Unsupervised Synonym Extraction for Document Enhancement in E-commerce Search

Hanqing Lu*
luhanqin@amazon.com
Amazon.com Inc
Palo Alto, USA

Yunwen Xu*
yunwenx@amazon.com
Amazon.com Inc
Palo Alto, USA

Qingyu Yin*
qingyy@amazon.com
Amazon.com Inc
Beijing, China

Tianyu Cao
caoty@amazon.com
Amazon.com Inc
Palo Alto, USA

Boris Aleksandrovsky
skelab@amazon.com
Amazon.com Inc
Palo Alto, USA

Yiwei Song
ywsong@amazon.com
Amazon.com Inc
Palo Alto, USA

Xianlong Fan
fanxl@amazon.com
Amazon.com Inc
Beijing, China

Bing Yin
alexbyin@amazon.com
Amazon.com Inc
Palo Alto, USA

ABSTRACT

The vocabulary gap between search queries and product descriptions is an important problem in modern e-commerce search engines. Most of the existing methods deal with the vocabulary gap issues by rewriting user-input queries. In this work, we describe another way to address vocabulary gap issues in the e-commerce search systems. In particular, we propose an unsupervised synonym extraction framework for document enhancement. Comparing to existing methods, the purposed framework has two main differences: 1) instead of extending and rewriting user-input queries, we make the synonym tokens searchable by adding them into the text descriptions of products; 2) the whole process is unsupervised and doesn't require any human labels. A two-phase unsupervised synonym discovery framework is proposed for extracting synonym rules from the search log data. We demonstrate the effectiveness of our approach by two experiments: 1) we do online A/B testing experiments in multiple countries, which show significant improvements in key business metrics; 2) we conduct a human audit to evaluate the quality of the extracted synonym rules, which indicates that 85.5% of the extracted rules are high quality pairs.

CCS CONCEPTS

• **Information systems → Query intent**; • **Computing methodologies → Information extraction**.

KEYWORDS

synonym, e-commerce, query understanding

*Both authors contributed equally to this paper.

1 INTRODUCTION

E-commerce customers express their purchase intents in several ways, some of them may use a different vocabulary than that of the product description [15]. For instance, customers could use "cat rx food" to express their needs for "cat prescription food". Though the word "rx" and "prescription" have the same meaning in the given context, the query "cat rx food" returns much less relevant products due to the vocabulary gap between query terms and product descriptions (as shown in Figure 1). Admittedly, the vocabulary gap issue heavily hurts the performance of keyword-based e-commerce search systems which perform in a "matching and ranking" manner. Typically, in the matching stage, a search engine relies on lexical similarity between product descriptions and customer keywords. In the above example, only "prescription" is indexed for most of the relevant products, but not "rx". In other words, these products are not searchable with token "rx". Therefore, the input query "cat rx food" can not find these products, which leads to matching failure and bad customer experience.

One straightforward way to alleviate this vocabulary gap issue is to rewrite query terms by using synonyms [16]. In this way, relevant products that match the customers' search intent can be retrieved. In the previous example, if we replace the term "rx" in the query with its synonym "prescription", more related products can be retrieved and shown to customers. In literature, query rewriting techniques have been widely explored [1, 11, 15, 25, 27]. Some of these methods utilize user behavior information in the form of click-through rate [5], query co-occurrence [12], and co-clicked query similarity [1, 12] to rewrite queries, and some others employ external knowledge base [17, 19, 22] to expand search keywords. Recently, [16] propose a novel query rewriting method by utilizing
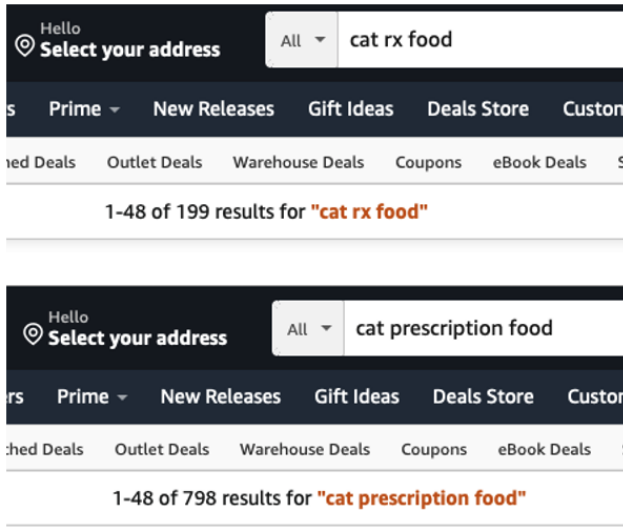
Figure 1: Search with 'rx' (199 results) returns less relevant products than search with 'prescription' (798 results).

