Learning to Rewrite Negation Queries in Product Search

Mengtian Guo
UNC at Chapel Hill
mtguo@email.unc.edu

Mutasem Al-Darabsah
Amazon Search
mutasema@amazon.com

Choon Hui Teo
Amazon Search
choonhui@amazon.com

Jonathan May
Amazon Search
jnatmay@amazon.com

Tarun Agarwal
Amazon Search
tagar@amazon.com

Rahul Bhagat
Amazon Search
rbhagat@amazon.com

Abstract

In product search, negation is frequently used to articulate unwanted product features or components. Modern search engines often struggle to comprehend negations, resulting in suboptimal user experiences. While various methods have been proposed to tackle negations in search, none of them took the vocabulary gap between query keywords and product text into consideration. In this work, we introduced a query rewriting approach to enhance the performance of product search engines when dealing with queries with negations. First, we introduced a data generation workflow that leverages large language models (LLMs) to extract query rewrites from product text. Subsequently, we trained a Seq2Seq model to generate query rewrite for unseen queries. Our experiments demonstrated that query rewriting yields a 3.17% precision@30 improvement for queries with negations. The promising results pave the way for further research on enhancing the search performance of queries with negations.

1 Introduction

Online shopping has become increasingly popular in recent years. Retail stores, such as Amazon, eBay, and AliExpress, rely on product search engines to retrieve products that fulfill the user's needs given the query. Providing high-quality results is essential for user satisfaction.

Handling negations has long been recognized as a challenging task in information retrieval (Koopman and Zuccon, 2014; Peikos et al., 2023; Weller et al., 2023). In product search, a search engine that fails to recognize the negation intent can return products that violate the search intent. For instance, the results retrieved by popular retail stores given the query "men sneakers no laces" often contain the undesired product feature of having shoe "laces".

Numerous methods have been proposed to tackle negations in search by either separately indexing the negated content (Limsopatham et al., 2012; Koopman and Zuccon, 2014; Taylor and Harabagiu, 2018) or filtering search results based on the negated content (Merra et al., 2023). Negations in product search pose a unique challenge due to the vocabulary gap between the user's query and the product text fields. For instance, the negation expression "no laces" in the above example indicates a preference for shoes without laces which can be fulfilled by a "slip-on" shoes product that may not even mention term "laces" in its product text. The observation of **vocabulary gap between the negation expression in a query and description of product feature in product text** in product search motivated us to explore the approach of query rewriting for enhancing the search quality on queries with negation. We adopted the generative paradigm, which is to train a Seq2Seq model to generate query rewrites given the original query.

To train the query rewriting model, a dataset containing high-quality query and query rewrite pairs is needed. Considering the limited user behavior data associated with negation queries and the search model's poor understanding of negation intents (Gowriraj et al., 2023; Li et al., 2022; Zhang et al., 2022), we introduced a novel approach that utilizes large language models (LLMs) to extract query rewrites from product text. This approach can extract query rewrites from limited user behavior data and leverages the semantic understanding capability of LLMs. The core idea involves prompting LLM to identify feature descriptions in product text that align with the negation expression in the query. Subsequently, we generate query rewrites by replacing the negation span with the extracted feature description. Through experiments, we demonstrated that query rewriting can lead to remarkable improvements in the search performance of queries with negations.

The main contributions of this work are summarized as follows:

Proceedings of the 31st International Conference on Computational Linguistics: Industry Track, pages 575–582
January 19–24, 2025. ©2025 Association for Computational Linguistics

- We introduced a query rewriting approach to enhance product search engines' performance on queries with negation.

- We proposed an approach to mine high-quality query rewrites from user behavior data based on LLMs.

- The offline quantitative analysis demonstrated the effectiveness of our data generation and query rewriting approach.

2 Related Work

2.1 Handling Negation in Information Retrieval

Handling negations has been recognized as a challenging task in information retrieval, both for non-neural methods such as Indri (Koopman and Zuccon, 2014), BM25 (Peikos et al., 2023) and for neural information retrieval methods such as bi-encoder. Weller et al. (2023) show that most current neural information retrieval methods fail to recognize the negation intent in search queries.

