

# Leitura do artigo

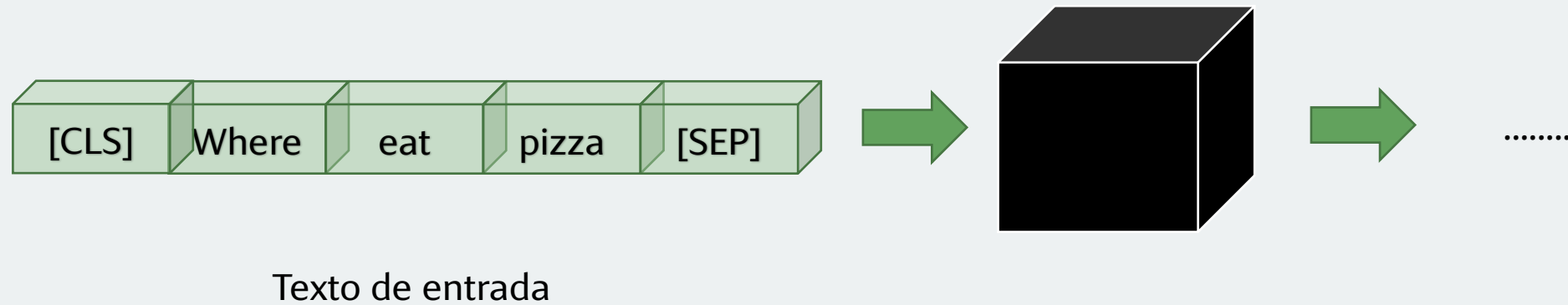
**SPLADE: Sparse Lexical and Expansion  
Model for First Stage Ranking**  
*(Thibault Formal et al)*

# Conceitos importantes

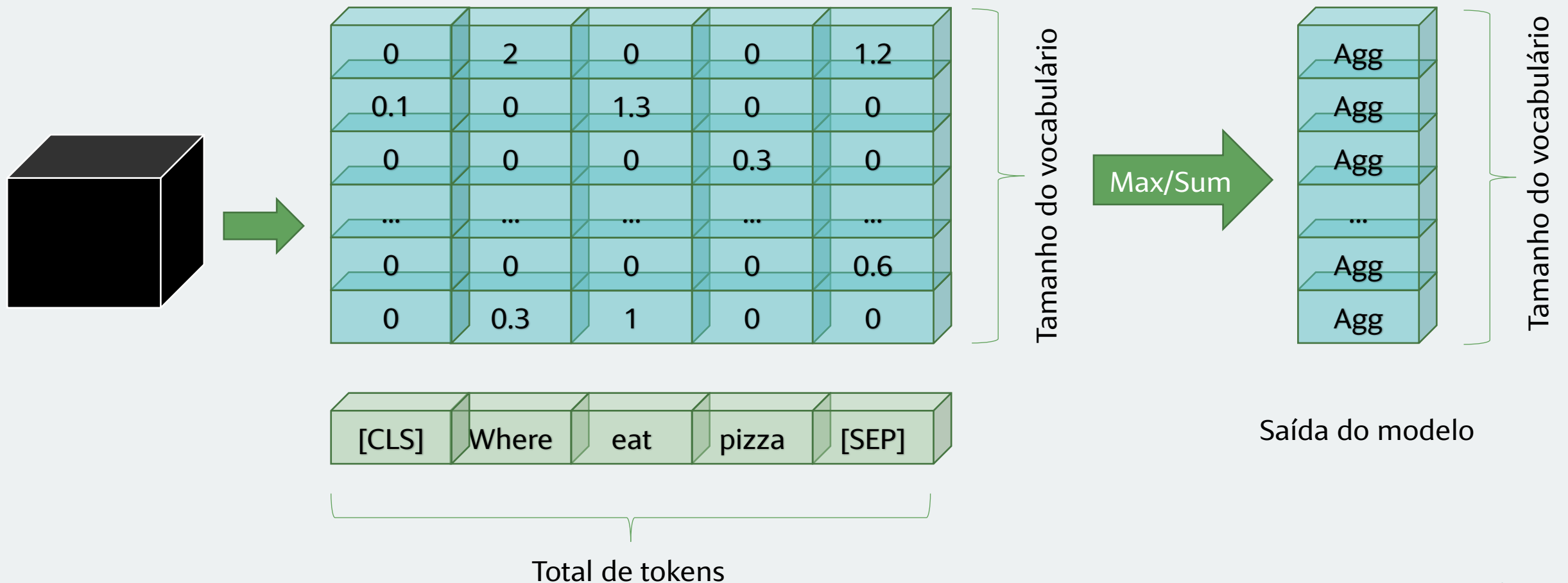
## *Representação esparsa:*

Tokens são representados por um vetor de tamanho  $n$ , onde a maior parte dos elementos é zero. No caso do artigo, os elementos diferentes de 0 apontam para palavras do vocabulário.

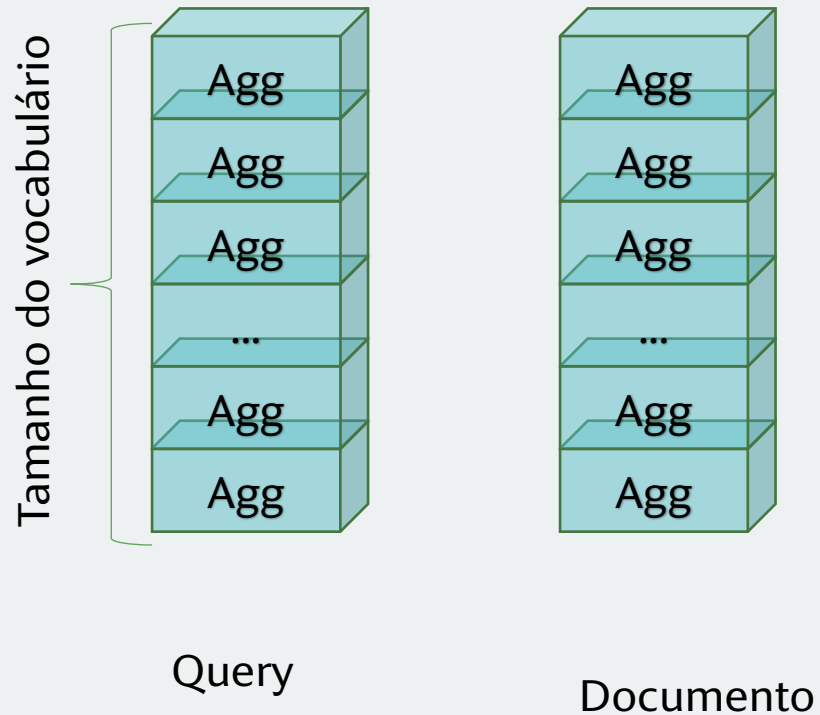
# Conceitos/contribuição do artigo



# Conceitos/contribuição do artigo



# Conceitos/contribuição do artigo



$\text{Score}(d, q) = \text{produto escalar entre } d \text{ e } q$

No SpladeV2 os elementos do vetor da query são  $\{0, 1\}$ .

- ⇒ Se o corpus for pequeno, dá pra guardar a matriz de documento e resolver com uma multiplicação de matrizes.
- ⇒ Se o corpus for grande, dá pra resolver com um índice invertido guardando como chave o token (ou token\_id) e a lista de documento e score do documento

# Dúvidas básicas

O artigo apresenta resultados bons para MS MARCO e TREC DL 2019. Mas os testes feitos no exercício com SUM no TREC COVID mostraram resultados ruins. Porque?

# Obrigado

Leandro Carísio  
carisio@gmail.com