Augmentation of Local Government FAQs using Community-based Question-answering Data

Yohei Seki
yohei@slis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

Masaki Oguni
m-oguni@klis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

Sumio Fujita
sufujita@yahoo-corp.jp
Yahoo Japan Corporation
Tokyo, Japan

ABSTRACT

To reduce the cost of administrative services, many local governments provide a frequently asked questions (FAQ) page on their websites that lists the questions received from local inhabitants with their official responses. The number of Q&A items posted on the FAQ page, however, will vary depending on the local government. To address this issue, we propose a method for augmenting local government FAQs by using a community-based Q&A (cQA) service. We also propose a new FAQ augmentation task to identify the regional dependence of Q&A to achieve the goal mentioned above. In our experiments, we fine-tuned the bidirectional encoder representations from transformers (BERT) model for this task, using a labeled local-government FAQ dataset. We found that the regional dependence of Q&As can be identified with high accuracy by using both the question and the answer as clues and with fine tuning for the deeper layers in BERT.

CCS CONCEPTS

• **Information systems** → *Question answering*; • **Human-centered computing** → Accessibility technologies.

KEYWORDS

FAQ augmentation, community-based QA (cQA), local government, bidirectional encoder representation from transformers (BERT)

ACM Reference Format:
Yohei Seki, Masaki Oguni, and Sumio Fujita. 2020. Augmentation of Local Government FAQs using Community-based Question-answering Data. In *The 22nd International Conference on Information Integration and Web-based Applications & Services (iiWAS '20), November 30-December 2, 2020, Chiang Mai, Thailand*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3428757.3429137

1 INTRODUCTION

Many local governments provide frequently asked questions (FAQ) pages on their websites, aiming to reduce the cost of staff time required to respond to questions from the local-area inhabitants. By using the FAQ page, inhabitants can often resolve their own issues about civic life without having to contact the local government office directly. However, the number of Q&As posted on the FAQ page will differ among the various local governments. For example, the FAQ page provided by Yokohama City, an ordinance-designated city in Japan, contained 2,733 Q&As in January, 2020. In contrast, some medium-sized local governments provide only a few Q&As, with some not even having an FAQ page. One of the solution is to enrich the FAQs relevant to administrative services. However, manually adding the Q&As will incur substantial costs.

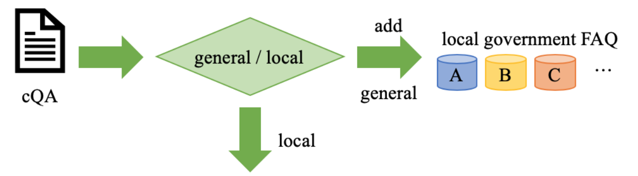

Figure 1: Overview of local government FAQ augmentation

In this study, we propose a method for augmenting local government FAQs using a community-based question-answering (cQA) service. A cQA service is an Internet-based service whereby users can post questions and obtain answers from other users. The largest cQA service in Japan is Yahoo! Chiebukuro[1], which posts more than 700 million Q&As in various categories. Many users use the cQA service because of its ease of use and post their questions according to their real-time requirements. Therefore, the questions in cQA cover a wider range of topics than those included in local government FAQs, which are typically asked by inhabitants via emails or phone calls to their local government offices. Moreover, a cQA service tends to provide more detailed responses, often answered by category-expert users, than those in local government FAQs, which are answered by local office staff. Therefore, by augmenting the local government FAQ by using cQA as an additional resource, we should be able to provide more fruitful Q&As.

We aim to augment the local government FAQs using data from a cQA service by identifying regional dependencies in the Q&As, as described in Figure 1. To enrich the FAQ pages of any local government, we aim to add Q&As that are not regionally dependent[2] (i.e., a general Q&A). In this paper, we address this goal by formalizing a new FAQ-augmentation task to identify the presence or absence of regional dependence in a set of Q&As, and evaluate the results.