A lot of methods have been proposed to detect negations in text content (Chapman et al., 2001; Mehrabi et al., 2015; Councill et al., 2010; Khandelwal and Sawant, 2020; Merra et al., 2023). However, less attention has been paid to improving the search quality with negation queries. Researchers have explored indexing, filtering, and learning-based approaches. One proposed approach is to distinguish terms within the negative context during indexing, e.g. creating a negated version of terms within the negative context (Limsopatham et al., 2012; Koopman and Zuccon, 2014; Taylor and Harabagiu, 2018). In product search, Merra et al. (2023) proposed to remove products from the search results when the negation content of the query appears in the product text. Wang et al. (2022) proposed to train the semantic retrieval model using negative queries generated by partially negating the original query. An auxiliary loss is added to capture the change in search intent. These methods focus on how to handle negation in the retrieval model, assuming that the negation content and documents use the same vocabulary. In this work, we took a different perspective and addressed the vocabulary gap between queries and documents.

2.2 Query Rewriting

Query rewriting is a fundamental topic in information retrieval. Relevance feedback-based approaches use explicit or implicit user feedback on search results to expand the query with additional terms (Salton and Buckley, 1990). A subset of work uses a two-phase approach (Li et al., 2022; Xiao et al., 2019; Tan et al.). In the first phase, candidate queries are generated based on various signals such as the clicked document, surrounding queries in a session, and collaborative filtering. In the second phase, candidates are ranked using a ranking model based on hand-crafted features (He et al., 2016; Tan et al.), semantic similarity (Li et al., 2022; Xiao et al., 2019), or user profile(Li et al., 2022). Another approach is to train a Seq2Seq model to generate rewrites. People have modified Seq2Seq model training for product search by integrating knowledge graphs (Farzana et al., 2023), query understanding results (Wang et al., 2021), and by modeling search intent (Zhang et al., 2022). In this paper, we adopted the Seq2Seq approach. Since our focus is to demonstrate the efficacy of query rewriting for queries with negation, we implement a generic Seq2Seq model architecture as a proof of concept.

In e-commerce, people have leveraged various data sources to generate query rewriting candidates for model training, including the rewrites generated by users (Wang et al., 2021; Zuo et al., 2022; Farzana et al., 2023) and historical queries from other users. For instance, Zhang et al. (2022) mapped infrequent queries to more popular queries with similar intents. However, relying solely on user-issued queries may fail to close the vocabulary gap between search queries and product text. Moreover, the poor performance of search engines on negation queries often leads to abandoned search sessions without user engagement, which limits the availability of user behavior data for rewrite mining.

Researchers have used generative large language models (LLMs) to rewrite queries in conversational search (Yu et al., 2020; Gowriraj et al., 2023; Mao et al., 2023). These methods utilized LLMs for free-text generation to expand or summarize search context. In this work, we leveraged LLMs to extract content from product text fields such as product title for query rewriting.

3 Methodology

3.1 Overview of Negation Query Rewriting

To bridge the vocabulary gap between the negation content and product text, we intended to rewrite the negation content using the vocabulary adopted by the product text. An intuitive idea is to replace the query's negation span with the corresponding feature description in the product text. In this work, we use the product title as the sole text field of a product as it succinctly covers the product type, brand, and key product attributes or features. The following example illustrates the idea.

Original query: *Screen protector iphone 14 no fingerprints*

Purchased product title: *Mothca Matte Glass Screen Protector for iPhone 14/iPhone13/13 Pro Anti-Glare & Anti-Fingerprint Tempered Glass Clear Film Case Friendly Easy*

Query rewrite: *Screen protector iphone 14* ***Anti-Fingerprint***

In this example, we leveraged the (query, product title) pairs associated with user actions (i.e. purchase). We identified the product feature "Anti-Fingerprint" from the product title that corresponds to "no fingerprints". A query rewrite is generated by replacing "no fingerprints" with "Anti-Fingerprint". This approach does not rely on pre-existing rewrites created by users. It ensures that the generated query rewrites align with the product text vocabulary, bridging the vocabulary gap.

