

Definitions of Entity Tags for Product Titles

This guide is for the task of evaluation and correction of NER tags in *product titles*. A product title is like a headline provided by an on-line seller that describes the product for sale. It contains various information about the product.

Before evaluating the NER tagging, annotators are encouraged to interpret the whole title and identify the *main product phrases*, and other elements that should take the other NER labels listed here. Here are the definitions of the various entities. Examples in Vietnamese are listed at the end of the page.

1. Product

Product refers to the minimal word or phrase that meaningfully refers to the product being sold in the product title. There may be one or more instances referring to the same product. This can be specific references or more general ways to refer to the product like in the following:

máy bơm lốp oto , xe hơi điện tử michelin 4399 ml cao cấp + tặng kèm bảng ghi số điện thoại gắn kính lái

Many product phrases have a very general first-word or first-syllable, like ‘máy’, ‘vải’, ‘giá’, ‘kệ’, ‘giá treo’, tủ, etc. which does have an individual meaning but are *not sufficient* to describe a product. As such, annotators should detect the *minimal span available* in the title - this would be in the form of an uninterrupted sequence of word(s) – that refers to the product. This is called the product phrase. It is possible that some titles contain only one general word that refers to the product, and no other adjacently available words that help to make the product more specific. If so, this ‘general’ word would still need to be tagged as a product entity.

For example, “plastic bag” when referring to common polythene household, non-fashion bags, should be tagged as one entity span: ‘túi nhựa’. Many more examples are listed at the end of this document.

All product phrases – no matter how general or specific the word is – for e.g. ‘tools’ vs. ‘monkey wrench’, need to be tagged as a Product entity. However, words that do not contribute to the product definition - like ‘beautiful’ in ‘*beautiful* titanium engagement ring’ – should not be included in the product phrase. Also note that some of these product ‘modifiers’ – extra information describing the product – could potentially be tagged with other NER tags (explained below). For example, ‘women’s winter coat’ – would correctly be labeled as :

women – Consumer_Group
winter – Season
coat – Product_Core

In the phrase ‘*beautiful* titanium engagement ring’, ‘engagement’ is also a potential Occasion, according to its definition.

In short, when determining product phrase, consider:

- a. what’s the minimal meaningful product span

- b. are any other words in the span necessary in the product phrase to determine it
- c. if the other words are not 'necessary', are they eligible to be tagged with other NER tags.
- d. Consider the consistency of product phrase tagging across different titles, that sell the same product.

2. Brand

Brand words that refer to the actual product's brand require this tag. Be careful not to tag brand words for supported products (e.g. 'leather case for *Canon* camera', where 'Canon' may not be the brand of the case). Also, to verify this tag, it is often necessary to look at the link of the product (or search online) what the actual brand is. This should not be confused with product's seller or online-store name. OEM should not be tagged with Brand.

3. Color

Some titles contain words or phrases that refer to the main color of the item sold. This can be a single word or a phrase. Note that color words are only for items that are possible to be in various colors (e.g. 'rust red'). If a color word is part of the product, e.g. 'red onion', in a title that sells get onion, 'red' here should not be a color but part of the product phrase. In English, 'purplish pink' is a single color; so for your respective language, use your best native judgement to evaluate the span of the color entity. Looking at the product page itself will usually help in the judgement.

4. Material and Fabric

- a. **Material** : These are words and phrases that refer to the major materials used to make the product. Ensure that the whole material elements are tagged in one span (e.g. 'aluminium alloy', 'hard acrylic', 'white pearl', 'stainless steel') if there are multiple words to the material. For materials that refer to types of cloth, the tag Fabric should be used. Often, referring to the product page itself is necessary.
- b. **Fabric**: This tag is used for titles having a fabric as a major material. Fabric refers to woven cloth, and includes examples like 'chiffon', 'nylon spandex'.

Note: When there is ambiguity tagging a word between fabric and material (assuming the material is 'cloth-like'), the Fabric tag should be used.

5. Consumer_Group

This refers to targeted users of the product. They can refer to phrases/words referring to intended users of a certain gender, age-group or age-range, social roles, or job holders.

