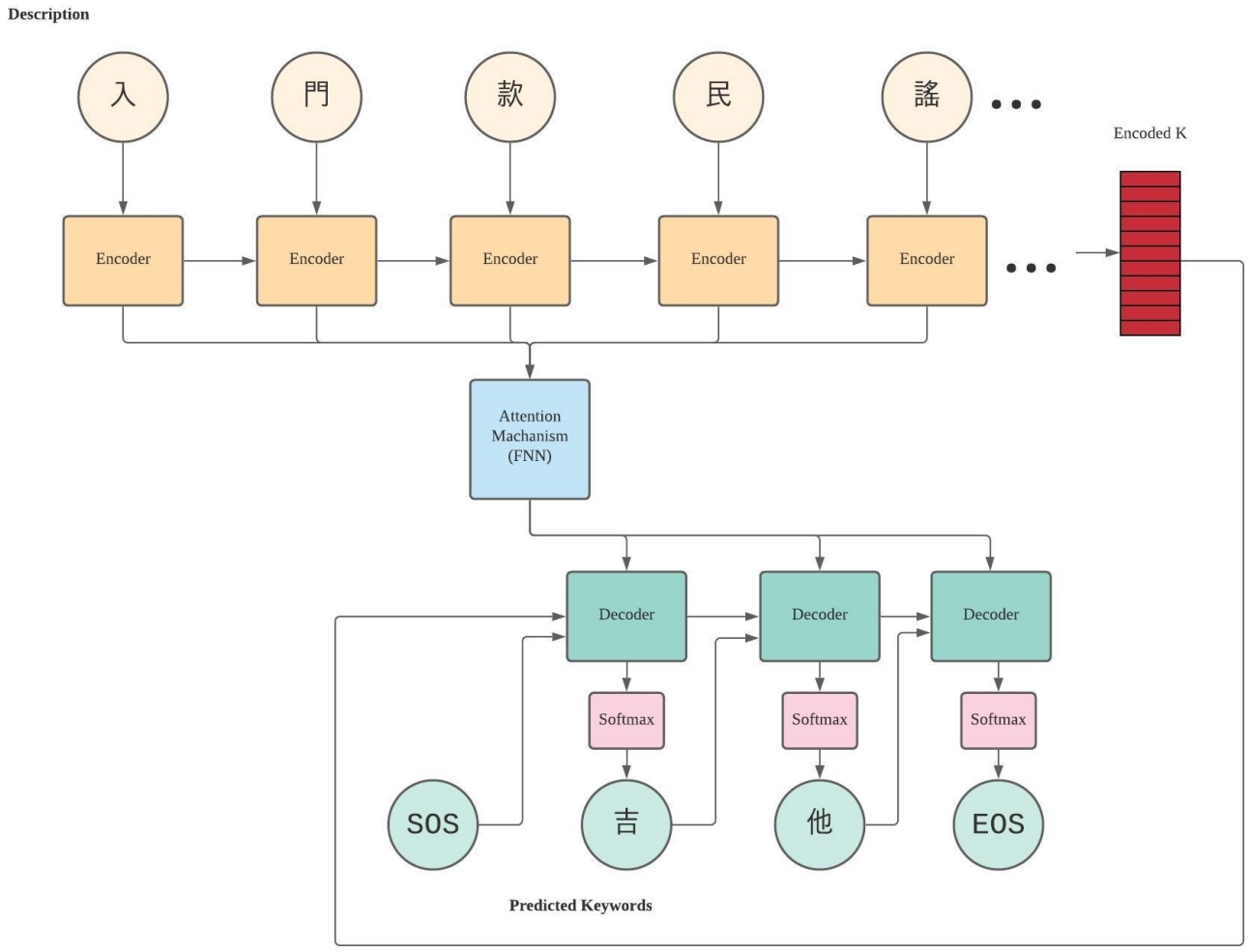
Keyword extraction

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.

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.