PRODUCT NAME ENTITY RECOGNITION

Important Date

Deadline of submission will be 2 weeks after the candidates received the test contents.

Task Description

A big challenge for Shopee is to extract the product name entity from product title. For example, from a product title “Apple iPhone 6 (A1586) 16GB golden 4G”, we shall extract the core term “iPhone 6”, in order to serve the purpose of search, recommendation and keywords extraction etc.

Candidates are required to deliver a complete program. Given a title ***s*** of a product ***p***, the program should generate a core term ***t*** can fully represent the product ***p***

We will provide some sample data as follows:

Category	Product name	Core terms	Major Descriptive Terms	Brand
Mobile & Gadget	9.7 inch Onda V975i Tablet PC Screen Protector Film	Screen Protector Film	Tablet PC	Onda
Mobile & Gadget	BLUBOO Picasso Smartphone Silicone Protective Back Cover	Back Cover	Silicone Protective	BlUBOO Picasso
Mobile & Gadget	DoogeeX5/DoogeeX5 PRO Flip Cover Protective +Tempered Glass	Flip Cover	Tempered glass	Doogee X5
Mobile & Gadget	YI 4K Action Camera 2 Ambarella A9SE75 Sony IMX377 12MP	Camera	action	YI

And provide some testing product title. Your task will be writing a program to automatically extract the core terms out of the product titles.

Optional task: extract the Brand or major descriptive terms

Data

We provide a sample of product titles and their corresponding core terms. The sample can be used as case studies to further understand the problem. The sample is available in google drive.

Usage of external data sources is encouraged for training your model. However, internal training data can also be provided upon request. Please state clearly your intentions when making request.

The test set will be given in English. It contains 200 product titles (100 mobile & gadget and 100 Toys, Kids & Babies). The test data is also available in google drive.

Submission

Each submission has to contain the following items:

1. A result file contains the extracted core terms for the test set
2. A description of steps, data and algorithms that were used. Explanation of the reason behind the chosen solution is encouraged
3. Source codes and instruction
4. Submit through github is preferable

candidates are encouraged to use their programming languages, as long as they are comfortable using them. If candidates use any non-standard library, please include in the documents.

Similarly, candidates can implement new algorithms or re-implement existing algorithms. It is allowed to use third party software/service. However, their usage should be well documented.

Evaluation

Evaluation will be done by human judgement. Beside the result quality, the choice of algorithm, source code quality, solution explanation, proper documentation etc are all taken into consideration.