6. Occasion and Season

- a. **Season**: Words and phrases referring to time-related situations, typically a yearly occurrence. This would include expressions referring to:
 - the 4 seasons of the year, which may also appear with the year number ('2019 Spring')

- numbers referring to the year ('2018 fashion')
- annual cultural and social festivals including – 'Xmas', 'Ramadan', 'Tết', 'spring festival', 'Valentine's', and 'father's day', 'thanksgiving'.

- b. **Occasion:** This refers to situations of use for the product which are less time-bound compared to Season. It includes place of use and type of activity. For example, 'outdoor use only', 'foldable pocket mirror for travel',

Note: When encountering ambiguity between Season and Ambiguity, assuming there is time-bound elements involved, the Season tag should be used.

7. Pattern

This tag is intended to be used for physical appearance of the whole product. This include shapes or cutting of the product, or parts of the product (e.g. collar types, fitting styles, layers/levels, fitting 'body hugging', 'middle hole'). Prints on the product (e.g. 'floral', 'doreamon', 'Kiss Me', 'smiling face', 'horizontal lines', 'embroidery edge', 'plaid',) which are on the surface of products are also tagged with this.

8. Style

Style refers to the *overall* visual effect of the product, particularly for products with more aesthetic values like apparel and decoration (rather than tools). These include elements like 'European pastoral', 'cute', 'sexy', 'Bohemian', 'feminine', 'boyfriend', 'school girl' etc. These refer to an *overall impression* which is less tangible, and do not refer to any specific physical design of the product. For example, 'tassel' referring to presence of actual tassels on the product would be classified as Pattern, and not Style.

Note: When facing expressions or terms unclear between Style and Occasion in your respective native language, the Occasion tag is advised to be used.

9. Place_of_Origin

These refer to proper names of cities, provinces, regions, or countries that the product comes from or is manufactured in. Note the difference between 'Chair made in Japan', and 'Japanese design stool from China'. The first being *Place_of_Origin*, while the latter a *Style*.

VIETNAMESE PRODUCT TITLE EXAMPLES

E.g 1: The example below sells "ốp lưng" as the product, not 'iphone 6'

Ốp lưng dẻo 3d hình thiên nga shengo cho iphone 6 (hồng)	
Text	Category
Ốp lưng	product
dẻo	pattern
thiên nga	pattern
shengo	brand
hồng	color

E.g 2: The example below sells “gối cổ” as the product, not ‘voucher’

gối cổ hình gấu panda + tặng phiếu tích điểm tmark + tặng voucher giảm giá tại khách sạn du lịch đà lạt giá rẻ thành tín trị giá 120.000	
Text	Category
gối cổ	product
gấu panda	pattern

NOTE: Giveaways and free products need not be tagged with NER elements.

E.g 3: The example below sells more than 1 product “bao da” and ‘kính cường lực glass’, but not ‘ipad air’

bộ bao da dành cho ipad air 1 kaku paris (xanh dương) và kính cường lực glass	
Text	Category
bao	product
da	material
kính cường lực glass	product
xanh dương	color
kaku	brand

E.g 4:

nước hoa nam calvin klein euphoria eau de toilette 100 ml	
Text	Category
nước hoa	product
nam	Consumer_Group
calvin klein	brand

E.g 5: The example below sells “máy bơm lốp” as the product, not ‘oto’ nor ‘xe hơi’

máy bơm lốp oto , xe hơi điện tử michelin 4399 ml cao cấp + tặng kèm bảng ghi số điện thoại gắn kính lái	
Text	Category
máy bơm lốp	product
michelin	brand

E.g 6: The example below sells “túi ngủ” as the product and rephrase it multiple times.

