

### 3\_NLP\_token

December 26, 2024

```
[1]: text = "Just in a few 20,00 years, natural language processing has evolved into  
↳something so powerful and impactful, which no one could have imagined."
```

```
[5]: #white space Tokenizer  
print(text.split())
```

```
['Just', 'in', 'a', 'few', '20,00', 'years,', 'natural', 'language',  
'processing', 'has', 'evolved', 'into', 'something', 'so', 'powerful', 'and',  
'impactful,', 'which', 'no', 'one', 'could', 'have', 'imagined.']
```

```
[8]: #Loading model  
import spacy
```

```
[11]: nlp = spacy.load('en_core_web_sm')
```

```
[12]: doc = nlp(text)
```

```
[13]: doc #document, not token
```

```
[13]: Just in a few 20,00 years, natural language processing has evolved into  
something so powerful and impactful, which no one could have imagined.
```

```
[14]: type(doc)
```

```
[14]: spacy.tokens.doc.Doc
```

```
[15]: #spacy Tokenizer (token as object no string)  
print([token for token in doc])
```

```
[Just, in, a, few, 20,00, years, ,, natural, language, processing, has, evolved,  
into, something, so, powerful, and, impactful, ,, which, no, one, could, have,  
imagined, .]
```

```
[16]: #Type of token  
[type (token) for token in doc]
```

```
[16]: [spacy.tokens.token.Token,  
spacy.tokens.token.Token,  
spacy.tokens.token.Token,
```

[illegible]

```
[17]: #Type of token
print([token.text for token in doc])
```

```
['Just', 'in', 'a', 'few', '20,00', 'years', ',', 'natural', 'language',  
'processing', 'has', 'evolved', 'into', 'something', 'so', 'powerful', 'and',  
'impactful', ',', 'which', 'no', 'one', 'could', 'have', 'imagined', '.']
```

[ ]:

```
[18]: #Compare
      #white space Tokenizer
      print(text.split())
      #Sapcy token
      print([token.text for token in doc])
```

```
[['Just', 'in', 'a', 'few', '20,00', 'years,', 'natural', 'language',  
'processing', 'has', 'evolved', 'into', 'something', 'so', 'powerful', 'and',  
'impactful,', 'which', 'no', 'one', 'could', 'have', 'imagined.']]  
[['Just', 'in', 'a', 'few', '20,00', 'years', ',', ',', 'natural', 'language',  
'processing', 'has', 'evolved', 'into', 'something', 'so', 'powerful', 'and',  
'impactful', ',', ',', 'which', 'no', 'one', 'could', 'have', 'imagined', '.']]
```

[ ]: