QueryNER: Segmentation of E-Commerce Queries

Chester Palen-Michel, Lizzie Liang, Zhe Wu, Constantine Lignos

Aspect-Value Extraction:

`High - end [speaker cover] for [B & W] [805d] 1 pair made of [velvet suede] made to order`

QueryNER Segmentation:

`[High - end] [speaker cover] for [B & W] [805d] [1 pair] [made of velvet suede] [made to order]`

Overview

- Query segmentation dataset for e-commerce
- 17 entity types
- Nearly 10,000 annotated queries
- Query segmentation allows us to determine:
  - Which portions of the query are most important to relevance?
  - Which portions of the query might be most safely dropped?
  - Can we weight different spans for relevance rather than tokens?
  - Can we link spans to a knowledge graph (e.g. known brands)?
- Dataset available on GitHub and Hugging Face

Corpus Content

	Queries	Entities	Tokens
Train	7,841	17,505	28,457
Dev	871	1,930	3,124
Test	933	2,317	3,610

Table 1: Dataset splits. Mean entity length is 1.6 tokens. Mean query length is 3.63 tokens.

Entity Type	Count
Core Product Type	8,310
Modifier	3,367
Creator	2,217
Department	1,652
Product Name	1,345
Content	1,301
UoM	862
Color	691
Shape	607
Material	569
Occasion	397
Condition	178
Quantity	104
Price	51
Origin	40
Time	32
Product Number	31

Table 2: Counts of entity types

Prior Work

- Past work focuses on aspect-value extraction which has limited token coverage (Papenmeier et al., 2021)
- Many prior datasets are unreleased (e.g. Faranza et al., 2023; Joshi et al., 2015)

Baselines

	Precision	Recall	F1
BERT	$60.94_{\pm 0.5}$	$60.17_{\pm 0.4}$	$60.56_{\pm 0.4}$
XLM-R	$60.45_{\pm 0.5}$	$59.75_{\pm 0.5}$	$60.10_{\pm 0.5}$
BERT-cont.	$61.78_{\pm 0.4}$	$60.82_{\pm 0.3}$	$61.29_{\pm 0.3}$

Table 3: Baseline results of BERT, XLM-R, and BERT with continued pre-training on the rest of the ESCI e-commerce dataset (Reddy et al, 2023)

Data Augmentation

Transformation	Example
Original	airforce 1 women shoes white
Shuffled	shoes women white airforce 1
Butterfinger	airvorce 1 women shoes white
Numeric	airforce 6 women shoes white
Color	airforce 1 women shoes green
Mention Replacement	zerogrand boys shoes leopard
All Transformations	shofs boys maple zerogrand

- Transformed data to simulate real query variation
- Training on additional transformed data mitigates impact on performance
- Model trained on original and augmented data maintains good performance on the original test set while improving performance on augmented data

Acknowledgments

This work was supported by the grant "Improving Relevance and Recovery by Extracting Latent Query Structure" from eBay to Brandeis University.

References

Reddy, Chandan and Choudhary, Nurendra and Maquez, Lluis and Valero, Fran and Rao, Nikhil and Zaragoza, Hugo and Bandyopadhyay, Sambaran and Biswas, Arnab. 2023. Shopping queries dataset: A large-scale ESCI benchmark for improving product search. *arXiv preprint arXiv:2206.06588*

Papenmeier, Andrea and Kern, Dagmar and Hienert, Daniel and Sliwa, Alfred and Aker, Ahmet and Fuhr, Norbert. 2021. Dataset of natural language queries for e-commerce. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval,* pages 307-311.

Farzana, Shahla, and Zhou, Qunzhi and Ristoski, Petar. 2023. Knowledge graph-enhanced neural query rewriting. In *Companion Proceedings of the ACM Web Conference 2023*, pages 911-919.

Joshi, Mahesh and Hart, Ethan and Vogel, Mirko and Ruvini, Jean-David. 2015. Distributed word representations improve NER for e-commerce. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 160–167, Denver, Colorado. Association for Computational Linguistics.