synonym expansions that are extracted from both the user behavior data and the external knowledge base. However, given the fact that there are multiple ways of describing the same entity regardless of the level of formality, it is sometimes hard to define proper expansions or substitute terms for a given query. One alternative way to obtain the target query is utilizing the well-performing queries that are defined as head queries with high click-through rate [7, 15]. Nevertheless, the vocabulary gap issue may not be fully resolved since we cannot guarantee that the rewritten query keywords actually appear in the product description of relevant products. In the meanwhile, most of the recent approaches for query rewriting are supervised methods [15, 16] that require human labels. This fact sets the barrier for applying them to product search system that runs on multiple locales where different languages are used.

To better address the above problems, in this paper, we propose a document-side synonym expansion method to deal with the vocabulary gap in e-commerce search engine. Instead of extending and rewriting customer-input keywords, we try to directly add the synonym tokens into the product description. In particular, these additional tokens for document enhancement are generated by applying the synonym expansions to product title text. For each generated synonym rule, the source word/phrase is from product title and the target word/phrase is from the related query. Therefore, by indexing additional informative tokens extracted from the related queries, the vocabulary mismatch between products and queries can be alleviated. At the first place, we generate highly engaged product title-query pairs from customer query logs. After that, we extract n-gram pairs (one from product title and the other from the associated query) as synonym candidates based on both the *co-occurrence frequency* and the *transition probability*. To ensure semantic similarity of the n-gram terms in candidate pairs, we employ a BERT-based [6] embedding model to filter out the irrelevant n-gram pairs. This provides us with the synonym rules. For indexing, we apply the synonym rules to expand product title

text and add these additional tokens to products' indices. The whole process is unsupervised. In search time, these additional tokens for products will be utilized in the lexical matching stage, so that these products can be matched even though the tokens in search keyword do not match that of the original product description.

We favor this strategy in two aspects. First, our method has more direct impact to bridge the vocabulary gap between products and queries since we directly add tokens that from related customer keywords to products. Second, the whole process is unsupervised, which means we don't need human labeled synonyms to train models. Thus, our method could be easily applied to large-scale production system. We run online A/B testing experiments in multiple locales using different languages and show the efficiency of our approach.

The major contributions of this paper are listed as follows:

- We proposed an document-side synonym expansion method to resolve the vocabulary gap issue between product texts and customer-input keywords in e-commerce search.
- The proposed method is unsupervised and can be easily applied to large-scale production search system.
- We empirically evaluated the effectiveness of our method by running an online A/B testing experiment.

The rest of this paper is organized as follows. Section 2 describes our unsupervised synonym discovery method for document enhancement. Section 4 presents our experiment details. We describe related work in Section 5, and the last section is about the conclusion.

2 METHODOLOGY

In this section, we introduce the technical details of the proposed unsupervised synonym discovery method and how we apply those synonyms to enhance the documents. We first give an overview of the whole method. After that, we introduce in details about the candidate generation process. Then we illustrate the filtering strategy that eliminates low-quality synonyms. At last, we explain how we apply these synonyms on the document side to help the matching process of e-commerce search.