To generalize this idea to unseen user queries, we first built a dataset with query rewrite examples (Section 3.2) and subsequently trained a Seq2Seq model to learn to rewrite user queries (Section 3.3).

3.2 Negation Query Rewriting Dataset Generation

In this section, we first describe the algorithm to detect negation spans in search queries. We then present the approach for identifying the corresponding product feature descriptions.

3.2.1 Query Negation Span Detection

Let q be a user query. We define the negation cue (NC_q) as a set of tokens expressing a negation, and the negation scope (NS_q) as the set of tokens affected by the negation. $N_q = (NC_q, NS_q)$ is the negation span in user query.

Example query: *Screen protector iphone 14* ***no*** *fingerprints*

For instance, in the example query, $NC_q = [\text{"}no\text{"}]$ is the negation cue and $NS_q = [\text{"}fingerprints\text{"}]$ is the negation scope. $N_q =$ "$no\ fingerprints$" is the negation span.

We leveraged the ND4Q model introduced by Merra et al. (2023) for negation span detection. In our implementation, we employed the trained model parameters shared by the authors of the ND4Q paper. This model demonstrates a robust performance in detecting negation spans, achieving an accuracy of 95.38% on the negation query dataset collected by the authors (Merra et al., 2023).

3.2.2 Query Rewrite Generation through LLM

To identify the semantically similar counterparts of negation spans in product text, the applied model must understand the underlying intent conveyed by the negation. Advanced LLMs can provide this understanding with the rich semantic knowledge acquired through pre-training. We leveraged LLMs by prompting the model to identify phrases within product text that mirror the negation span. Specifically, we utilized Flan-T5-XL (Chung et al., 2022) for this purpose. The prompt was structured in the following format:

Prompt Template: *In <product title> which phrase is equivalent to <negation span>?*

Example Prompt: *In 'Mothca Matte Glass Screen Protector for iPhone 14/ iPhone13/ 13 Pro Anti-Glare & Anti-Fingerprint Tempered Glass Clear Film Case Friendly Easy' which phrase is equivalent to 'no fingerprints'?*

Example Answer: *Anti-Fingerprint*

3.2.3 Query Rewrite Generation through Removing the Negation Span

We observed that sometimes LLMs fail to extract a valid rewrite as the product title lacks a feature that is semantically similar to the negation span. This observation suggests that some features corresponding to the negation spans might not require explicit statements. In this case, eliminating the negation span has the potential to reduce the search model's confusion without compromising the essential information required for identifying relevant products. Hence, we generated a query rewrite by removing the negation span from the original search query. This strategy has proven effective in improving search performance by Peikos et al. (2023). An example is illustrated as follows:

Original query: *Screen protector iphone 14 no fingerprints*

Query rewriting: *Screen protector iphone 14*

3.3 Query Rewriting Model Training

3.3.1 Model Architecture

We leveraged the model architecture FELIX, a text-editing approach for text generation proposed by Mallinson et al. (2020). Text-editing models are efficient in low-resource settings and fast at inference time compared to traditional Seq2Seq models.

FELIX decomposes the text-editing task into two sub-tasks: *tagging* to determine the edit operations on the input tokens, and *insertion* to fill in the missing tokens in the output that are absent in the input. The tagging model is a token classification model that assigns one of three labels to each token (`KEEP`, `DELETE`, `INSERT`). When an `INSERT` tag is predicted, k `[MASK]` tags ($k = 5$ in our implementation) are inserted in the intermediate sequence, signaling the insertion model to infill it with a span of a maximum of k tokens. The insertion model is based on a Masked Language Model (MLM). Both the tagging model and the insertion model are based on a 12-layer BERT-base model.

3.3.2 Handling Class Imbalance in Tagging Model Training

During model training, a notable class imbalance issue was observed with the token labels. The ratio between the three labels - `KEEP` : `DELETE` : `INSERT` - is 76:5:1. This imbalance is inherent to the query rewriting task, as the majority of tokens are expected to remain unchanged. We experimented with three strategies to cope with class imbalance: 1) model pretraining; 2) upsampling; and 3) modifying loss functions.

