown in Figure 4, we did not observe a clear trend between k and the performance of **EQR** across different types of retrievers. We report the performance for $k = 12$ for all types of retrievers.

B. Ablation Studies

To test the robustness of **EQR**, we conducted two ablation studies on the prompts of **EQR** using the TAS-B dense encoder.

B.1. Few-shot Examples

We evaluated the performance of **EQR** with and without few-shot examples. The results did not show a significant difference, with the prompt using few-shot examples performing slightly better on recall-based metrics, while the version without few-shot examples performed better on MAP.

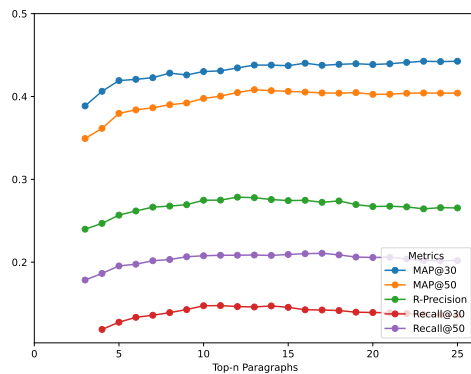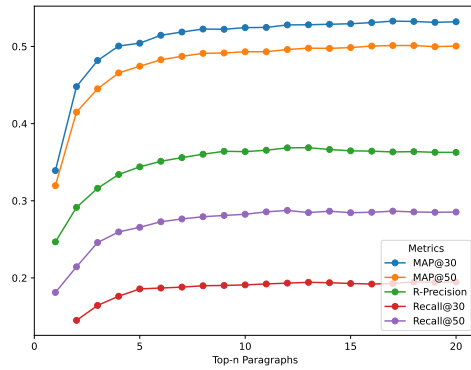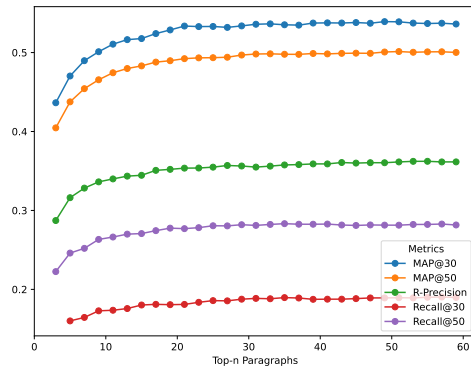|	w/o few-shots	w/ few-shots
MAP@30	0.708	0.704
MAP@50	0.662	0.661
Recall@30	0.255	0.260
Recall@50	0.359	0.366
R-Precision	0.446	0.455

Table 2
Comparative performance of **EQR** w/ and w/o few-shot examples using the TAS-B retriever
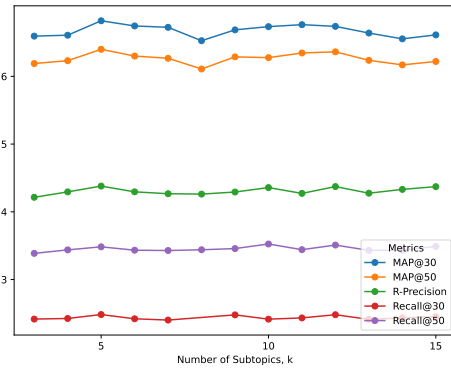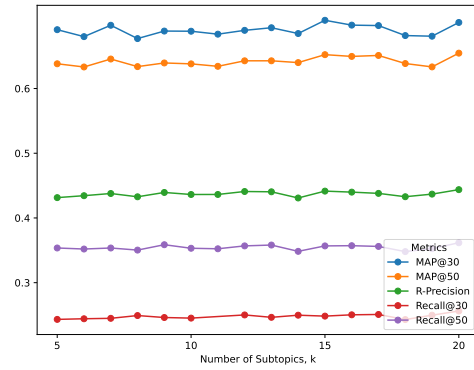
B.2. Target Destinations List

We evaluated the performance of **EQR** with and without the 774 target destination list provided in the prompt to define a valid range of choices for the LLM to select as examples. The results were mixed: the version with the target destination list performed better on MAP and Recall@30, while the version without it performed better on Recall@50.

	w/o city list	w/o city list
MAP@30	0.701	0.704
MAP@50	0.653	0.661
Recall@30	0.256	0.260
Recall@50	0.373	0.366
R-Precision	0.455	0.455

Table 3
Comparative performance of **EQR** w/ and w/o target destination list using the `TAS-B` retriever

Figure 4: Effect of the number of subtopics in EQR across different retrievers: Dense - `TAS-B` (top) and Dense - `MiniLM` (bottom)

C. Examples

In this section, we provide additional examples of various QR methods and results as shown in Tables 4,5,6,7, and 8.