2.1 Overview

Figure 2 gives an overview of the proposed synonym discovery framework. In the high-level, our synonym rules are extracted from two different data sources: dictionaries and query logs. In this work, we employed two dictionaries to generate general synonym candidates: the WordNet synset [18] and Wikipedia redirects data. For query log mining method, we generated synonym candidates by utilizing the related query-product pairs. In order to better utilize the customer feedback, we first characterized the relatedness of each query and product by using customer behavior information, i.e., click and purchase actions and then selected related query-product pairs. In this way, we can collect plenty number of related query and product pairs and then employ these parallel data to generate n-gram pairs, where the first n-gram term comes from the customer keywords and the other one comes from the correspondence product title. Among these pairs, we only selected ones with high co-occurrence frequency and transition probability as the candidate synonym pairs. Finally, the candidate synonym pairs
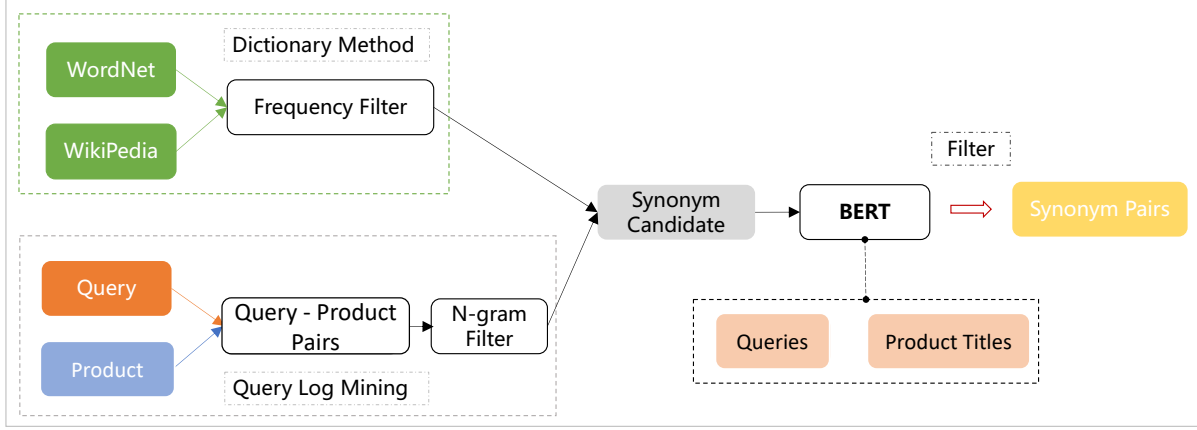
Figure 2: Illustration of our method.

generated by using these two data sources were combined to be the ultimate candidate set.

In order to capture the similarity of n-gram terms in candidate synonym pairs at the semantic level, we employed BERT [6] to generate the representation vectors of n-gram tokens. Specially, we trained the BERT model by using the data collected from search queries and product titles. For each n-gram pair, we fed them into the BERT and apply the average-pooling over the output hidden states to generate correspondence representation vectors. We then calculated the cosine similarity between terms in each n-gram pair and kept the ones with scores bigger than the specific threshold. In this way, we can collect millions of synonym rules and apply these rules for document enhancement. In practice, we expanded the definitions of product by applying synonym rules on each word in the product title and made these additional words searchable. Consequently, these additional words can be utilized in the matching stage, so that the vocabulary gap between queries and products can be alleviated.

2.2 Synonym Candidate Generation

In this subsection, we describe how to generate the candidate synonym pairs from different data sources in details, including the dictionary method and query log mining.

2.2.1 Dictionary Method. We extracted general synonym candidates from the WordNet synset and Wikipedia redirects data. WordNet is one of the authoritative sources for English. It mostly provides synonyms for single words, such as [handbag, bag, wrinkle, purse, pocketbook] are all defined as synonyms. Many of them are good candidates for noun phrases. Wikipedia redirects data is based on

Wikipedia title redirects links, and are also generated manually. Wikipedia[1] covers a vast range of topics, some of them are helpful in search. An example of useful redirect pair is "cba" to "Chinese basketball association", which links the acronym from âĂIJcbsâĂİ to its complete form.Then word alignment and phrase extraction are implemented to prepare substitute words (such as [cba, Chinese basketball association]).

