Stochastic Optimization of Text Set Generation for Learning Multiple Query Intent Representations

Helia Hashemi
University of Massachusetts Amherst
hhashemi@cs.umass.edu

Hamed Zamani
University of Massachusetts Amherst
zamani@cs.umass.edu

W. Bruce Croft
University of Massachusetts Amherst
croft@cs.umass.edu

ABSTRACT

Learning multiple intent representations for queries has potential applications in facet generation, document ranking, search result diversification, and search explanation. The state-of-the-art model for this task assumes that there is a sequence of intent representations. In this paper, we argue that the model should not be penalized as long as it generates an accurate and complete set of intent representations. Based on this intuition, we propose a stochastic permutation invariant approach for optimizing such networks. We extrinsically evaluate the proposed approach on a facet generation task and demonstrate significant improvements compared to competitive baselines. Our analysis shows that the proposed permutation invariant approach has the highest impact on queries with more potential intents.

CCS CONCEPTS

• Information systems → Query representation; **• Computing methodologies → Natural language generation**.

KEYWORDS

Query representation learning; facet generation; text set generation

ACM Reference Format:
Helia Hashemi, Hamed Zamani, and W. Bruce Croft. 2022. Stochastic Optimization of Text Set Generation for Learning Multiple Query Intent Representations. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557666

1 INTRODUCTION

Learning effective query representations has always played a key role in information retrieval (IR) systems. Early approaches for query representation mostly focused on term-based representations (e.g., TF-IDF weighting in vector space models [32]). Their semantic representations have also been studied in latent semantic indexing (LSI) [4], bag-of-words embedding-based models [45, 46], and contextual embedding-based models [10, 41]. State-of-the-art solutions for obtaining accurate query representations fine-tune large language models, e.g, BERT [5] and BART [20], on a downstream retrieval task. These approaches often learn a single representation for each query or query term. However, this is a sub-optimal solution for representing ambiguous or faceted queries that can be associated to multiple different intents. To tackle this problem, Hashemi et al. [8] recently proposed NMIR – an encoder-decoder framework for learning multiple vectors for a query, each representing a potential query intent. NMIR aims at learning multiple representations by generating different query intent descriptions. Despite the strong effectiveness achieved by NMIR, it ignores the *permutation invariance nature of query intents*. In other words, it assumes that the query intents should be generated as a sequence. With this assumption, a method that learns accurate query representations and precisely generates the query intent descriptions but in a different order than the ground truth will be significantly penalized by the loss function. In this paper, we propose a solution to address this fundamental shortcoming.

The proposed approach, named PINMIR,[1] looks at the query intents as a *set* rather than a sequence. Given the *unordered structure of sets*, PINMIR uses a permutation invariant loss function for optimization and thus learns more accurate query representations. Permutation invariant losses often consider all possible permutations of the predicted output which quickly becomes computationally inefficient as the number of intents increases. To address this issue, we also propose a stochastic variation of our permutation invariant loss. Besides the loss function, we use a simple solution based on resetting the positional embedding of transformer decoders for permutation invariant decoding.

As a pioneering work for learning multiple intent representations of the query, Hashemi et al. [8] demonstrated that query facet generation can be successfully used for extrinsic evaluation of the learned query representations. Following their advice, we evaluate our model on a query facet generation task.[2] Our experiments on the large-scale MIMICS dataset demonstrate the effectiveness of the proposed solutions compared to state-of-the-art baselines. Our experiments suggest that the proposed permutation invariant approach has the highest impact on queries with more intents. We also show that our stochastic loss is as effective as an exact permutation invariant loss, while being more efficient.

It is notable that although we lay out our proposed optimization approach based on NMIR for learning multiple query representations, it can be simply adoptable for any sequence-to-sequence model that generates a unordered set of text pieces.

2 RELATED WORK

This section briefly reviews prior work related to multiple query representation learning and set neural networks.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557666

[1] stands for the Permutation Invariant Neural Multiple Intent Representation model.

[2] In this paper, we use "facets" and "intent descriptions", interchangeably.

