

**Theme:** Machine Learning (algorithm)

**Topic:** Cross-lingual Knowledge Base Construction and Application in Global E-Commerce Scene

### **Background**

AliExpress is a global e-commerce app/website managed by Alibaba, which sells goods to over 200 countries and cover 18 kinds of languages. Complex usage and meaning drift of languages impels us to construct a multi-lingual e-commerce domain knowledge base, which currently used in product searching and shopping guide. We have two kinds of research topic below:

- Knowledge Base Construction

Different from traditional entity/relation extraction, E-commerce search queries and item titles are noisy, short and disordered, document or sentence based model doesn't work well in our area. Fortunately, we have abundant interaction corpus between queries and items collected from user behavior, so we are now trying to extract relations and properties of products from query session or query-item pair. As we know, we are the first team who conducts multi-lingual knowledge extraction process from interaction information.

- Knowledge Base Application

At the same time we are building knowledge base, we are also applying it to understand our queries and items, so we can represent queries and titles as a product-centered graph, integrate symbolic rules and product embedding into one model, transfer query-item matching problem as a graph matching procedure, improve searching experience.

For researcher, we have rich domain corpus come from E-commerce searching logs, and we offer real-world experiment environment. We hope our cooperation can contribute well-defined problems and innovatively algorithms .

### **Target**

- A well-defined E-Commerce knowledge extraction framework and a

innovatively cross-lingual mining algorithm.

- A Graph-based semantic matching model, using multi-lingual e-commerce knowledge base
- Submit 1-2 papers of IE/IR domain to top class academic meeting(CCFA)
- Release part of datasets of multi-lingual knowledge base

### **Related Research Topics**

- Weakly supervised/Semi-supervised Entity/Relation Extraction
- Product Knowledge Graph (Amazon)
- Deep semantic matching model (DSSM based)
- Deep Graph Matching
- Knowledge Augmented language model

### **Suggested Collaboration Method**

AIR (Alibaba Innovative Research), one-year collaboration project.