Model Pretraining The tagging model was initialized with a publicly available pretrained BERT-base checkpoint. Before training on the negation query rewriting dataset, we employed a second-stage model pretraining on a noisy query rewriting dataset, aiming to adapt the model to the tagging task. In assembling this pretraining dataset, our goal was to gather an expanded collection of easily obtainable query rewriting examples with reduced class imbalance.

The pretraining dataset was drawn from four distinct sources to gather a greater variety of examples. First, we extracted a random sample of frequent search queries, assuming that no rewriting is needed for this query set. Second, we extracted user-generated query rewrites. Within a search session, if a user issues a query q and, within 60 seconds, follows it with query q' along with a click or purchase action, we record the query rewrite pair (q, q'). Third, for each tail query q, we identified the head query q' with the highest cosine similarity in the embedding space, forming a query rewrite pair (q, q'). Last, we used a query relaxation model to remove tokens with low importance, resulting in query rewrite q'.

Upsampling We applied query-level upsampling. Most queries remain unchanged in the generated dataset. We upsampled the modified queries, i.e. with negation spans rewritten or removed, to match the number of unchanged queries. This increased the occurrence of `DELETE` and `INSERT` labels in the training set.

Loss Function We employed focal loss during model training (Lin et al., 2018). Focal loss was designed to address the issue of class imbalance, particularly in scenarios where one class is significantly more prevalent than the other. The idea behind focal loss is to down-weight the contribution of well-classified examples and give higher importance to examples that are hard to classify or are misclassified.

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} log(p_t) \tag{1}$$

$$p_t = P(y = t|X) \tag{2}$$

Here, p_t is the predicted probability of the target class, α_t and γ are both hyperparameters. We used the default hyperparameters proposed in the original paper ($\alpha_t = 0.25$, $\gamma = 2$).

4 Experiments

4.1 Evaluation Metrics

We evaluated the impact of query rewriting by comparing the search results of the original query and the query rewrite given the same search model. Specifically, we evaluated the impact on precision utilizing a production-grade semantic matching model within a large e-commerce search engine. We chose to focus on precision due to the observed low precision on negation queries, which significantly impacts user experience.

Given a search query q, the semantic matching model retrieves a set of products $D = \{d_1, d_2, \cdots, d_K\}$. For each product, we evaluate its relevance to the search query using a relevance judgment model $r(q, d_i)$. If the product is relevant to the query, $r(q, d_i) = 1$.

$$Precision@K = \frac{\sum_{i=1}^{K} r(q, d_i)}{K} \tag{3}$$

For the set of product $D' = \{d'_1, d'_2, \cdots, d'_K\}$ retrieved by a query rewrite q', product relevance is evaluated concerning the original query, i.e. $r(q, d'_i)$. $Precision@K'$ can then be evaluated. We set $K = 30$ in our experiment. The impact of query rewrite is calculated as:

$$\Delta Precision@K = Precision@K' - Precision@K \tag{4}$$

4.2 Dataset Generation

We extracted approximately 2.3 million negation queries associated with purchases from the anonymous search logs of a large e-commerce site. We focused on negation queries with $Precision@30 < 100\%$ when processed by a semantic matching model. Queries with negation span that are already present in the purchased product title were left unaltered. The rest underwent the query rewrite generation process as described in Section 3.2.2 and Section 3.2.3.

To maintain data quality, we evaluated the generated query rewrites and retained only those that improved Precision@30. For queries where the rewrite led to a precision decrease, we assumed that no rewriting was needed. The entire data generation process is summarized in Figure 1.

As shown in Table 1, the resulting dataset contains 281K query rewrites generated by the LLM, 398K query rewrites generated by removing the negation spans, and 1.6M unaltered queries. The unaltered queries encompass cases with negation spans present in the product title or the rewrites failed to improve precision@30. The unaltered queries are integrated into model training, enabling the model to recognize the cases where query rewriting is not needed. We randomly sampled a validation set containing 50K queries, and the remaining queries constitute the training set.