Query Representation. Traditionally, queries were represented based on term occurrences and frequencies [32]. However, these models suffer from the vocabulary mismatch problem. Several studies have tried to address this issue mostly with query expansion and pseudo-relevance feedback [2, 18, 31, 47]. Latent semantic indexing (LSI) is one of the early approaches for learning a semantic representation for queries and documents. It calculates a term frequency matrix given a piece of text and uses singular value decomposition for embedding the given text in a semantic space. Word embedding models, e.g., word2vec [26] and GloVe [30], learn word representations by predicting the next word in a text. Thus, queries can be represented based on their individual query term embeddings [45]. Most recently large language models, such as BERT [5], are used for representing queries and documents. All these approaches only generate a single representation per query or query term. To the best of our knowledge, NMIR [8] is the only neural approach that learns multiple intent representations for search queries. However, this approach ignores the permutation invariance nature of query intents. Such a simplifying assumption leads to a sub-optimal solution [50]. This paper addresses this shortcoming by introducing a permutation invariant variation of NMIR.

Query Facet Extraction and Generation. In our experiments, we use query facet generation as an extrinsic evaluation methodology. Early work on facet extraction and generation focused on facet extraction and generation for e-commerce and digital libraries using external resources and metadata [3, 11, 17, 21, 35]. These models, however, are not applicable to an open domain setting, such as web search [39]. To adopt facet generation methods to open domain, a number of methods conduct local analysis on the top retrieved documents in response to the query. For instance, Kong and Allan proposed a number of supervised methods for facet extraction from the web [12–14]. Dou et al. [6] proposed a facet extraction approach that use a hybrid model called QDMiner. This paper uses these models as baselines. There is also a line of research that studies query variations [22]. For instance, Xue and Croft [42] modeled queries as a distribution over query variations. Learning multiple query vectors has potential impact on all of these tasks.

Set Neural Networks. Set neural networks can be considered as set-input networks and set-output networks. Most existing models focus on set-input problems, where the input of the network is a set of items. An algorithm designed for set-input problems should satisfy two conditions. First, it should be permutation invariant, meaning the model's prediction should remain the same under any permutation of the input. Second, such model should take variable input size. Therefore, existing network architectures such as MLP and RNN cannot be used for input sets [27, 28, 38]. One line of work to handle set inputs uses pooling architectures for permutation invariant mapping [24, 33, 34, 36]. Their core idea is to apply the neural function F to each set item individually, and apply a pooling permutation invariant function (e.g., sum or average). Zaheer et al. [44] discuss the structure of *set pooling* methods and prove that they are a universal approximator for any set function. More recently, attention-based approaches come to the play for the set networks [9, 40, 43]. For instance, Lee et al. [19] proposed Set Transformer which allows the model to encode pairwise or higher order interactions between items in a set.

Set-output networks are less explored. To design a set-output network, the model needs to satisfy two condition. First, the model must be permutation-equivariant; meaning the generation of a particular permutation of output should be as probable as any other permutation. Second, the loss function should be permutation invariant. Recently, Zhang et al. [49] introduced a model for permutation-equivariant set generation. Following their work, researchers worked on a transformer variant for predicting a set of object properties[15, 23]. The majority of these approaches study computer vision problems and do not focus on text set generation.

3 METHODOLOGY

Learning multiple representations for a single query or generally speaking a piece of text is not a straightforward task. This is even more challenging when the number of representations varies among different instances. NMIR proposed by Hashemi et al. [8] is the pioneering work in this area and also the current state-of-the-art approach for learning multiple representations for search queries. In this section, we introduce PINMIR that extends NMIR by satisfying a permutation invariance constraint. We first briefly introduce NMIR and then describe our extension. Note that the proposed optimization approach is not restricted by the network architecture and can be applied to other networks beyond NMIR. Our proposed training schema is adaptable to any other network that generates a set of unordered text pieces.

Problem Statement. Let $Q = \{q_1, q_2, \ldots, q_n\}$ be a training query set with n queries, and $D_i = \{d_{i1}, d_{i2}, \ldots, d_{im}\}$ be the top m retrieved documents in response to the query q_i using an arbitrary retrieval model M. In addition, let $F_i = \{f_{i1}, f_{i2}, \ldots, f_{ik_i}\}$ denote the set of all intent descriptions (facets) associated with the query q_i, where k_i is the number of query intents and can vary across queries. The task is to learn k_i representations $R_i = \{R_{i1}, R_{i2}, \ldots, R_{ik_i}\}$ for the query q_i, each associated with a query intent in F_i.

