Leveraging Knowledge Graph and DeepNER to Improve UoM Handling in Search

Qunzhi Zhou, Zhe Wu, Jon Degenhardt, Ethan Hart, Petar Ristoski,
Aritra Mandal, Julie Netzloff, and Anu Mandalam

eBay Inc, San Jose, CA, 95125, USA
{qunzhou,zwu1,jdegenhardt,ejhart,pristoski,
arimandal,jnetzloff,amandalam}@ebay.com

Abstract. Understanding Unit of Measurements (UoM) is critical to e-commerce search engines. It is a challenging problem since numeric values and unit symbols are typically treated as regular text tokens. In this paper, we introduce a framework that utilizes Knowledge Graph (KG) and deep learning based Named Entity Recognition (NER) to provide better semantic understanding of UoMs.

Keywords: E-commerce, Knowledge Graph, Information Retrieval

1 Introduction

Query understanding that allows identifying relevant items from millions of products is fundamental for the success of any e-commerce platform. However, UoMs haven't been extensively studied in the context of search queries. A UoM usually consists of two parts, a numerical quantifier and a unit symbol. Identifying UoMs in text is challenging for several reasons: (i) UoMs can be represented in multiple measurement systems, e.g., metric and imperial system, (ii) UoMs have abbreviations that can be ambiguous, e.g., *g* for *gram* or *gigabyte*, (iii) user input doesn't always follow standard representation, e.g., *64.5" TV*, *64 1/2 inch TV*, *164cm TV* all have the same query intent, and results should be identical.

Several UoM ontologies have been introduced in the literature [1], most of which are tailored to a specific application or domain, making it difficult to reuse [2], or they provide only a catalog of abstract UoM concepts. Furthermore, most of existing work focuses on extracting UoMs from semi-structured data, like tabular data, while identifying UoMs in natural language, especially short texts like search queries, remains a challenge.

To address these challenges, we introduce a novel approach to support semantic query matching for UoMs based on Entity Linking over a KG, which contains a comprehensive set of unit entities and relations.

2 Approach

Our approach for matching UoMs in text consists of three main components, (i) knowledge graph, (ii) entity resolution, (iii) query rewrite. Our KG integrates product data with real-world knowledge domains including brands, colors, materials, UoM, etc. The UoM sub-graph captures around 800 entities and relations between them, including label variants, physical quantities, SI base units, and
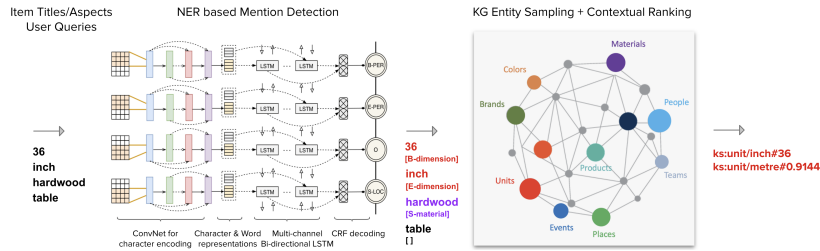
Fig. 1: UoM entity resolution for product data and search queries

conversion rules. A generic entity resolution service was developed to link text mentions in product item and query data to KG entities. As shown in Figure 1, a deep hybrid neural network NER model [3] and the knowledge graph are leveraged to incorporate context-based features to recognize and disambiguate UoM entities. NER tags are mapped to ontology classes and used as features to rank the relevance of entity candidates sampled from the KG. To support UoM entity matching, the search engine leverages the resolution service in both item data and user query processing flow. For each recognized UoM phrase, two UoM URIs are generated including the original URI and the canonical URI which represents the equivalent measurement in SI base unit. Both URIs are indexed for items and included in query rewrites.

3   Results

The proposed approach was evaluated in an offline and online setting. In the offline setting, a set of 10,000 random search queries were used to measure the change in recall and precision. The human evaluation with independent annotators showed considerable increase in recall, while the relevance remained the same. Secondly, we conducted an online A/B test, exposing the new experience to millions of users. For the treated traffic, we observed statistically significant drop in search abandonment rate and decrease in low recall search sessions. This matches well with the observed offline recall increase of 5%+ for the random queries containing UoMs.

4   Conclusion

We have shown that through the use of KG and NER, we can improve UoM handling for query understanding and improve item ranking as a result. Given the genericness of the solution, our next step is to extend the application over domains such as sizes, colors and materials.

References

1. Markus D Steinberg, Sirko Schindler, and Jan Martin Keil. Use cases and suitability metrics for unit ontologies. In *OWL: Experiences and Directions*. 2016.
2. Foppiano et al. Automatic identification and normalisation of physical measurements in scientific literature. In *ACM Symposium on Document Engineering*, 2019.
3. Yingwei Xin, Ethan Hart, Vibhuti Mahajan, and Jean-David Ruvini. Learning better internal structure of words for sequence labeling. 2018.