Figure 3: Effect of top-n paragraphs across different retrievers: Dense - `TAS-B` (top), Dense - `MiniLM` (middle), Sparse - BM25 (bottom)

Method	Reformulation / Results
Q2E	extreme sports; hiking trails; rock climbing; water sports; skydiving
Top Results	*Galapagos Islands, Queenstown (New Zealand), Cape Town, Denver, Hong Kong, ...*
GQR	Destinations known for offering a wide range of adventurous experiences
Top Results	*Rio de Janeiro, São Paulo, Washington, D.C., Albuquerque, Charlotte, ...*
Q2D	1. Queenstown, New Zealand: Known as the 'Adventure Capital of the World,' offering everything from bungee jumping and skydiving to skiing and snowboarding 2. Interlaken, Switzerland: This city is a hub for adventure sports, offering activities like paragliding, skydiving, and ice climbing. 3. Banff, Canada: Located in the heart of the Canadian Rockies, Banff is a paradise for outdoor enthusiasts.
Top Results	*Queenstown (New Zealand), New York City, Auckland, Santiago, Dunedin, ...*
EQR (k=3)	Mountain Adventures - Cities that offer hiking, climbing, and skiing opportunities in nearby mountain ranges such as Queenstown (New Zealand) and Aspen. Water Sports - Coastal cities known for exceptional surfing, diving, and sailing activities such as Honolulu and Gold Coast. Jungle Expeditions - Locations that provide guided tours and expeditions into dense jungles and rainforests such as Manaus and Belize City. Desert Safaris - Cities that offer desert experiences, including dune bashing and camel rides such as Dubai and Abu Dhabi. Extreme Sports - Cities that host a range of extreme sports from bungee jumping to paragliding such as Queenstown (New Zealand) and Interlaken.
Top Results	*Queenstown (New Zealand), Puerto Vallarta, Tenerife, Santiago, Gold Coast, ...*

Table 4

Query: Top cities for adventure seekers

Method	Reformulation / Results
Q2E	youth-friendly activities; budget accommodations; group tours; adventure parks; cultural experiences
Top Results	*Aarhus, San Francisco, San Diego, Washington, D.C., Rochester (New York), ...*
GQR	Cities suitable for celebrating a high school graduation with friends.
Top Results	*Shanghai, San Francisco, New York City, Milan, Philadelphia, ...*
Q2D	1. New York City, USA: Iconic landmarks like Times Square and Central Park offer vibrant learning and celebration opportunities. 2. London, England: Rich in history, London offers educational visits to sites like the British Museum and Tower of London. 3. Tokyo, Japan: Tokyo blends tradition and futurism, offering unique experiences in culture, technology, and cuisine.
Top Results	*New York City, London, Manchester, Philadelphia, Rome, ...*
EQR (k=3)	Adventure Activities: Cities that offer exciting outdoor activities and adventures ideal for energetic young adults, such as Queenstown and Interlaken. Cultural Hotspots: Cities rich in cultural experiences, museums, and historical sites, providing educational value, such as Rome and Athens. Beach Destinations: Popular coastal cities with vibrant beach scenes and nightlife suitable for young travelers, such as Miami and Cancun.
Top Results	*San Francisco, Milan, Shanghai, New York City, Athens, ...*

