

Text Normalization Challenge

English Language

Amir Shabani
9413088021

Description

- What is Text Normalization?
- Where can it be used?

A baby giraffe is 6ft tall and weighs 150lb.

A baby giraffe is six feet tall and weighs one hundred fifty pounds.

A diagram illustrating text normalization. Two blue arrows point from the top sentence to the bottom sentence. The first arrow points from '6ft' to 'six feet', and the second arrow points from '150lb' to 'one hundred fifty pounds'.

Train Data

```
In [1]: import pandas
```

```
In [2]: train = pandas.read_csv('en_train.csv')
```

```
In [3]: train.shape
```

```
Out[3]: (9918441, 5)
```

```
In [4]: train.head()
```

```
Out[4]:
```

	sentence_id	token_id	class	before	after
0	0	0	PLAIN	Brillantaisia	Brillantaisia
1	0	1	PLAIN	is	is
2	0	2	PLAIN	a	a
3	0	3	PLAIN	genus	genus
4	0	4	PLAIN	of	of

Test Data

```
In [1]: import pandas
```

```
In [2]: test = pandas.read_csv('en_test.csv')
```

```
In [3]: test.shape
```

```
Out[3]: (1088564, 3)
```

```
In [4]: test.head()
```

```
Out[4]:
```

	sentence_id	token_id	before
0	0	0	Another
1	0	1	religious
2	0	2	family
3	0	3	is
4	0	4	of

Evaluation

```
In [3]: train.loc[train['sentence_id'] == 44]
```

Thus, during the years 1573-1595 they rebuilt the old castle.



fifteen seventy three

fifteen ninety five

id,after

...

44_5, "fifteen seventy three"

44_6, "to"

44_7, "fifteen ninety five"

...

Prizes

Three cash prizes will be awarded to the top teams:

- 1st place - \$12,000
- 2nd place - \$8,000
- 3rd place - \$5,000