For our experiments, we fine-tuned the bidirectional encoder representations from transformers model (BERT) [4], using a local government FAQ dataset labeled to indicate the presence or absence of regional dependence. In addition, we extracted documents

[1] https://chiebukuro.yahoo.co.jp/
[2] We use the term "region" in the sense of "city," and "Q&A without regional dependence" refers to a Q&A that can be used across cities under the same national polity.

iiWAS '20, November 30-December 2, 2020, Chiang Mai, Thailand

ACM ISBN 978-1-4503-8922-8/20/11…$15.00
https://doi.org/10.1145/3428757.3429137

embeddings (DEs) by several methods and analyzed their effectiveness. We also compared the effect on identifying the regional dependence of the Q&As for three cases: using the question only, using the answer only, and using both question and answer as input clues.

Our contributions are summarized as follows. (1) We propose a method for augmenting local government FAQs using a cQA dataset. (2) We evaluate a task involving regional dependence using a deep neural network and provide a labeled corpus using a cQA dataset.

2 RELATED WORK

Sakata et al. [11] proposed a method for presenting appropriate Q&As from the local government FAQ pages that were relevant to queries found in administrative dialogue systems. However, if there is no appropriate Q&A relevant to the queries, the dialog system cannot provide appropriate answers to the inhabitant's question. Therefore, it is necessary to enrich the local government FAQ, but this will be expensive if done manually. To address this, Chatterjee et al. [2] proposed an FAQ augmenter for enterprise FAQ pages. They use multi-task deep neural networks to estimate the similarity between Q&A pairs and the query. Many other researchers [1, 3, 14] have conducted question retrieval tasks for cQA-based data to bridge lexical gaps in questions. In this study, we propose a method for augmenting local government FAQ from a cQA service dataset based on the similarity between cQA and original local government FAQs. To estimate the similarity, we use BERT [4] to cover the questions with lexical gap to the original questions.

To augment the local government FAQ using cQA dataset, we define a new task to identify regional dependency in cQA. Yan et al. [13] utilized a cQA dataset for the task-oriented dialogue system by classifying user's intent in questions. However, as far as we know, there have been no studies to identify regional dependency in cQA. For this task, we apply document embedding method for questions and answers using BERT. In pre-training models such as BERT, linguistic properties such as morphological, local-syntax, and longer-range semantics tend to be treated at different layers, such as the word embedding layer, deeper contextual layers, or surface layers in each of these cases [6, 10]. In Section 6, we discuss and clarify which layers are effective for fine-tuning to identify regional dependency in cQA.

3 LOCAL GOVERNMENT FAQS DATASET

We collected Q&As from the FAQ pages of websites for three ordinance-designated cities in Japan, namely Yokohama[3], Sapporo[4], and Kobe[5]. A total of 4,718 Q&As were collected from each city: 2,733 from Yokohama, 1,098 from Sapporo, and 887 from Kobe.

3.1 Regional Dependence Identification for Augmenting Local Government FAQs

To our knowledge, no cQA dataset related to local government FAQs has been published to date. Because cQA services contain a wide range of Q&As, it is difficult to narrow down the Q&As to those relevant to local government FAQs. Therefore, we have created a cQA dataset comprising Q&As related to administrative services by estimating their similarity to local-government FAQ datasets, as described in Section 4. In this dataset, we label whether or not each input cQA is regionally dependent, which determines whether the Q&A can be applied across regions. In our experiments (see Section 6), we use the local government dataset as training data and cQA dataset as test data for evaluating the accuracy in identifying regional dependence in Q&As.

3.2 Labeling of the Local Government FAQs dataset

To label the local government FAQs, we randomly extracted 1,500 of the 4,718 Q&As for labeling. The remaining unlabeled 3,218 Q&As were used for the augmentation exercise (see Section 4). We conducted the labeling task via a crowdsourcing platform, Yahoo! Crowdsourcing[6]. We asked the crowd workers to label the Q&As in terms of regional dependence by suggesting labeled examples. We requested that a "local" label be assigned to Q&As having regional dependence and a "general" label otherwise. The examples for "general" and "local" labeled questions are as follows.

- (General) Is it able to use national health insurance for traffic accident?
- (General) When I live abroad for while, should I continue to join national pension plan?
- (Local) How to apply housing loan credit for city tax?
- (Local) Give me the information about available (childcare) nursery schools.

