lvl	container	tag	opt	description	type	source
10	document	document		container for document	dict	TET
				created at (DD.MM.YY		
11	document	createdAt		HH:MM:SS)	str	system
11	document	createdBy		created by module	str	system
		-		last modified at (DD.MM.YY		
11	document	modifiedAt	х	HH:MM:SS)	str	system
11	document	modifiedBy	Х	last modified by module	str	system
11	document	noFonts	Х	number fonts	int	TET
11	document	noLines		number lines in document	int	TET
11	document	noLinesFooter	Х	number footer lines	int	LT
11	document	noLinesHeader	Х	number header lines	int	LT
11	document	noLinesToc	Х	number toc lines	int	LT
11	document	noListsBullet	Х	number bulleted lists	int	LT
11	document	noListsNumber	Х	number numbered lists	int	LT
	document	noPages		number pages in document	int	TET
		-				
11	document	noParas		number paragraphs in document	int	TET
11	document	noSentences	х	number sentences in document	int	spaCy
11	document	noTables		number tables in document	int	LT
11	document	noWords		number words in document	int	TET
19	document	config	Х	container for configuration	dict	TET
19	document	fonts	Х	container for fonts	list	TET
19	document	headings		container for headings	list	TET
19	document	listsBullet		container for bulleted lists	list	TET
19	document	listsNumber		container for numeredd lists	list	TET
19	document	pages		container for pages	list	TET
19	document	params	Х	container for parameters	dict	TET
20	config	config	Х	container for configuration	dict	TET
21	config	parser		container for module parser	dict	TT
				-		
21	config	tokenizer		container for module tokenizer	dict	TET
				include the configuration data in		
22	config	jsonInclConfig		the JSON file	bool	TET
				include the font data in the JSON		
22	config	jsonInclFonts		file	bool	TET
				include the heading data in the		
22	config	jsonInclHeading		JSON file	bool	TET
				include the bulleted list data in		
22	config	jsonInclListBullet		the JSON file	bool	TET
				include the numbered list data in		
22	config	jsonInclListNumber		the JSON file	bool	TET
				include the parameters in the		
22	config	jsonInclParams		JSON file	bool	TET
				Improves the readability of the		
22	config	jsonIndent		JSON file	int	TET
22	config	jsonSortKeys		Sort the keys in ascending order	bool	TET

Version: 0.9.8 Page: 1/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
				maximum Levenshtein distance		
22	config	ltFooterMaxDistance		for a footer line	int	TET
22	config	ltFooterMaxLines		maximum number of footers	int	TET
				Maximum Levenshtein distance		
22	config	ltHeaderMaxDistance		for a header line	int	TET
22	config	ltHeaderMaxLines		Maximum number of headers	int	TET
				number of lines following the		
				heading to be included as		
22	config	ltHeadingFileInclNoCtx		context into the JSON file	bool	TET
				number of lines following the		
		ltHeadingFileInclRegex		heading to be included as		
22	config	р		context into the JSON file	bool	TET
				maximum level of the heading		
22	config	ltHeadingMaxLevel		structure	int	TET
				minimum number of pages to		
22	config	ltHeadingMinPages		determine the headings	int	TET
	_			file with rules to determine the		
22	config	ltHeadingRuleFile		headings	bool	TET
				tolerance of vertical indentation		
22	config	ltHeadingToleranceLlx		in percent	int	TET
	6.			minimum number of entries to		
22	config	ltListBulletMinEntries		determine a bulleted list	int	TET
	ç.			file with rules to determine the	l	
	config	ltListBulletRuleFile		bulleted lists tolerance of vertical indentation	bool	TET
	t:-	ltListBulletToleranceL			haal	TET
22	config	lx		in percent	bool	TET
				if it is set to true the regular		
				if it is set to true, the regular expression for the numbered list		
22	config	ltListNumberFileInclRe		is included in the JSON file	bool	TET
22	Coming	gexp		is included in the 330N file	DOOI	ILI
				minimum number of entries to		
22	config	ltListNumberMinEntries		determine a numbered list	bool	TET
	comig	Tellistramperminentites		file with rules to determine the	5001	1.51
22	config	ltListNumberRuleFile		numbered lists	bool	TET
	comig	ltListNumberToleranceL		tolerance of vertical indentation	5001	1.51
22	config	1x		in percent	bool	TET
				F		
				if it is set to true, the empty ells		
		   ltTableFileInclEmptyCo		are included in the separate		
22	config	lumns		JSON file with the tables	bool	TET
				maximum number of pages for		
				the search of the TOC (from the		
22	config	ltTocLastPage		beginning)	bool	TET
		_		-		
22	config	ltTocMinEntries		minimum number of TOC entries	bool	TET
	-			ignore the tokens which are		
22	config	spacyIgnoreBracket		brackets?	bool	TET

