Facts and Figures of Spacy model

Comparison between time and accuracy

Two peer-reviewed papers in 2015 confirmed that spaCy offers the **fastest syntactic parser in the world** and that **its accuracy is within 1% of the best** available. The few systems that are more accurate are 20× slower or more.

SYSTEM	YEAR	LANGUAGE	ACCURACY	SPEED (WPS)
spaCy v2.x	2017	Python / Cython	**92.6**	*n/a* ?
spaCy v1.x	2015	Python / Cython	91.8	13,963
ClearNLP	2015	Java	91.7	10,271
CoreNLP	2015	Java	89.6	8,602
MATE	2015	Java	92.5	550
Turbo	2015	C++	92.4	349

Parse Accuracy

Parsing in NLP is the process of determining the syntactic structure of a text by analyzing its constituent words based on an underlying grammar (of the language).

SYSTEM	YEAR	TYPE	ACCURACY
spaCy v2.0.0	2017	neural	94.48
spaCy v1.1.0	2016	linear	92.80

Detailed Speed Comparison

Here we compare the per-document processing time of various spaCy functionalities against other NLP libraries. We show both absolute timings (in ms) and relative performance (normalized to spaCy). Lower is better.

	ABSOLUTE (MS PER DOC)			RELATIVE (TO SPACY)		
SYSTEM	TOKENIZE	TAG	PARSE	TOKENIZE	TAG	PARSE
spaCy	0.2ms	1ms	19ms	1x	1x	1x
CoreNLP	0.18ms	10ms	49ms	0.9x	10x	2.6x
ZPar	1ms	8ms	850ms	5x	8x	44.7x
NLTK	4ms	443ms	*n/a*	20x	443x	*n/a*

Reference

https://spacy.io/usage/facts-figures