We assigned labeling tasks to the workers so that each Q&A was labeled by three workers. We set five Q&As to be labeled per task, with one worker able to participate in up to ten labeling tasks. That is, each worker labeled no more than 50 Q&As. Table 1 shows the labeling results for the local government FAQ dataset.

Table 1: Labeling results for local government FAQ dataset.

Labeling pattern	Number of cases
general: 3, local: 0	262
general: 2, local: 1	317
general: 1, local: 2	422
general: 0, local: 3	499
Total	1,500

From Table 1, 579 Q&As were labeled as "general" by two or more workers out of three, and 921 Q&As were labeled as "local" by two or more workers out of three. Many Q&As labeled as "local" described the facilities managed by the particular local government.

To assess the reliability of the labeling criterion, we calculated Fleiss's κ coefficient [5], which indicates the degree of agreement among the responses from three or more annotators. The result was 0.321 in our case (i.e., "fair agreement" [9]). We also conducted an analysis of examples where the responses differed between workers. We checked the free description field for those Q&As with only one "local" annotation. Here, the "local" annotators commented that they were not sure if the same system applied to other regions or whether

[3]https://www.city.yokohama.lg.jp/
[4]https://www.city.sapporo.jp/
[5]https://www.city.kobe.lg.jp/
[6]https://crowdsourcing.yahoo.co.jp/

the system could be applied nationwide. They therefore used a "local" annotation, even though some of these Q&As should actually have been labeled as "general". We also analyzed those Q&As that only one worker labeled as "general". We found that these workers commented that they thought the local specific facility could be used nationwide. Finally, we confirmed that the majority of the answers to these Q&As were consistent with the answers we expected.

4 EXTRACTING CQA DATASET RELEVANT TO LOCAL GOVERNMENT FAQS

Here, we describe how to extract a cQA dataset relevant to the local government FAQ described in Section 3. As mentioned above, cQA contains a wide range of Q&As, making it difficult to extract those Q&As that are relevant to administrative services. To address this issue, we extracted a cQA dataset by using the local government FAQ data described in Section 3.2 in the following procedure.

(1) Extract document embeddings (DEs) from the local government FAQ dataset and the cQA corpus.
(2) Extract a cQA dataset similar to the Q&As used in local government FAQs.

4.1 Extracting a cQA dataset similar to Q&As in local government FAQs

To estimate the similarity, we extracted the DEs based on a Japanese BERT model[7], which had been previously trained using the Japanese Wikipedia corpus. To extract the DEs, we used MeCab [8] (with NEologd[8] as the dictionary) to segment the Q&A into morphemes. We applied byte-pair encoding [12] to the morphemes, and added a special starting token (called "[CLS]") to the beginning of the sentence. We then applied the token sequence to the BERT model. The output from the model was a token embedding corresponding to each token. Here, we used the embedding corresponding to the [CLS] token extracted from the last layer of BERT as the DE.

In extracting the cQA data similar to the Q&As in the local government FAQs, we used the 3,218 unlabeled Q&As already collected (see Section 3.2). We used the L2 norm between vectors as the distance between the DEs. We extracted 5,000 Q&As from the Yahoo! Chiebukuro (cQA) corpus based on the smallest distances between the DEs for Q&As in the cQA corpus posted from 2015 to 2017 and the DEs for the 3,218 Q&As in the unlabeled local-government FAQ data. We discuss the evaluation of this strategy in Section 4.2.

4.2 Evaluation for Extracting cQAs Relevant to Local Government FAQs

We investigated the extracted questions in Section 4.1 and compared the top five question categories for the questions relevant and irrelevant to the administrative services. The results are shown in Table 2.

From this table, we found that "administrative procedures" was the majority category in relevant questions, while "railways& stations", "legal consultation", and "care& welfare" categories appeared in both relevant and irrelevant questions. To estimate the effectiveness of our proposed method, we compared the precision for extracting relevant questions with our proposed method and with sampling 100 questions in each category. The results are shown in Table 3. Note that our proposed model outperforms the sampling method with statistical significance (using one-tailed t-test, at significance level 5%).

