

# Project Proposal: Knowledge Graph of Perfume Commodities

Xiaoke Li 6488706306

Ziquan Cheng 9916764285

## Domain

We plan to build a ***perfume comparison and recommendation system*** using perfume knowledge graphs. The knowledge graph will be constructed by information including perfumes' brands, names, smells, selling platforms, volumes, prices, user feedbacks, ratings, delivery options, photos, etc. With this knowledge graph, we can figure out what the most popular perfumes are, link identical perfume items between different online shopping sources, as well as compare features like price, ratings, delivery options and some other features to give users recommendations based on users preference.

It is always tiring for people to compare different perfumes over different platforms with all of their different features. Thus, our KG can save people a lot of energy when they are struggling to decide which perfume to buy.

## Datasets

The knowledge graph will contain

- (unstructured) Data mined from general online shopping sources like Amazon, Ebay, Sephora, FragranceNet, and some brands' official websites.
- (structured) Tabular perfume data from Kaggle.

## Technical Challenges

We might meet following technical challenges:

- Issues with crawlers on the website.
- Issues with entity linking problems.
- Data insufficient or missing problem.
- System build and Frontend design.

Here's the detailed explanation:

The first possible technical challenge we will meet is the powerful anti-crawler when crawling data from online shopping websites. Some websites may have a login requirement, we have to save cookies or use some tricks to get the comments, rating, and some other detailed info.

Then the entity linking challenge. Some products may have very different titles and they are not likely to be linked with each other only through title comparison although they are the same entities, while others may have similar titles but are not likely to be the same entities.

Challenges of data insufficient and missing problems. Some attribute values like comments and ratings may be empty or insufficient on some sources, leading to the data missing problem.

System and Frontend design is another challenge. An excellent recommendation system tends to have high efficiency and accuracy. The method to send a query, fetch the result, and the way to better display the result in the front-end returned from the knowledge graph are all important to consider.