3.1 A Brief Overview of NMIR

Our approach extends NMIR [8] which uses an encoder-decoder transformer architecture. Let $\phi(\cdot)$ and $\psi(\cdot)$ denote a text encoder and decoder, respectively. The encoder takes the query q_i and documents D_i as input. The model assumes that the top retrieved documents are relevant to the query (similar to the pseudo-relevance feedback assumption), and the retrieval model M retrieves a diverse set of documents. The core idea is to use the query and the top retrieved documents, and extract some information from them to generate all query intent descriptions or facets. The representations led to different query facets can be used as the multiple representations of the query. The more precise facets get generated, the more accurate multiple representations are expected. NMIR clusters the top retrieved documents and assigns each cluster to a query intent description $f_{ij} \in F_i$ using a greedy algorithm, say γ:

$$C_i^* = \gamma\left(\phi(d_{i1}), \phi(d_{i2}), \ldots, \phi(d_{im}), \phi(f_{i1}), \phi(f_{i2}), \ldots, \phi(f_{ik_i})\right)$$

where $C_i^* = \{C_{i1}^*, C_{i2}^*, \ldots, C_{ik_i}^*\}$ is a set of document sets. Each C_{ij}^* is a set of documents from D_i that are assigned to f_{ij} by γ which uses k-means to cluster the document representations produced by the encoder. The number of clusters in training time is defined by the number of query intents in ground truth. However, at inference time the number of clusters are unknown. NMIR considers two scenarios. In the first scenario, it assumes that the number of clusters is constant for all queries, and in the second scenario, it uses non-parametric K-Means [25] to handle dynamic number of clusters.

The decoder's input for generating the j^{th} intent description is a concatenation of the query string and the first $j - 1$ intent descriptions, separated by a special token. NMIR uses the cross entropy loss function of sequence-to-sequence models [37], thus expects that the predictions follow the same order as the ground truth. NMIR parameters are initialized using BART pre-trained parameters [20] and Guided Transformer [7] is used for adjusting the query representation based on each document cluster. The proposed optimization solution is orthogonal to the network architecture choice, therefore, we refer the reader to [8] for more information on the NMIR architecture.

3.2 The Permutation Invariant NMIR

Despite its strong performance, NMIR still suffers from some limitations. First, it uses the standard sequence-to-sequence optimization, as as result, it assumes that the query intents are ordered and it tries to optimize the model to produce intent descriptions in the same order as it appears in the ground truth. Second, NMIR uses a greedy algorithm for assigning each cluster to a ground truth query intent during training. Therefore, the model's performance depends on this heuristic cluster-intent assignment algorithm. In this paper, we introduce a permutation invariant optimization solution for text generation, when each element of the set is a piece of text. We explain our model as a variant of NMIR, where the performance of the model is not sensitive to the order of generated query intent descriptions. In this model, we no longer need the intent-cluster matching algorithm, since the order of generated intents do not matter. A side benefit is that in reality, sometimes documents address more than one query intent and assigning only one intent to a document would be sub-optimal.

First, we need to define a permutation invariant loss function for training the model. Common permutation invariant loss functions include Chamfer loss and Hungarian loss. Chamfer loss is based on Chamfer distance that was first introduced in computer vision [1]. Although it is more efficient, it is not applicable to our task due to the design of decoder for text generation. The reason is that the decoder generates the output token-by-token and the closest ground truth facet is not known until the facet is fully generated. Therefore, we extend the Hungarian loss [16] for text set generation. The proposed loss function for a query q_i is computed as follows:

$$
\begin{aligned}
L(\hat{F}_i, F_i) &= \min_{F'_i \in \pi(F_i)} L_{CE}(\hat{F}_{ij}, F'_{ij}) \\
&= \min_{F'_i \in \pi(F_i)} \frac{1}{k_i} \sum_{j=1}^{k_i} \sum_{t=1}^{|f'_{ij}|} -\log p(f'_{ijt} | v, f'_{ij1}, f'_{ij2}, \cdots, f'_{ijt-1})
\end{aligned}
$$

where $\pi(F_i)$ denotes all permutations of ground truth intents for query q_i. Therefore, the size of $\pi(F_i)$ is equal to $k_i!$. The loss function L_{CE} is the average sequence-to-sequence loss for generating each facet description, and v denotes the encoder representation. Intuitively, the proposed loss function computes all permutations of ground truth set and considers the one with the minimum loss value, which is the loss value for the closest ground truth ordering to the generated set. Therefore, the original ordering of ground truth text would not impact the loss value.