4.3 Semantic Matching Model

We used a DSSM-style bi-encoder model similar to Nigam et al. (2019) which produces embedding vector by taking the average embeddings of input tokens. In this work, we use a BERT-based encoder (Devlin et al., 2018; Reimers and Gurevych, 2019) that has 2 BERT layers with 4 attention heads each.

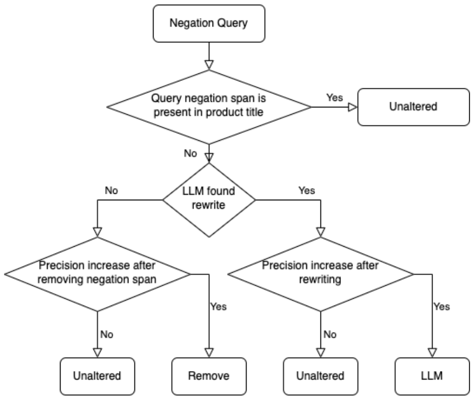

Figure 1: Data generation process

	Training	Validation
Total	2.2M	50K
Case 1 (Unaltered)	1.6M (70%)	35K (70%)
Case 2 (Remove)	389K (17%)	8.7K (17%)
Case 3 (LLM)	275K (12%)	6.3K (13%)

Table 1: Number of instances belonging to different types of rewriting and data partitions in the resulting dataset.

The query and product encoders share the same model weights but accept different token lengths. Following (Nigam et al., 2019), we use only query keywords and product title as model input. The model is trained using the InfoNCE (Van den Oord et al., 2018) loss function equipped with Additive Margin Softmax (Wang et al., 2018).

4.4 Results

Table 2 presents the changes in precision@30 when applying query rewriting on negation queries within the validation set. Notably, all original negation queries in the validation set have $precision@30 < 100\%$. In addition, we reported the impact on queries on which the semantic matching model performs poorly, i.e. those with $precision@30 < 88\%$.

The mined query rewrites yield a 5% improvement in precision@30 for queries in the validation dataset and an 8% improvement for queries with $precision@30 < 88\%$. On the other hand, the query rewrites generated by the FELIX model with pre-training demonstrate a 3.17% improvement in precision@30 for queries in the validation dataset and a 5.94% improvement for queries with $precision@30 < 88\%$. A slight performance decrease is observed without the pre-training phase.

These results indicate that query rewriting effec-

	Dataset	Not Pre-trained	Pre-trained
All queries	+5%	+2.87%	+3.17%
Prec@30 <88%	+8%	+5.45%	+5.94%

Table 2: Changes in precision@30 with query rewriting.

		Predicted rewrite	
		No	Yes
Actual rewrite	No	23K (0)	12K (-3.12%)
	Yes	3K (0)	12K (+15.67%)

Table 3: Impact of query rewriting on queries that require or do not require rewriting. The rows represent whether a query requires rewriting in the gathered dataset, while the columns represent whether the model rewrites the query.

tively addresses the vocabulary gap between negation content and product text, enabling the semantic matching model to better understand the search intent behind negation. The Seq2Seq model can generalize the query rewriting patterns mined from search data to unseen queries, generating query rewrites that more precisely capture the customer's search intent than the original query. Consequently, this model can serve as a preprocessing step, refining user queries before feeding into the search engine or model.

As mentioned previously, a majority of the queries in the dataset remain unaltered. We examined the model's impact on queries that require or do not require rewriting. In Table 3, the rows represent whether a query requires rewriting in the gathered dataset, while the columns represent whether the model rewrites the query. In each cell in the table, the number of queries falling into the category and their change in precision@30 are shown. The analysis is done on the pre-trained model.

We can see that the model can recognize the majority of queries that require rewriting. Focusing on these queries, the model-generated rewrites result in a notable 15.67% improvement in precision@30, which is equivalent to approximately 5 more relevant products within the top 30 results. However, the model also rewrites a substantial portion of queries that should remain unaltered. For these queries, the model's rewrites lead to a decrease in precision (-3.12%). In other words, inappropriate query rewriting may result in more irrelevant products. To address this issue, further investigation is needed to minimize the negative impact on queries that do not require rewriting. In this work, the query rewriting model was trained to keep the original query when rewriting is not needed. An alternative strategy is to apply a binary classification model or leverage search behaviors (e.g. user-initiated query rewrite) to identify queries that require rewriting.

