

пусть у нас  $K$  слов в през.

$$\dim(\text{emb}) = n$$

$$\dim(\mathcal{E}, K, V) = m$$

}  $\Rightarrow$

$K$ -ые  $Q, K, V$  размера  $n \times m$

прелюды:

- 1) формируем эмб каждого слова на  $Q, K, V$
- 2) получаем  $K \times n \cdot n \cdot m = K \times m$  3 матрицы
- 3) перемножаем  $Q(K \cdot m) \cdot K^T(m \cdot K) = (K \times K)$  веса
- 4) делим  $ei$  на  $\sqrt{m}$  - для дилатации
- 5) берем на строки  $(K \times K)$  softmax
- 6) умножаем  $(K \times K) \cdot V(K \times m) = (K \times m)$  - получаем

Input

Embedding

Queries

Keys

Values

Score

Divide by  $8 (\sqrt{d_k})$

Softmax

Softmax  
X

Value

Sum

Image source: <https://jalammar.github.io/illustrated-transformer/>

Thinking

Machines

$x_1$

$x_2$

$q_1$

$q_2$

$k_1$

$k_2$

$v_1$

$v_2$

$q_1 \cdot k_1 = 112$

$q_1 \cdot k_2 = 96$

14

12

0.88

0.12

$v_1$

$v_2$

$z_1$

$z_2$

**STEP 1:** create Query, Key, Value

**STEP 2:** calculate scores

**STEP 3:** divide by  $\sqrt{d_k}$

**STEP 4:** softmax

**STEP 5:** multiply each value vector by the softmax score

**STEP 6:** sum up the weighted value vectors

Self-Attention