túi ngủ mùa đông cho bé , bộ đồ ngủ trưa văn phòng , túi ngủ cho bé , túi ngủ du lịch dã ngoại phượt - văn phòng , túi ngủ dày dặn cực ấm , bền nhẹ . nơi bán túi ngủ văn phòng giá rẻ , uy tín , chất lượng nhất - bh bởi ken99 mẫu 71	
Text	Category
túi ngủ	product
mùa đông	season
bé	Consumer_group
đồ ngủ	product
trưa	Occasion
văn phòng	Occasion
túi ngủ	product
bé	Consumer_group
túi ngủ	product
du lịch	Occasion
dã ngoại	Occasion
phượt	Occasion
văn phòng	Occasion
túi ngủ	product
dày dặn	pattern

E.g 7: The example below sells “giá treo điện thoại” as the product, not ‘điện thoại’ nor ‘ô tô’.

giá treo dien thoai tren giuong - giá treo điện thoại trên ô tô đẹp , chất lượng , giá rẻ	
Text	Category
gia treo dien thoai	product
giá treo điện thoại	product

E.g 8: The example below sells “chai tẩy gỉ sét và làm bóng đồ dùng inox cao cấp siêu mạnh số 1 nhật bản”. However, ‘inox’ is not the material for this product.

chai tẩy gỉ sét và làm bóng đồ dùng inox cao cấp siêu mạnh số 1 nhật bản	
Text	Category
chai tẩy gỉ sét	product
nhật bản	Place_of_origin

NOTE: In example 8, ‘cao cấp’ is *not* tagged as Style as the phrase refers to quality and not aesthetic value, as the product is a cleaning agent. Refer to definition of Style.

E.g 9: The example below sells “túi đựng”. However, ‘du lịch’ is not the occasion for this product.

túi đựng chăn màn quần áo - túi đựng quần áo du lịch - túi vải đựng đồ đa năng loại lớn nằm ngang sun plaza	
Text	Category
túi đựng chăn	Product
túi đựng quần áo	Product

túi	Product
vải	Fabric
loại lớn	pattern
nằm ngang	pattern
sun plaza	brand

E.g 10: The example below sells “bột đậu đen”. However, ‘đen’, ‘vàng’ are not the colors for this product.

bột đậu đen nông sản vàng 500gr	
Text	Category
bột đậu đen	product
nông sản vàng	brand

Data Format requirement

The format proposed is for the purpose of both evaluation and re-annotation/correction.

A row-by-row representation of the sentence will be provided along with the full title, category name, and product URL. The NER tag labels generated by the NER model will be shown – in the BIE format (standard suffixes to indicate NER spans).

Annotators are not to make the changes only to the NER tag labels in the second column. No lines should be deleted or changed. Changes can be done directly on the relevant lines of the second column.

Here is an English and a Vietnamese example:

(red indicates changed tags)

2212 Intuch Rage Against The Machine Ratm 1992 Band Logo Men'S T-Shirt Causal Short Sleeved Men'S T Shirt
Tee Black
category: Fashion
url: <https://www.lazada.com.ph/i265516187-s373035040.html>

Corrected Tagging:

Intuch
Rage
Against
The
Machine
Ratm
1992
Band
Logo
Men Consumer_Group_B
'S
T-Shirt Product_B
Causal Style_B
Short Pattern_B
Sleeved Pattern_E
Men Consumer_Group_B
'S

T Product_B
ShirtProduct_E
Tee Product_B
Black Color_B

In the above, 'T-Shirt' and 'Short Sleeved' were not tagged, while the sequence 'T Shirt Tee' was in one span (corrected into 2)

0001 Ốp lưng dẻo 3D hình thiên nga Shengo cho iPhone 6 Plus(Hồng)
category: Ốp lưng
url: <https://www.lazada.vn/products/op-lung-deo-3d-hinh-thien-nga-shengo-cho-iphone-6-plushong-i100902467-s101031840.html?spm=a2o4n.searchlist.list.1.702553a1W2z9Yg&search=1>

Corrected Tagging:

Ốp Product_B
Lưng Product_E
dẻo Pattern_B
3d
Hình
Thiên Pattern_B
nga Pattern_E
shengo Brand_B
cho
iphone
6
(
hồng Color_B
)

In this example, wrong tags for 'hình' and 'iphone' were removed.