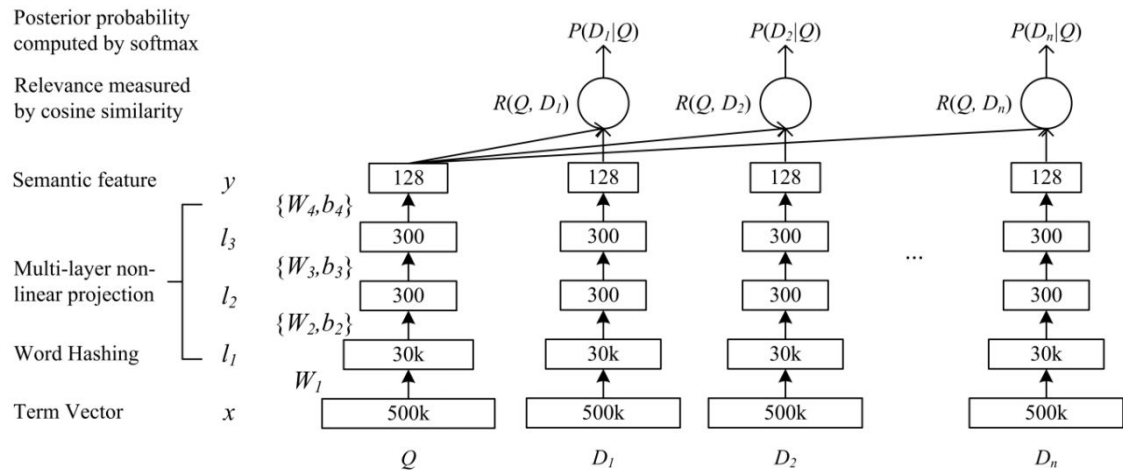


Summary



1. Bags of word for query and documents transform to the bags of n-grams – input vectors.
2. By the 1st layer l_1 the input vectors are convoluted.
3. $l_2 - l_3$ – fully-connected (? #полносвязные) layers.
4. Another conclusion to 128 dim for better training results.
5. Calculate the cosine similarity between query vector and each of documents.
6. Calculate the softmax output of the model: $P(D|Q) = \text{softmax}(yP)$, $y = \text{const.}$ (for all pairs Q : relevant documents + 4 unclicked documents)
7. Train the model by minimize $-\log[\prod_{\text{all pairs } Q:\text{relevant documents}} P(D|Q)]$.