Table 2: Top five question categories for relevant and irrelevant questions to administrative services in cQA

Relevant Questions		Irrelevant Questions	
Category	%	Category	%
Administrative Procedures	19.1	Care& Welfare	13.7
Railways& Stations	16.4	Postal& Delivery	9.8
Legal Consultation	15.5	Legal Consultation	7.8
Care& Welfare	10.0	Railways& Stations	7.8
Traffic Accident	6.4	Aquarium	6.5

Table 3: Evaluation for extracting questions relevant to local government FAQs by question categories

Question Category	Proposed method	Sampling method
Administrative Procedures	0.84	0.67
Railways& Stations	0.60	0.23
Legal Consultation	0.59	0.02
Care& Welfare	0.34	0.29

We also evaluate the precision of our ranking method from top rank until the bottom (5,000) rank. The result is shown in Figure 2. From this result, we found that our ranking strategy was able to extract cQAs stably with precision around 0.4 by the bottom rank. Finally, we selected 1,964 Q&As as the cases relevant to the local government FAQs.

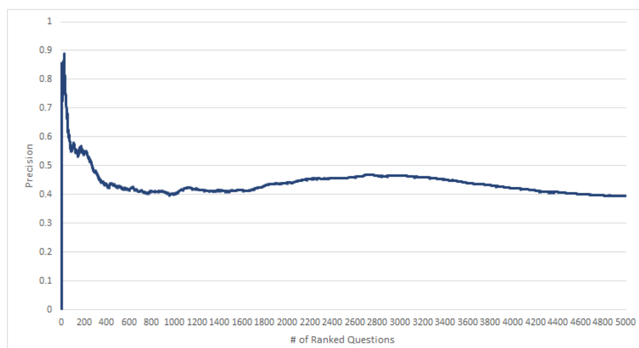

Figure 2: Transition of precision with question ranking to extract cQA relevant to local government FAQs

4.3 Labeling of the cQA dataset

Here, we describe the results of labeling the 1,964 Q&As using the Yahoo! Crowdsourcing platform. As in Section 3.2, we assigned labeling tasks to each worker so that each Q&A was labeled by three workers. Again, considering the load on the worker, we set five Q&As to be labeled per task, with any one worker participating

[7]https://github.com/cl-tohoku/bert-japanese
[8]https://github.com/neologd/mecab-ipadic-neologd

in a labeling task up to ten times. As a result, 1,443 Q&As were labeled "general" by the majority of three workers, and 521 Q&As were labeled "local" by the majority.

5 DE METHODS FOR REGIONAL DEPENDENCE IDENTIFICATION

In this section, we describe the various DE methods used in the experiments described in Section 6. To identify regional dependence, we extract DEs from BERT using several methods and compare them. We consider three cases. First, if we use both question and answer (the Q+A case), we apply "[CLS] Question [SEP] Answer [SEP]" as the input to the model, where "[SEP]" is a special separating token. Second, if we use only the question (the Q case), we apply "[CLS] Question [SEP]" as the input to the model. Third, if we use only the answer (the A case), we apply "[CLS] Answer [SEP]" as the input to the model. In our experiments, we investigated the CLS, MAX, and AVE methods (described in the following) for all three of these cases as to their performance in identifying regional dependence in Q&As. We also investigated the SEP_M, SEP_A, and SEP_C methods for the Q+A case.

- **CLS**
  A method that uses the embedding corresponding to the [CLS] token as the DE.
- **MAX**
  A method that uses the maximal pooling of token embeddings, except for embeddings corresponding to the [CLS], [SEP], and [PAD] tokens, as the DE.
- **AVE**
  A method that uses the average pooling of token embeddings, except for embeddings corresponding to the [CLS], [SEP], and [PAD] tokens, as the DE.
- **SEP_M**
  A method that uses the maximal pooling of embeddings corresponding to the [SEP] tokens for the Q+A case as the DE.
- **SEP_A**
  A method that uses the average pooling of embeddings corresponding to the [SEP] tokens for the Q+A case as the DE.
- **SEP_C**
  A method that uses the concatenated vector of embeddings corresponding to the [SEP] tokens for the Q+A case as the DE.