2.2.2 Query Log Mining. Admittedly, query logs are from customers' historical searches and contain rich customer feedback information. In order to extract synonym pairs from query logs, we need to generate query and related product pairs in the first place. A straightforward way to accomplish this is to directly utilize query and purchased (or clicked) products. However, since the number of purchased products is limited, the coverage of synonym pairs generated by this approach will be relatively low. Therefore, we employ a query-product graph building method to generate plenty number of parallel query-product pairs.

Specifically, we build a query-product bipartite graph from query logs, where the nodes present queries and products. For a given query node and a product node, an edge exists only when at least one of click or purchase actions are made for the query and correspondence product. The value of each edge is the linear combination of the action attributes, including purchases and clicks. After that, queries and products nodes can be clustered by the label propagation algorithm [32]. In this way, we collected queries and products that are interconnected among them but not with other nodes in the graph.
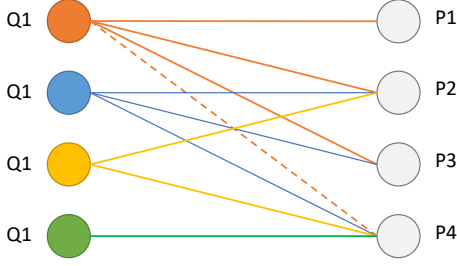
---

[1]https://en.wikipedia.org

Figure 3: Illustration of the proposed query-product graph, where Q is the query node and P presents the product node. A solid line indicates a direct connection and a dotted line denotes an indirect connection.

For each query and product pair, we can gain direct connections by using the edges. Nevertheless, we can also find indirect connections in this graph. To determine indirect matches, we vectorized the queries using a weight vector by using the connected products in the same graph and computed the cosine similarity between each pair of query nodes within a given cluster. Figure 3 shows an example of our approach where we show a cluster with various queries and products. In this example, Q1-P1, Q1-P2 and Q1-P3 are direct matches; the pair Q1-P4 is an indirect match. If the similarity is above a minimum threshold, we collected a set of products that are connected to at least one query in the given query pair, and then connect both queries to all the products in the set. Finally, we can gain plenty number of query-product pairs and the synonyms should appear within the user keywords and the correspondence product titles.

After generating query-product pairs, we extracted n-gram pairs directly from these parallel sequence of tokens: one from the user keywords and another one from product title. For a given pair of n-gram tokens, source A (from product title) and target B (from user keyword), we calculated both the *co-occurrence frequency*, $co(A, B)$ and *transition probability* $p(A, B)$. The *transition probability* can be obtained by using the following formulation:

$$p(A, B) = Count(A, B) / \sum_{B' \in S(A)} (Count[A, B']) \qquad (1)$$

where $S(A)$ presents all the n-grams associate with A. In particular, *transition probability* indicates the exclusive association of the given n-gram pairs, by which we can make sure the association is not very spread. For each n-gram pair, we calculated the multiplication of $co(A, B)$ and $p(A, B)$, and dropped out low-scoring ones.

2.3 Filtering

We employed the BERT model[6] to filter candidate synonym pairs generated from different data sources. BERT has been attracted much attention in recent NLP tasks. In this work, we firstly trained a BERT model on the search queries and product title dataset, then we fine-tuned it by using Query-Intent labels [13]. After that, we used the BERT model to generate 768-dimension embedding for each n-gram in our candidate synonym pairs. We filtered out all the synonym pairs with cosine similarity score of their embeddings less than 0.6.

3 DEPLOYMENT: DOCUMENT ENHANCEMENT WITH SYNONYMS

For each product, we extracted its title and used the generated synonym rules to find the synonym tokens. We added all these synonym tokens to a new index field of a given product to make them searchable. We didn't apply synonym expansion to all the product text fields because many text fields contain noisy/defect tokens, and applying synonym expansion to these tokens could introduce more defects.