This loss function can be quite expensive to compute, since it requires us to repeat this process for every permutation of the query intents. We propose to use a **stochastic variation** of this loss that instead of iterating over all possible permutations, takes s samples from the permutation set and computes the loss based on the sampled query intent sequences. Our experiments show that the stochastic loss performs comparably to the non-stochastic variation of the loss, which is computationally expensive.

Position Resetting. We highlight that in our task in contrast to the standard assumption in *set networks*, although the order does not matters between the set elements, it matters within each individual element. In other words, the order that the model generates different query intent descriptions does not matter, but it is important that sequence of tokens in each query intent description get generated legitimate, both semantically and syntactically. To help the model capture this concept, we modify the standard decoder architecture in transformer. The decoder generates tokens one-by-one and each token becomes the decoder's input for generating the next token. The standard transformer decoder uses *position embedding* for every token. However, in PINMIR, we reset the position embedding of decoder for every intent description. In other words, the position at the start of every new intent description is equal for all intents. In that case, the decoder representations for every permutation of a given set of intents would be identical.

4 EXPERIMENTS

Following Hashemi et al. [8], we evaluate the model on a query facet generation task. The task is defined as generating a number of textual facet descriptions for a given multi-faceted query.

Data. We use the MIMICS collection[3] in our experiments. It consists of three datasets. We use MIMICS-Click–the largest MIMICS dataset in our experiments, which consists of over 400K unique web search queries. We dedicate a random sample of 80% of data to the training, 10% to the validation, and 10% to the test set.[4] The top retrieved documents in response to each query is obtained by the Bing's public web search API.[5] We only use the documents' snippets to represent a document.

Evaluation Metrics. We use four different metrics to evaluate our model. The first group is "term overlap" that has been used for the facet extraction models [12]. We compare the precision, recall, and macro-averaged F1 score between the model prediction and the ground truth. The second group is "exact match". This group, again, computes the precision, recall, and macro-averaged F1 score for the exact facet descriptions. The third metric, is based on Set BLEU [8]. It extends BLEU [29] to a set of output text by selecting the best matched ordering R^* from all permutations of R such that $R^* = \arg\max_{R' \in \text{perm}(R)} \frac{1}{M} \sum_{i=1}^{M} \text{BLEU-4}(R'_i, G_i)$, where i denotes the facet index, G_i is the i^{th} ground truth facet, and $M = \max(|G|, |R|)$. Finally, the Set BLEU score is calculated as

[3] MIMICS is publicly available at https://github.com/microsoft/MIMICS.

[4] The NMIR paper [8] mentions that it uses MIMICS-Manual for evaluation. However, it was a typo and it uses 10% of the MIMICS-Click dataset.

[5] The MIMICS SERP data is available at http://ciir.cs.umass.edu/downloads/mimics-serp/MIMICS-BingAPI-results.zip.

Table 1: Results for the query facet generation experiment. The superscript * denotes statistically significant improvements compared to all the baselines using two-tailed paired t-test with Bonferroni correction at 99% confidence level.

#		Term Overlap			Exact Match			Set BLEU				Set BERT-Score		
facets	Model	Prec	Recall	F1	Prec	Recall	F1	1-gram	2-gram	3-gram	4-gram	Prec	Recall	F1
variable	QDist	0.0969	0.1564	0.1195	0.0017	0.0023	0.0019	0.1999	0.1134	0.0360	0.0107	0.6772	0.6855	0.6100
	QFI	0.1461	0.1748	0.1571	0.0057	0.0061	0.0059	0.2763	0.1269	0.0421	0.0140	0.7069	0.7113	0.6144
	QFJ	0.1807	0.2041	0.1894	0.0069	0.0067	0.0067	0.2484	0.1065	0.0242	0.0090	0.7196	0.6708	0.5871
	QDMiner	0.2060	0.2456	0.1894	0.0076	0.0083	0.0079	0.2893	0.1226	0.0301	0.0126	0.7220	0.7025	0.6285
	BART	0.4307	0.4618	0.4481	0.0474	0.0516	0.0486	0.4459	0.4003	0.3896	0.3351	0.7623	0.6932	0.6558
	NMIR	0.4851	0.5673	0.4968	0.0790	0.0842	0.0784	**0.5187**	0.4748	0.4470	0.4192	0.8003	0.7487	0.6928
	PINMIR	**0.4891**	**0.5691**	**0.5107***	**0.0798**	**0.0856**	**0.0795**	0.5173	**0.4763**	**0.4491**	**0.4246***	**0.8173***	**0.7524***	**0.7199***
max	QDist	0.1557	0.1593	0.1440	0.0023	0.0024	0.0023	0.3387	0.1048	0.0439	0.0176	0.7165	0.7802	0.7192
	QFI	0.1605	0.1941	0.1720	0.0058	0.0050	0.0050	0.3539	0.1524	0.0523	0.0203	0.7603	0.8127	0.7584
	QFJ	0.1767	0.1348	0.1451	0.0055	0.0057	0.0053	0.3735	0.1675	0.0564	0.0234	0.7731	0.8136	0.7714
	QDMiner	0.2176	0.1443	0.1773	0.0069	0.0066	0.0065	0.4275	0.1826	0.0657	0.0234	0.7758	0.8036	0.7792
	BART	0.3043	0.4124	0.3558	0.0282	0.0263	0.0275	0.5087	0.4406	0.3969	0.3445	0.7633	0.8017	0.7660
	NMIR	0.3877	**0.4559**	0.4121	0.0613	0.0584	0.0596	0.6313	0.5628	0.5222	0.4871	0.8442	0.8870	0.8405
	PINMIR	**0.4712***	0.4302	**0.4423***	**0.0731***	**0.0689***	**0.0677***	**0.6505***	**0.5732***	**0.5411***	**0.4895***	**0.8731***	**0.8873**	**0.8740***