6 EXPERIMENT

In this section, we describe our experiments with the task of identifying the regional dependence in cQAs.

6.1 Experimental Dataset

We used the two datasets described in Section 3.2 and Section 4.3 in our experiments. The local-government FAQ dataset was used to fine-tune BERT and the cQA dataset was used as test data for the evaluation.

6.2 Experimental Method

We compared the DE methods described in Section 5 as to their performance in identifying regional dependence in test data. We investigated the fine tuning of BERT by varying the number of targeted layers. We then compared the results of identifying regional dependence in cQAs for the three cases: Q, A, and Q+A.

6.3 Results

Table 4 gives the results for the identification of regional dependence in cQAs. From the results, we found that progressively better results were obtained for many of the methods when fine-tuning was applied to the deeper layers from the final (12th) layer to the seventh layer. In contrast, the results deteriorated as fine-tuning was applied to layers closer to the surface than the sixth layer. In general, deep learning models in natural language processing tend to extract morphological features of sentences in the surface layers (close to the input) and semantic features of sentences in deeper layers (close to the output) [7]. In our Q&A data, it is the semantic features of sentences that differ according to their regional dependence, but no differences appear in the morphological features of sentences. We can conclude that fine-tuning to the surface layer will make little contribution to the identification results because morphological features are irrelevant to regional dependence in the Q&A data. Therefore, we focus on the results obtained when fine tuning is applied to the last (12th) layer only, from layers 12 to 11, layers 12 to 10, layers 12 to 9, layers 12 to 8, and layers 12 to 7.

First, we compare the results for all three cases (Q+A, Q, and A). Comparing the CLS, MAX, and AVE results, the best results were obtained for the Q+A case.[9] In investigating those results where the accuracy improved significantly for the Q+A case, we found that either the question or the answer included expressions that indicated some regional dependency. This suggests that, for the Q and A cases, the item was wrongly judged to be "local," whereas, for the Q+A case, it could be more accurately judged as "general." From these results, we can conclude that using both questions and answers leads to more effective identification of regional dependence in the cQAs.

Next, we compare the results for each method for each DE method for the Q+A case. The best results were obtained when fine tuning was performed from the last layer to the seventh layer and the MAX method was used. (Note that in Devlin et al. [4], the DEs were extracted using the CLS method described in this paper.) We compared a variety of methods, using the CLS method as the baseline, and found that the MAX method gave significantly better results than the CLS method[10].

7 CONCLUSION

In this study, we have proposed and investigated the task of identifying regional dependence in cQAs for the purpose of augmenting local government FAQs. From our results, we found that using the clues contained in both the cQA questions and answers coupled with fine tuning of only the deeper layers of our model provided

[9]The difference between the results for the Q+A case and the results for the other two cases is statistically significant at p<0.01.
[10]The difference between the results using the MAX method and those for the CLS method is statistically significant at p<0.05.

Table 4: Comparison results for identifying the regional dependence in cQA (F-values)

Fine-tuning	Question and Answer (Q+A case)						Question only (Q case)			Answer only (A case)		
layers	CLS	MAX	AVE	SEP_M	SEP_A	SEP_C	CLS	MAX	AVE	CLS	MAX	AVE
None	0.006	0.059	0.027	0.463	0.379	0.480	0.001	0.640	0.349	0.011	0.053	0.047
12	0.944	0.954	0.959	0.949	0.947	0.948	0.836	0.857	0.857	0.916	0.922	0.925
12 to 11	0.939	0.961	0.961	0.955	0.956	0.952	0.910	0.928	0.918	0.904	0.931	0.926
12 to 10	0.963	0.966	**0.968**	0.955	0.956	0.952	0.871	0.898	0.889	0.930	0.933	0.937
12 to 9	**0.968**	**0.969**	0.965	0.963	0.961	0.967	0.892	0.908	0.892	0.939	0.932	0.932
12 to 8	0.960	0.965	0.950	0.966	0.965	0.958	0.818	0.933	0.926	0.935	0.918	0.929
12 to 7	0.967	**0.972**	0.942	0.960	**0.968**	**0.970**	0.899	0.892	0.919	0.938	0.926	0.927
12 to 6	0.963	0.948	0.934	0.942	0.953	0.964	0.925	0.810	0.869	0.941	0.930	0.931
12 to 5	N/A	0.965	0.937	N/A	N/A	0.963	0.932	N/A	0.858	0.928	0.937	0.929
12 to 4	0.954	**0.969**	0.967	0.955	N/A	0.953	N/A	N/A	0.929	0.939	0.847	0.933
12 to 3	N/A	0.965	0.965	N/A	N/A	0.959	N/A	N/A	N/A	0.911	0.847	0.929
12 to 2	**0.969**	0.966	0.942	N/A	N/A	0.949	N/A	N/A	0.930	0.926	0.932	0.937
12 to 1	N/A	0.962	0.962	N/A	0.957	0.960	N/A	N/A	N/A	0.941	0.937	0.924