4.5 Deployment Considerations

The query rewriting process inevitably introduces extra latency. In practice, query rewrites can be pre-computed offline for a list of head and torso queries, which are repeatable and cover most of the query coverage. Those query rewrites can be extracted directly using LLMs as described in Section 3.2. Whenever a real-time query is requested, a lookup to query rewrite cache is performed to collect the corresponding query rewrite, and the resulting queries are then passed to the search engine. If the query is not found in the cache, the trained query rewriting model can be leveraged to generate a query rewrite. In this paper, we adopted a BERT-based query rewriting model to demonstrate the effectiveness of the approach. However, in practice, lightweight query rewriting models, which have been proposed and applied to e-commerce platforms (Zhang et al., 2022; Zuo et al., 2022), can be leveraged considering latency requirements. Note that LLMs, which require significant computational resources and time, is only used offline.

5 Conclusion

In this work, we applied query rewriting to enhance product search performance on queries with negation. We introduced a method to extract query rewrites using LLMs, which produces high-quality query rewrites that increases search precision by 5%. This LLM-based data generation method can generalize to other use cases that are subject to vocabulary gap. Our experiments demonstrated that the trained query rewriting model yields a 3.17% precision improvement on queries with negation. The result implies that negation queries in product search are subject to vocabulary gaps, and query rewriting enables the semantic matching model to better understand the user's search intent. This work thus shows that query rewriting can significantly improve the search precision on queries with negation.

References

W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301–310.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling instruction-finetuned language models. *arXiv preprint*.

Isaac Councill, Ryan McDonald, and Leonid Velikovich. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59, Uppsala, Sweden. University of Antwerp.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

Shahla Farzana, Qunzhi Zhou, and Petar Ristoski. 2023. Knowledge Graph-Enhanced Neural Query Rewriting. In *Companion Proceedings of the ACM Web Conference 2023*, pages 911–919, Austin TX USA. ACM.

Srinivas Gowriraj, Soham Dinesh Tiwari, Mitali Potnis, Srijan Bansal, Teruko Mitamura, and Eric Nyberg. 2023. Language-Agnostic Transformers and Assessing ChatGPT-Based Query Rewriting for Multilingual Document-Grounded QA. In *Proceedings of the Third DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering*, pages 101–108, Toronto, Canada. Association for Computational Linguistics.

Yunlong He, Jiliang Tang, Hua Ouyang, Changsung Kang, Dawei Yin, and Yi Chang. 2016. Learning to Rewrite Queries. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1443–1452, Indianapolis Indiana USA. ACM.

Aditya Khandelwal and Suraj Sawant. 2020. NegBERT: A transfer learning approach for negation detection and scope resolution. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5739–5748, Marseille, France. European Language Resources Association.

Bevan Koopman and Guido Zuccon. 2014. Understanding negation and family history to improve clinical information retrieval. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, SIGIR '14, pages 971–974, New York, NY, USA. Association for Computing Machinery.

Sen Li, Fuyu Lv, Taiwei Jin, Guiyang Li, Yukun Zheng, Tao Zhuang, Qingwen Liu, Xiaoyi Zeng, James Kwok, and Qianli Ma. 2022. Query Rewriting in TaoBao Search. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3262–3271, Atlanta GA USA. ACM.

Nut Limsopatham, Craig Macdonald, Richard McCreadie, and Iadh Ounis. 2012. Exploiting term dependence while handling negation in medical search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 1065–1066, Portland Oregon USA. ACM.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2018. Focal Loss for Dense Object Detection. *arXiv preprint*. ArXiv:1708.02002 [cs].

Jonathan Mallinson, Aliaksei Severyn, Eric Malmi, and Guillermo Garrido. 2020. FELIX: Flexible Text Editing Through Tagging and Insertion. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1244–1255, Online. Association for Computational Linguistics.