4 EXPERIMENT

4.1 End-to-end Online Experiments

To evaluate the effectiveness of the document side synonym expansion method, we performed an online A/B test on a global e-commerce site. We first extracted over one million synonym pairs using the method described in Section 2. Then, we used these rules to expand the text in the product description. Specifically, in the treatment group, we indexed the products with additional synonym tokens, making them searchable even when different words were used in search queries. While the synonyms were not utilized in the control group. We considered the impact on both business metrics, such as sales, as well as recall improvement, characterized by the percent of searches resulting in less than one page of results (sparse result rate). As a guardrail, we also employed human evaluation for search defect on a few thousand randomly sampled top-ranked query, result pairs in each country [2].

In all countries, we found the proposed method statistically significantly improved key business metrics. These results indicated our method led to a better retrieval performance and can help customers fulfill their shopping missions when the vocabulary gap existed. Notably, the human evaluated defect rate was equal or better in all countries. This was a strong evidence that we injected reasonably relevant synonym tokens to the search index.

4.2 Synonym Quality Evaluation

In addition to the overall recall improvement by applying the document enhancement in search, we also directly evaluated the quality of the synonyms generated by our approach detailed in Section 2.2 and 2.3. Since the process is unsupervised, we conducted human audits on a few hundred expanded texts focusing on the relatedness of synonyms. Specifically, for each applied synonym expansion example, we assigned one of the three labels to the rules and context: "irrelevant", "related" (relevant in the context), or "fully interchangeable". After audit, we found 14.5% expansions were irrelevant, 33.1% were relevant in the context. 52.4% were fully interchangeable. This means about 85.5% expansions are semantically related and are effective in addressing the vocabulary gap.

5 RELATED WORK

In this section, we will give a brief introduction about the recent work on 1) query rewrite and 2) synonym discovery.

---

[2]Each query, result pair is manually judged as relevant or irrelevant, the defect rate is defined by the percentage of irrelevant pairs.

5.1 Query Rewrite

Query Rewriting has been shown to be effective in improving search results on queries with the vocabulary gap. A variety of customer feedback based techniques have been proposed for commercial search engines [31]. Basically, these method utilize implicit user feedback in the form of click-through rate [5], query co-occurrence [12], and co-clicked query similarity [1, 12] to rewrite queries. However, such a method may fail for queries with no or few results, which is known as pseudo-relevance feedback. In recent years, deep learning techniques have been extensively studied for query rewriting [9, 11, 15, 23]. [23] first regard the query rewriting task as machine translation, where a monolingual translation system is for generating expansion terms. Together with query-snippet pairs from user query logs, these extracted terms used to train a statistical machine translation model, which is employed to generate the final rewrited queries. However, most of the existing query rewriter models are sub-optimal for search engines. To better deal with this issue, [11] propose a learning-based query rewrite method that consists of a candidate generating phase and a candidate ranking phase. Experimental results on a Yahoo search engine demonstrate the effectiveness of their method. [15] propose a novel query rewriting method, aiming at rewriting vocabulary gap queries to well-performing ones. Specially, a Co-Attentive MaLSTM is proposed to better measure the similarity between query pairs at the semantic level. Though these methods gain great performance in solving vocabulary gap issues in the search engine, we would like to try to deal with the vocabulary gap problem by directly adding the searchable tokens into product description information, which we think, is a more straightforward way of bridging the vocabulary gap between queries and products.

