

Texto

O gato comeu o rato!

Padronização

Texto padronizado

o gato comeu o rato

Tokenização

Tokens

“o”, “gato”, “comeu”, “o”, “rato”

Indexação

Índice dos tokens

3, 27, 42, 3, 39

Embedding

**Representação
vetorial dos tokens**

1

0

0

1

0

0

0

1

0

0

0

1

0

0

0

0

0

0

0

0

0

0

0

0

1