$\frac{1}{M}\sum_{i=1}^{M}$ BLEU-n(R_i^*, G_i) for all n-grams. The last metric, BERT-Score [48], is a neural based metric that use BERT to find the semantic similarity between a single text prediction and a set of facets in the ground truth. Similar to Set BLEU, we use Set BERT-Score as $\frac{1}{M}\sum_{i=1}^{M}$ BERT-Score(R_i^*, G_i).

Results and Discussion. We compare our model with the following baselines: (1) the query variations generated by the QDist model proposed by Xue and Croft [42], (2, 3) two effective graphical models proposed by Kong and Allan [14] for facet extraction in web search, named QFI and QFJ, (4) a hybrid method for query facet extraction for web search, named QDMiner [6], (5) fine-tuned BART model [20] for facet generation that uses a pre-trained transformer encoder-decoder architecture and is also used in our model, and finally (6) the recent NMIR model of Hashemi et al. [8] which is the same as our model without consideration the permutation invariance nature of intents.

Each query in MIMIMCS contains between two and five facets. The majority of queries in this dataset only have two facets. Each query in our dataset contains an average of 2.81 facets per query. The results for our first set of experiments on this dataset are reported in Table 1 (# facets = variable). The proposed method generally outperforms all the baselines. The improvements in terms of exact match are marginal, while we observe significant improvements for term overlap F1, BLEU 4-gram, and Set BERT-Score.

Intuitively, we expect a permutation invariant loss to have higher impact on queries with more facets. In our second set of experiments, we solely focus on the queries with 5 facets (i.e., the maximum number of facets in MIMICS). According to Table 1, we observe substantially larger improvements in queries with five facets. The improvements are statistically significant in nearly all cases, except for term overlap recall and Set BERT-Score recall. This observation demonstrates that the permutation invariant model has higher impacts on the queries with more intents.

As mentioned in Section 3.2, this paper proposes the Stochastic Hungarian loss for efficiency reasons. In our experiments, we observe no statistically significant difference between the effectiveness of a model trained with Hungarian loss compared to its stochastic variation (with three samples). Hungarian loss achieves a term overlap F1 of 0.4724 for queries with three facets while this value for the Stochastic Hungarian loss is 0.4731. We made similar observations for other metrics too, which are not reported due to space limitations. Therefore, both exact and stochastic Hungarian losses perform comparably. This is while the stochastic variation can be used for larger number of facets efficiently by sampling from all permutations.

5 CONCLUSIONS AND FUTURE WORK

In this work, we introduced a model that learns generating a set of text pieces in an permutation invariant manner. We explained our model compared to NMIR, an existing model recently proposed by Hashemi et al. [8], to learn multiple representations for a search query. NMIR, despite performing strongly, suffers from some design limitations. In particular, the NMIR's solution to achieve multiple representations for a query is to generate all the query intents associated with that query. However, the model expects the output to be exactly in the same order as it appears in the ground truth. We introduced a novel variation of NMIR that compensates its drawbacks, and makes the model permutation invariant regarding different intents. We trained the model with a stochastic manner with a new loss function that we introduced. By resetting the positional embedding for each intent description generated by the model, we ensured that the model's decoder is also permutation invariant. We showed that our model outperforms competitive baselines for a facet generation task.