Table 5

Query: Cities for a high school graduation trip

Method	Reformulation / Results
Q2E	wellness centers; yoga retreats; meditation centers; hermal baths; sunshine beaches; nature reserve; peaceful countryside
Top Results	*Palm Springs, Aruba, Cape Town, Gold Coast, ...*
GQR	Destinations known for their tranquil environments, wellness centers, and natural beauty, ideal for a relaxing and restorative getaway.
Top Results	*London, San Francisco, Shanghai, Rochester (New York), San Diego, ...*
Q2D	1. Sedona, Arizona, USA: Famous for its red rock scenery and arts community, Sedona also offers wellness retreats and spas. 2. Bali, Indonesia: Known for its beaches, rice terraces, and tranquil atmosphere, Bali is perfect for relaxation. 3. Kyoto, Japan: Renowned for temples, tea houses, and gardens, Kyoto provides a peaceful escape from daily life.
Top Results	*Santo Domingo, São Paulo, Tenerife, Jakarta, Cabo San Lucas, ...*
EQR (k=3)	Spa Retreats - Cities known for offering luxurious spa services that combine relaxation with a variety of wellness treatments, perfect for rejuvenating the mind and body, such as Budapest and Bali. Nature Escapes - Destinations surrounded by stunning natural landscapes, ideal for outdoor activities like hiking and sightseeing, and offering a peaceful break from the fast pace of life, such as Asheville and Queenstown (New Zealand). Beachfront Relaxation - Cities with serene and picturesque beaches, perfect for enjoying sunbathing, swimming, and rejuvenating by the sea, such as Maldives and Honolulu.
Top Results	*Mombasa, Santo Domingo, Aruba, Puerto Vallarta, Maldives, ...*

Table 6

Query: Cities for a rejuvenating retreat

Method	Reformulation / Results
Q2E	quaint villages; cobblestone streets; local markets; artisan shops; scenic views; historic downtown; peaceful retreats; cultural festivals; bed and breakfasts; picturesque landscapes
Top Results	*Albuquerque, Aurangabad, Aarhus, Ottawa, George Town (Malaysia), ...*
GQR	Quaint towns that exude charm, perfect for a peaceful and immersive getaway.
Top Results	*London, Shanghai, Birmingham, New Orleans, Rochester (New York), ...*
Q2D	1. Bruges, Belgium: Famous for its medieval architecture, cobbled streets, and canals, Bruges offers a charming mix of history, cozy cafes, and art museums. 2. Hallstatt, Austria: Nestled between a lake and mountains, Hallstatt is renowned for its breathtaking landscapes and ancient salt mines. 3. Carmel-by-the-Sea, California, USA: Known for its fairy-tale cottages, art galleries, and beaches, Carmel is perfect for relaxation in a creative setting.
Top Results	*Amsterdam, Lisbon, Brussels, Tallinn, Aarhus, ...*
EQR (k=3)	Historic Charm: Towns that provide a rich sense of history, featuring well-preserved architecture and deep-rooted local traditions, perfect for cultural exploration, such as Bathurst (New Brunswick) and Ljubljana. Natural Beauty: Small towns nestled in breathtaking natural surroundings, offering opportunities for outdoor activities like hiking, photography, and nature walks, such as Aspen and Queenstown (New Zealand). Cultural Festivals: Towns renowned for their distinctive local festivals, giving visitors an authentic insight into regional culture and traditions, such as Edinburgh and Pamplona.
Top Results	*Riga, Aarhus, Albuquerque, Edmonton, Montevideo, ...*

Table 7
Query: Charming small town cities

Method	Reformulation / Results
Q2E	off-the-beaten-path; secluded; quiet towns; remote; less touristy; undiscovered; peaceful; small towns; hidden gems; tranquil
Top Results	*São Paulo, Manchester, Brussels, Ibiza, Nice, ...*
GQR	Cities known for their tranquil atmosphere and less tourist traffic.
Top Results	*London, Shanghai, Paris, San Francisco, Buenos Aires, New York City, ...*
Q2D	1. Ljubljana, Slovenia: Known for its green spaces, pedestrian-friendly streets, and relaxed vibe, ideal for a quieter European experience. 2. Luang Prabang, Laos: A serene town blending French colonial architecture and Buddhist temples, set in a lush, mountainous landscape. 3. Reykjavik, Iceland: Though popular, it offers chances to escape crowds by exploring nearby natural wonders like the Golden Circle and Blue Lagoon.
Top Results	*Reykjavík, Helsinki, Ljubljana, Aarhus, Tallinn, ...*
EQR (k=3)	Remote Locations: Cities that are off the beaten tourist path, providing a sense of solitude and offering distinctive, memorable experiences, such as Iqaluit and Ålesund. Small Town Charm: Smaller cities known for their peaceful streets, intimate atmosphere, and lack of large tourist crowds, making them ideal for a slower-paced getaway, such as Bathurst (New Brunswick) and Lethbridge. Nature Escapes: Cities situated near expansive nature reserves and national parks, where visitors can easily disconnect from urban life and immerse themselves in the tranquility of the outdoors, such as Whitehorse and Aspen.
Top Results	*Brussels, Reykjavík, Ljubljana, Budapest, Venice, ...*

Table 8
Query: Best cities to avoid crowds