5.2 Synonym Discovery

Synonym discovery has been a popular research topic in natural language processing for years. A straightforward way of dealing with synonym discovery problem is to utilize the dictionaries such as WordNet [18] and ConceptNet [14] that contain many common and general sense synonyms constructed manually based on domain knowledge. After that, these extracted synonyms can be used to generate alternative keywords candidates for query expansion [29], or index expansion [4]. Meanwhile, some methods learn to generate synonyms in historical search logs [29, 30] and expand search queries by synonym patterns. With the recent success in deep learning technology [2, 6], deep neural network models for synonym discovery have been widely explored [10, 24]. [10] introduce a novel word-embedding-based approach to address the synonym extraction of multi-word terms. [16] propose an end-to-end synonym generation system that learns to generate synonym candidates by using a set of filtering components. A machine learning classifier is then employed to select correct synonym rules from the candidates. Finally, the generated synonym dictionary is incorporated for improving retrieval.

5.3 Document Expansion

Other than directly addressing vocabulary gap issues existing in search engine with enhancing query representations, there are also many studies focus on enriching document representations

[3, 8, 20, 21, 26, 28]. [26] first illustrated the effectiveness of document expansion method for speech retrieval, where a text collection related to the speech is utilized for enhancing the contents of the speech. [3] showed the advantage of using document expansion by comparing with query expansion in the same framework. [28] focused on the task of language model information retrieval and proposed a document expansion technique to deal with the insufficient sampling issues, where the documents are expanded with their neighborhood information. [8] further showed the effectiveness of aggressive document expansion for short documents. In recent study, [20] proposed a method of using neural network models for document expansion. In particular, a sequence-to-sequence model is employed to predict the possible queries for a given documents and these queries are then used as the external information to expands the contents of the documents. Compared with existing work, in our study, we focus on document expansion for e-commerce search, where the documents (products) are always short and the vocabularies used in the queries could be noisy. In addition, we explore to expand the contents of products by using only synonyms. As far as we know, this is the first time that synonyms are employed for document expansion.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented an unsupervised synonym extraction framework for enhancing the product document. Specifically, we proposed a two-phase unsupervised synonym discovery framework that extracts synonym rules from user behavior log data. The extracted synonyms were used to expand the product titles. We did multiple experiments to demonstrate the effectiveness of our approach. In the online A/B testing experiment, it showed significant increased purchase and conversion in all six countries. In the offline human audit, we identified that 85.5% of extracted rules are high quality pairs. For the next step, we will focus on incorporating knowledge graphs to further improve the quality of the extracted synonym rules.

REFERENCES

[1] I Antonellis, H Garcia-Molina, and C-C Simrank+ Chang. [n.d.]. Query rewriting through link analysis of the click graph. *Proceedings of VLDB (Dec 2008)* ([n. d.]), 408–421.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[3] Bodo Billerbeck and Justin Zobel. 2005. Document expansion versus query expansion for ad-hoc retrieval. In *Proceedings of the 10th Australasian document computing symposium*. Citeseer, 34–41.

[4] Adrian Boteanu, Adam Kiezun, and Shay Artzi. 2018. Synonym Expansion for Large Shopping Taxonomies. (2018).

[5] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma. 2002. Probabilistic query expansion using query logs. In *Proceedings of the 11th international conference on World Wide Web*. 325–332.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186.

[7] Doug Downey, Susan Dumais, and Eric Horvitz. 2007. Heads and tails: studies of web search with common and rare queries. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 847–848.

[8] Miles Efron, Peter Organisciak, and Katrina Fenlon. 2012. Improving retrieval of short texts through document expansion. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 911–920.

[9] Jianfeng Gao, Xiaodong He, Shasha Xie, and Alnur Ali. 2012. Learning lexicon models from search logs for query expansion. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.* Association for Computational Linguistics, 666–676.

[10] Amir Hazem and Béatrice Daille. 2018. Word Embedding Approach for Synonym Extraction of Multi-Word Terms. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

[11] 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.* 1443–1452.

[12] Rosie Jones, Benjamin Rey, Omid Madani, and Wiley Greiner. 2006. Generating query substitutions. In *Proceedings of the 15th international conference on World Wide Web.* 387–396.

