Project Proposal: Knowledge Graph of Perfume

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.

Our recommendation system will make recommend on these three dimensions:
- Random recommendation (based on comments/ratings)
- Lowest price recommendation (based on a particular item)
- Similar items recommendation (based on scent, price or brand)

Datasets
The knowledge graph will contain
- (unstructured) Data mined from general online shopping sources – Amazon, Sephora, FragranceNet.
- (structured) Tabular perfume data from Kaggle.

Otology
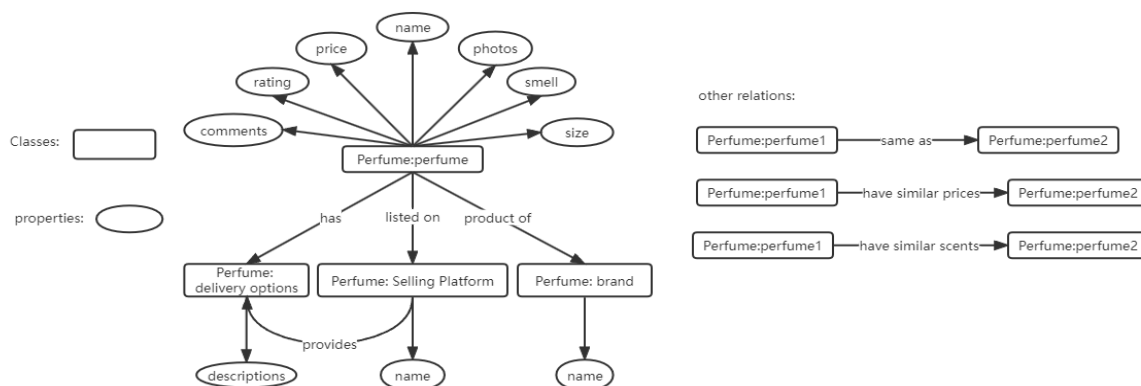
Evaluation

The target of evaluation of our model performance focuses on finding "same as" relations. We plan to use F2-score for evaluation, because we believe recall is twice more important than precision in this case.

Technical Challenges

We might meet following technical challenges:

- Entity Linking

  What is the best function/model to calculate similarity between 2 product titles. Some webpages may have very different string patterns in product titles but are displaying the same product, while other webpages may have very similar string patterns in titles but are not displaying the same product at all. We need to find a function/model that can best distinguish the same/different products through product titles.

- Sentiment Analysis and Tag Extraction

  In our first recommendation dimension, we leverage comments and ratings to determine what is the best perfume, thus we need to conduct sentiment analysis and tag extraction on comments in order to gain supplementation for ratings. It is not easy to select a good model to perform this task.

- Labeled Data Insufficiency

  Though it's easy for us to acquire perfume data from the web, there is no way for us to get labeled data for our sentiment analysis tasks. For the same reason, in evaluation, it will be hard to decide the model's performance of doing entity linking without labeled data.