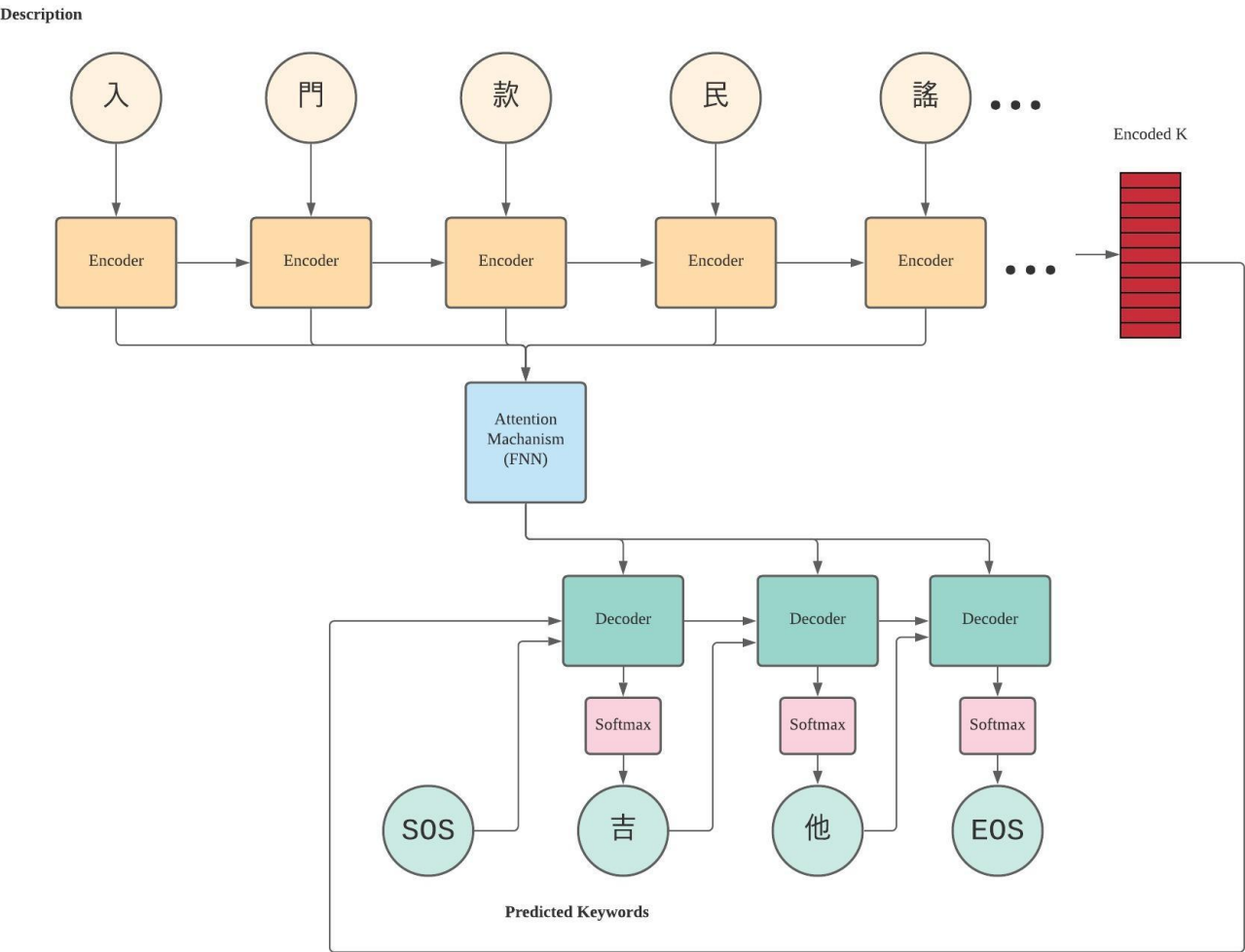# Keyword extraction

Model: Sequence to Sequence Neural Network

Concepts:

We use two RNNs, with one serving as an encoder and one as a decoder. First, we encode the description into a vector. Then, we use the result of the encoder and the last word(which is SOS at the beginning) to predict the next keyword.

Example:

Training data:

We use the customer-searched result as training data. If a customer searched a keyword and then clicks on a product, we bind the searched keyword and the description of the clicked product together. The model is then trained on 11 million rows of such data.

Producing multiple keywords:

We can produce multiple predicted keywords for each description by feeding the second and third most possible words for the first word at the start of the second and third keywords, respectively.

Result:

```
In [45]: evaluateRandomly(encoder1, attn_decoder1)
```

```
Input:   ALYA歐漾ALYA歐漾淨水ALYA歐漾 龍頭式五段過濾蔬果洗滌淨水器(含濾心1個) FF-5601
Target:  ▉
Prediction0:  淨心<EOS>
Probability0:  0.37 0.32 0.23
Prediction1:  濾芯<EOS>
Probability1:  0.14 0.47 0.94
Prediction2:  水果<EOS>
Probability2:  0.03 0.21 0.31

Input:   Mizuno陽光運動館(B1) MIZUNO 美津濃 WAVE RIDER 23 JUNIOR 大童鞋 女運動鞋 K1GC193316 [陽光樂活]
Target:  ▉
Prediction0:  mizuno wave 女<EOS>
Probability0:  0.24 0.54 0.22 0.57 0.16 0.39
Prediction1:  wave 女童鞋<EOS>
Probability1:  0.12 0.56 0.13 0.38 0.83 0.96
Prediction2:  女童鞋<EOS>
Probability2:  0.1 0.72 0.8 0.92

Input:   其他品牌松本清本舖【Braun 百靈】耳溫槍IRT6030，兩年保固，內附21個耳套，贈3D立體防塵口罩一盒(顏色隨機，送完為止)
Target:  ▉
Prediction0:  耳溫槍<EOS>
Probability0:  0.68 0.54 0.94 0.83
Prediction1:  百靈耳溫槍<EOS>
Probability1:  0.08 0.73 0.6 0.34 0.95 0.91
Prediction2:  口罩耳機<EOS>
Probability2:  0.02 0.6 0.47 0.23 0.44

Input:   其他品牌日韓小潼Simply 日本益生菌 30入/盒
Target:  ▉
Prediction0:  益生菌<EOS>
Probability0:  0.34 0.85 0.97 0.78
Prediction1:  日本<EOS>
Probability1:  0.21 0.56 0.31
Prediction2:  生日禮物<EOS>
Probability2:  0.04 0.78 0.48 0.41 0.95

Input:   其他品牌OSOME奧森鞋業Mconstar 月星 WagaMama系列 日本機能鞋 貴族點點 中童 運動鞋 I9640#藍色 奧森
Target:  ▉
Prediction0:  moonstar 童<EOS>
Probability0:  0.59 0.51 0.49 0.5
Prediction1:  童童鞋<EOS>
Probability1:  0.05 0.23 0.74 0.92
Prediction2:  小童鞋<EOS>
Probability2:  0.03 0.37 0.6 0.94
```

Reference:

https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html

https://arxiv.org/abs/1706.03762

Sample code:

Note: All the code here is re-contructed by me out of work-time and is not the one implemented in the company.

Link