the most effective combination for the identification task. A cQA service tends to provide more detailed responses than existing local government FAQs, making it a suitable candidate for improving their FAQ pages. The results of the experiments showed that it is possible to identify regional dependency with high accuracy. This suggests that the augmentation of the local government FAQ using cQA is indeed a feasible approach.

One issue for future work is to estimate the applicable region for Q&As identified as "local," which would enable those regionally dependent Q&As to augment the local-government FAQ pages in the estimated region.

ACKNOWLEDGMENTS

This work was partially supported by a JSPS Grant-in-Aid for Scientific Research (B) (#19H04420).

REFERENCES

[1] Xin Cao, Gao Cong, Bin Cui, Christian S. Jensen, and Quan Yuan. 2012. Approaches to Exploring Category Information for Question Retrieval in Community Question-Answer Archives. *ACM Trans. Inf. Syst.* 30, 2, Article Article 7 (May 2012), 38 pages. https://doi.org/10.1145/2180868.2180869

[2] Ankush Chatterjee, Manish Gupta, and Puneet Agrawal. 2020. FAQAugmenter: Suggesting Questions for Enterprise FAQ Pages. In *Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM '20)*. Association for Computing Machinery, New York, NY, USA, 829–832. https://doi.org/10.1145/3336191.3371862

[3] Zheqian Chen, Chi Zhang, Zhou Zhao, Chengwei Yao, and Deng Cai. 2018. Question retrieval for community-based question answering via heterogeneous social influential network. *Neurocomputing* 285 (2018), 117 – 124. https://doi.org/10.1016/j.neucom.2018.01.034

[4] 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)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[5] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (1971), 378–382.

[6] Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What Does BERT Learn about the Structure of Language?. In *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy, 3651–3657. https://doi.org/10.18653/v1/P19-1356

[7] Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What Does BERT Learn about the Structure of Language?. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 3651–3657.

[8] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying Conditional Random Fields to Japanese Morphological Analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Barcelona, Spain, 230–237.

[9] J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (1977), 159–174.

[10] Matthew Peters, Mark Neumann, Luke Zettlemoyer, and Wen tau Yih. 2018. Dissecting Contextual Word Embeddings: Architecture and Representation. In *Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2018)*. Brussels, Belgium, 1499–1509.

[11] Wataru Sakata, Tomohide Shibata, Ribeka Tanaka, and Sadao Kurohashi. 2019. FAQ Retrieval Using Query-Question Similarity and BERT-Based Query-Answer Relevance. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR' 19)*. Association for Computing Machinery, New York, NY, USA, 1113–1116. https://doi.org/10.1145/3331184.3331326

[12] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1715–1725.

[13] Zhao Yan, Nan Duan, Peng Chen, Ming Zhou, Jianshe Zhou, and Zhoujun Li. 2017. Building Task-Oriented Dialogue Systems for Online Shopping. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI' 17)*. AAAI Press, 4618–4625.

[14] Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, and Ming Zhou. 2014. Question Retrieval with High Quality Answers in Community Question Answering. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM ' 14)*. Association for Computing Machinery, New York, NY, USA, 371–380. https://doi.org/10.1145/2661829.2661908