[13] Heran Lin, Pengcheng Xiong, Danqing Zhang, Fan Yang, Ryoichi Kato, Mukul Kumar, William Headden, and Bing Yin. 2020. Light Feed-Forward Networks for Shard Selection in Large-scale Product Search. (2020).

[14] Hugo Liu and Push Singh. 2004. ConceptNetâĂŤa practical commonsense reasoning tool-kit. *BT technology journal* 22, 4 (2004), 211–226.

[15] Subhadeep Maji, Rohan Kumar, Manish Bansal, Kalyani Roy, Mohit Kumar, and Pawan Goyal. 2019. Addressing Vocabulary Gap in E-commerce Search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1073–1076.

[16] Aritra Mandal, Ishita Khan, and Prathyusha Senthil Kumar. 2019. Query Rewriting using Automatic Synonym Extraction for E-commerce Search. (2019).

[17] Rila Mandala, Takenobu Tokunaga, and Hozumi Tanaka. 1999. Combining multiple evidence from different types of thesaurus for query expansion. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval.* 191–197.

[18] George A Miller. 1998. *WordNet: An electronic lexical database.* MIT press.

[19] Roberto Navigli and Paola Velardi. 2003. An analysis of ontology-based query expansion strategies. In *Proceedings of the 14th European Conference on Machine Learning, Workshop on Adaptive Text Extraction and Mining, Cavtat-Dubrovnik, Croatia.* Citeseer, 42–49.

[20] Rodrigo Nogueira, Wei Yang, Jimmy Lin, and Kyunghyun Cho. 2019. Document expansion by query prediction. *arXiv preprint arXiv:1904.08375* (2019).

[21] Jeremy Pickens, Matthew Cooper, and Gene Golovchinsky. 2010. Reverted indexing for feedback and expansion. In *Proceedings of the 19th ACM international conference on Information and knowledge management.* 1049–1058.

[22] Yonggang Qiu and Hans-Peter Frei. 1993. Concept based query expansion. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval.* 160–169.

[23] Stefan Riezler and Yi Liu. 2010. Query rewriting using monolingual statistical machine translation. *Computational Linguistics* 36, 3 (2010), 569–582.

[24] Jiaming Shen, Ruiliang Lyu, Xiang Ren, Michelle Vanni, Brian Sadler, and Jiawei Han. 2019. Mining Entity Synonyms with Efficient Neural Set Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence,* Vol. 33. 249–256.

[25] Gyanit Singh, Nish Parikh, and Neel Sundaresan. 2012. Rewriting null e-commerce queries to recommend products. In *Proceedings of the 21st International Conference on World Wide Web.* 73–82.

[26] Amit Singhal and Fernando Pereira. 1999. Document expansion for speech retrieval. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval.* 34–41.

[27] Zehong Tan, Canran Xu, Mengjie Jiang, Hua Yang, and Xiaoyuan Wu. 2017. Query Rewrite for Null and Low Search Results in eCommerce. (2017).

[28] Tao Tao, Xuanhui Wang, Qiaozhu Mei, and ChengXiang Zhai. 2006. Language model information retrieval with document expansion. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference.* 407–414.

[29] Xing Wei, Fuchun Peng, Huihsin Tseng, Yumao Lu, and Benoit Dumoulin. 2009. Context sensitive synonym discovery for web search queries. In *Proceedings of the 18th ACM conference on Information and knowledge management.* 1585–1588.

[30] Xing Wei, Fuchun Peng, Huishin Tseng, Yumao Lu, Xuerui Wang, and Benoit Dumoulin. 2010. Search with synonyms: problems and solutions. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters.* Association for Computational Linguistics, 1318–1326.

[31] Jinxi Xu and W Bruce Croft. 2017. Quary expansion using local and global document analysis. In *Acm sigir forum,* Vol. 51. ACM New York, NY, USA, 168–175.

[32] Xiaojin ZhuÐČ and Zoubin GhahramaniÐČÐ¡. 2002. Learning from labeled and unlabeled data with label propagation. (2002).