Kelong Mao, Zhicheng Dou, Haonan Chen, Fengran Mo, and Hongjin Qian. 2023. Large Language Models Know Your Contextual Search Intent: A Prompting Framework for Conversational Search. *arXiv preprint*. ArXiv:2303.06573 [cs].

Saeed Mehrabi, Anand Krishnan, Sunghwan Sohn, Alexandra M. Roch, Heidi Schmidt, Joe Kesterson, Chris Beesley, Paul Dexter, C. Max Schmidt, Hongfang Liu, and Mathew Palakal. 2015. Deepen: A negation detection system for clinical text incorporating dependency relation into negex. *Journal of Biomedical Informatics*, 54:213–219.

Felice Antonio Merra, Omar Zaidan, and Fabricio De Sousa Nascimento. 2023. Improving the Relevance of Product Search for Queries with Negations. In *Companion Proceedings of the ACM Web Conference 2023*, pages 86–89, Austin TX USA. ACM.

Priyanka Nigam, Yiwei Song, Vijai Mohan, Vihan Lakshman, Weitian, Ding, Ankit Shingavi, Choon Hui Teo, Hao Gu, and Bing Yin. 2019. Semantic Product Search. *arXiv preprint*. ArXiv:1907.00937 [cs].

Georgios Peikos, Daria Alexander, Gabriella Pasi, and Arjen P. de Vries. 2023. Investigating the Impact of Query Representation on Medical Information Retrieval. In *Advances in Information Retrieval*, Lecture Notes in Computer Science, pages 512–521, Cham. Springer Nature Switzerland.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks.

Gerard Salton and Chris Buckley. 1990. Improving retrieval performance by relevance feedback. *Journal of the American society for information science*, 41(4):288–297.

Zehong Tan, Canran Xu, and Mengjie Jiang. Query Rewrite for Null and Low Search Results in eCommerce.

Stuart J. Taylor and Sanda M. Harabagiu. 2018. The Role of a Deep-Learning Method for Negation Detection in Patient Cohort Identification from Electroencephalography Reports. *AMIA Annual Symposium Proceedings*, 2018:1018–1027.

Aaron Van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. pages arXiv–1807.

Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu. 2018. Additive margin softmax for face verification. volume 25, pages 926–930. IEEE.

Yaxuan Wang, Hanqing Lu, Yunwen Xu, Rahul Goutam, Yiwei Song, and Bing Yin. 2021. QUEEN: Neural Query Rewriting in E-commerce.

Ziyue Wang, Aozhu Chen, Fan Hu, and Xirong Li. 2022. Learn to understand negation in video retrieval. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 434–443, New York, NY, USA. Association for Computing Machinery.

Orion Weller, Dawn Lawrie, and Benjamin Van Durme. 2023. NevIR: Negation in Neural Information Retrieval. *arXiv preprint*. ArXiv:2305.07614 [cs].

Rong Xiao, Jianhui Ji, Baoliang Cui, Haihong Tang, Wenwu Ou, Yanghua Xiao, Jiwei Tan, and Xuan Ju. 2019. Weakly Supervised Co-Training of Query Rewriting andSemantic Matching for e-Commerce. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 402–410, Melbourne VIC Australia. ACM.

Shi Yu, Jiahua Liu, Jingqin Yang, Chenyan Xiong, Paul Bennett, Jianfeng Gao, and Zhiyuan Liu. 2020. Few-shot generative conversational query rewriting. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 1933–1936, New York, NY, USA. Association for Computing Machinery.

Mengxiao Zhang, Yongning Wu, Raif Rustamov, Hongyu Zhu, Haoran Shi, Yuqi Wu, Lei Tang, Zuohua Zhang, and Chu Wang. 2022. Advancing Query Rewriting in E-Commerce via Shopping Intent Learning.

Simiao Zuo, Qingyu Yin, Haoming Jiang, Shaohui Xi, Bing Yin, Chao Zhang, and Tuo Zhao. 2022. Context-Aware Query Rewriting for Improving Users' Search Experience on E-commerce Websites. *arXiv preprint*. ArXiv:2209.07584 [cs].