In the future, we intend to evaluate the impact of permutation invariant models on document ranking, search result diversification, and clarifying question selection. We also believe that the proposed solutions can be generalized to many text set generation tasks. We will explore this direction in our future work.

Acknowledgements. This work was supported in part by the Center for Intelligent Information Retrieval and in part by a Bloomberg Data Science PhD Fellowship. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

[1] Harry G. Barrow, Jay M. Tenenbaum, Robert C. Bolles, and Helen C. Wolf. 1977. Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching. In *IJCAI*. 659–663.

[2] W. B. Croft and D. J. Harper. 1979. Using Probabilistic Models of Document Retrieval Without Relevance Information. *J. Doc.* 35, 4 (1979), 285–295.

[3] Wisam Dakka and Panagiotis G. Ipeirotis. 2008. Automatic Extraction of Useful Facet Hierarchies from Text Databases. *2008 IEEE 24th International Conference on Data Engineering* (2008), 466–475.

[4] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *J. Assoc. Inf. Sci.* 41, 6 (1990), 391–407.

[5] 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 (NAACL '19)*. ACL, Minneapolis, Minnesota, 4171–4186.

[6] Zhicheng Dou, Zhengbao Jiang, Sha Hu, Ji-Rong Wen, and Ruihua Song. 2016. Automatically Mining Facets for Queries from Their Search Results. *IEEE Trans. on Knowl. and Data Eng.* 28, 2 (2016), 385–397.

[7] Helia Hashemi, Hamed Zamani, and W. Bruce Croft. 2020. Guided Transformer: Leveraging Multiple External Sources for Representation Learning in Conversational Search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 1131–1140.

[8] Helia Hashemi, Hamed Zamani, and W. Bruce Croft. 2021. Learning Multiple Intent Representations for Search Queries. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, 669–679.

[9] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. 2018. Attention-based Deep Multiple Instance Learning. *CoRR* (2018).

[10] Omar Khattab and Matei Zaharia. 2020. ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 39–48.

[11] Christian Kohlschütter, Paul-Alexandru Chirita, and Wolfgang Nejdl. 2006. Using Link Analysis to Identify Aspects in Faceted Web Search.

[12] Weize Kong and James Allan. 2013. Extracting Query Facets from Search Results. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland) *(SIGIR '13)*. ACM, New York, NY, USA, 93–102.

[13] Weize Kong and James Allan. 2014. Extending Faceted Search to the General Web. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (Shanghai, China) *(CIKM '14)*. 839–848.

[14] Weize Kong and James Allan. 2016. Precision-Oriented Query Facet Extraction. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16)*. 1433–1442.

[15] Adam R. Kosiorek, Hyunjik Kim, and Danilo J. Rezende. 2020. Conditional Set Generation with Transformers. *CoRR* (2020).

[16] Harold W. Kuhn. 1955. *Naval Research Logistics Quarterly* 1–2 (1955), 83–97.

[17] K. Latha, K. R. Veni, and R. Rajaram. 2010. AFGF: An Automatic Facet Generation Framework for Document Retrieval. In *2010 International Conference on Advances in Computer Engineering*. 110–114.

[18] Victor Lavrenko and W. Bruce Croft. 2001. Relevance Based Language Models. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New Orleans, Louisiana, USA) *(SIGIR '01)*. ACM, New York, NY, USA, 120–127.

[19] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. 2018. Set Transformer. *CoRR* (2018).

[20] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7871–7880.

[21] Chengkai Li, Ning Yan, Senjuti B. Roy, Lekhendro Lisham, and Gautam Das. 2010. Facetedpedia: Dynamic Generation of Query-Dependent Faceted Interfaces for Wikipedia. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. 651–660.

[22] Binsheng Liu, Xiaolu Lu, and J. Shane Culpepper. 2021. Strong Natural Language Query Generation. *Inf. Retr.* 24, 4–5 (oct 2021), 322–346.

[23] Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. 2020. Object-Centric Learning with Slot Attention. *CoRR* (2020).

[24] David Lopez-Paz, Robert Nishihara, Soumith Chintala, Bernhard Schölkopf, and Léon Bottou. 2017. Discovering Causal Signals in Images. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI, USA) *(CVPR '17)*. IEEE Computer Society, 58–66.

[25] A. Meyerson. 2001. Online facility location. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. 426–431.

[26] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems* (Lake Tahoe, Nevada) *(NeurIPS '13)*. Curran Associates Inc., Red Hook, NY, USA, 3111–3119.

[27] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. 2013. Domain Generalization via Invariant Feature Representation.

[28] Junier Oliva, Barnabas Poczos, and Jeff Schneider. 2013. Distribution to Distribution Regression. In *Proceedings of the 30th International Conference on Machine Learning*.

[29] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (Philadelphia, Pennsylvania) *(ACL '02)*. ACL, USA, 311–318.

[30] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP '14)*. ACL, Doha, Qatar, 1532–1543.

[31] J. J. Rocchio. 1971. Relevance Feedback in Information Retrieval. In *The SMART Retrieval System: Experiments in Automatic Document Processing*. 313–323.

[32] G. Salton, A. Wong, and C. S. Yang. 1975. A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 11 (Nov. 1975), 613–620.

[33] Baoguang Shi, Song Bai, Zhichao Zhou, and Xiang Bai. 2015. DeepPano: Deep Panoramic Representation for 3-D Shape Recognition. *IEEE Signal Processing Letters* (2015).

[34] Jake Snell, Kevin Swersky, and Richard S. Zemel. 2017. Prototypical Networks for Few-shot Learning. *CoRR* (2017).

[35] Emilia Stoica, Marti Hearst, and Megan Richardson. 2007. Automating Creation of Hierarchical Faceted Metadata Structures. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*. 244–251.

[36] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik G. Learned-Miller. 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. In *Proceedings of the 2015 IEEE International Conference on Computer Vision* (Santiago, Chile) *(ICCV '15)*. IEEE Computer Society, 945–953.

[37] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (Montreal, Canada) *(NIPS'14)*. MIT Press, Cambridge, MA, USA, 3104–3112.

[38] Zoltán Szabó, Bharath K. Sriperumbudur, Barnabás Póczos, and Arthur Gretton. 2016. Learning Theory for Distribution Regression. *J. Mach. Learn. Res.* 17, 1 (jan 2016), 5272–5311.

[39] Jaime Teevan, Susan Dumais, and Zachary Gutt. 2008. Challenges for Supporting Faceted Search in Large, Heterogeneous Corpora like the Web. In *HCIR 2008*.

[40] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. 2016. Order Matters: Sequence to sequence for sets.

[41] Lee Xiong, Chenyan Xiong, Ye Li, Kwok-Fung Tang, Jialin Liu, Paul Bennett, Junaid Ahmed, and Arnold Overwijk. 2021. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In *International Conference on Learning Representations (ICLR '21)*.

[42] Xiaobing Xue and W. Bruce Croft. 2013. Modeling Reformulation Using Query Distributions. *ACM Trans. Inf. Syst.* 31, 2, Article 6 (may 2013), 34 pages.

[43] Bo Yang, Sen Wang, Andrew Markham, and Niki Trigoni. 2018. Attentional Aggregation of Deep Feature Sets for Multi-view 3D Reconstruction. *CoRR* (2018).

[44] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. 2017. Deep Sets. *CoRR* (2017).

[45] Hamed Zamani and W. Bruce Croft. 2016. Estimating Embedding Vectors for Queries. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval* (Newark, Delaware, USA) *(ICTIR '16)*. ACM, New York, NY, USA, 123–132.

[46] Hamed Zamani and W. Bruce Croft. 2017. Relevance-Based Word Embedding. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. ACM, New York, NY, USA, 505–514.

[47] Chengxiang Zhai and John Lafferty. 2001. Model-Based Feedback in the Language Modeling Approach to Information Retrieval. In *Proceedings of the Tenth International Conference on Information and Knowledge Management* (Atlanta, Georgia, USA) *(CIKM '01)*. ACM, New York, NY, USA, 403–410.

[48] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *Proceedigns of the 8th International Conference on Learning Representations (ICLR '20)*.

[49] Yan Zhang, Jonathon S. Hare, and Adam Prügel-Bennett. 2019. Deep Set Prediction Networks. *CoRR* (2019).

[50] Yan Zhang, Jonathon S. Hare, and Adam Prügel-Bennett. 2020. FSPool: Learning Set Representations with Featurewise Sort Pooling. *CoRR* (2020).