Version: 0.9.8 Page: 2/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
				ignore the tokens which are left		
22	config	spacyIgnoreLeftPunct		punctuation marks, e.g. "("?	bool	TET
		spacyIgnoreLineTypeFoo		ignore the tokens from line type		
22	config	ter		footer ?	bool	TET
		spacyIgnoreLineTypeHea		ignore the tokens from line type		
22	config	der		header ?	bool	TET
		spacyIgnoreLineTypeHea		ignore the tokens from line type		
22	config	ding		heading ?	bool	TET
		spacyIgnoreLineTypeLis		ignore the tokens from line type		
22	config	tBullet		bulleted list ?	bool	TET
		spacyIgnoreLineTypeLis		ignore the tokens from line type		
22	config	tNumber		numbered list ?	bool	TET
		spacyIgnoreLineTypeTab		ignore the tokens from line type		
22	config	le		table ?	bool	TET
				ignore the tokens from line type		
22	config	spacyIgnoreLineTypeToc		TOC ?	bool	TET
				ignore the tokens which are		
22	config	spacyIgnorePunct		punctuations ?	bool	TET
				ignore the tokens which are		
22	config	spacyIgnoreQuote		quotation marks ?	bool	TET
				ignore the tokens which are		
				right punctuation marks, e.g. ")"		
22	config	spacyIgnoreRightPunct		?	bool	TET
				ignore the tokens which consist		
22	config	spacyIgnoreSpace		of whitespace characters ?	bool	TET
				ignore the tokens which are part		
22	config	spacyIgnoreStop		of a "stop list" ?	bool	TET
22	config	spacyTknAttrCluster		brown cluster ID	bool	TET
22	config	spacyTknAttrDep_		syntactic dependency relation	bool	TET
22	config	spacyTknAttrDoc		the parent document	bool	TET
22	config	spacyTknAttrEntIob_		IOB code of named entity tag	bool	TET
				knowledge base ID that refers to		
				the named entity this token is a		
22	config	spacyTknAttrEntKbId_		part of, if any	bool	TET
22	config	spacyTknAttrEntType_		named entity type	bool	TET
				the syntactic parent, or		
22	config	spacyTknAttrHead		"governor", of this token	bool	TET
				the index of the token within the		
22	config	spacyTknAttrI		parent document	bool	TET
				the character offset of the token		
22	config	spacyTknAttrIdx		within the parent document	bool	TET
				does the token consist of		
22	config	spacyTknAttrIsAlpha		alphabetic characters?	bool	TET

Version: 0.9.8 Page: 3/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
				does the token consist of ASCii		
22	config	spacyTknAttrIsAscii		characters?	bool	TET
22	config	spacyTknAttrIsBracket		is the token a bracket?	bool	TET
22	config	spacyTknAttrIsCurrency		is the token a currency symbol?	bool	TET
22	config	spacyTknAttrIsDigit		does the token consist of digits?	bool	TET
		spacyTknAttrIsLeftPunc		is the token a left punctuation		
	config	t		mark, e.g. "(" ?	bool	TET
22	config	spacyTknAttrIsLower		is the token in lowercase?	bool	TET
22	config	212 0 21 1 1 1 1 2 1 2 1 2 1 2 1 2 1 2 1		is the taken out of vecabulary?	haal	TET
	config	spacyTknAttrIsOov		is the token out-of-vocabulary?	bool	TET
22	config	spacyTknAttrIsPunct		is the token punctuation?	bool	161
22	config	spacyTknAttrIsQuote		is the token a quotation mark?	bool	TET
		spacyTknAttrIsRightPun		is the token a right punctuation		
22	config	ct		mark, e.g. ")" ?	bool	TET
22	config	spacyTknAttrIsSentEnd		does the token end a sentence?	bool	TET
22	config	spacyTknAttrIsSentStar t		does the token start a sentence?	bool	TET
				does the token consist of		
22	config	spacyTknAttrIsSpace		whitespace characters?	bool	TET
22	config	spacyTknAttrIsStop		is the token part of a "stop list"?	bool	TET
22	config	spacyTknAttrIsTitle		is the token in titlecase?	bool	TET
22	config	spacyTknAttrIsUpper		is the token in uppercase?	bool	TET
22	config	spacyTknAttrLang_		language of the parent document's vocabulary	bool	TET
	config	spacyTknAttrLeftEdge		the leftmost token of this token's syntactic descendants base form of the token, with no	bool	TET
22	config	spacyTknAttrLemma_		inflectional suffixes	bool	TET
22	config	spacyTknAttrLex		the underlying lexeme	bool	TET
				sequential ID of the token's lexical type, used to index into		
22	config	spacyTknAttrLexId		tables, e.g. for word vectors	bool	TET
22	confi-			does the token resemble an	hes!	
22	config	spacyTknAttrLikeEmail		email address?	bool	TET
าา	config	an a grafflan A + + x I d l- a Norm		does the token represent a	haal	    TET
	config	spacyTknAttrLikeNum		number?	bool	TET
22	config	spacyTknAttrLikeUrl		does the token resemble a URL?	bool	TET
	config	spacyTknAttrLower_		lowercase form of the token text		TET
22	config	spacyTknAttrMorph		morphological analysis	bool	TET

Version: 0.9.8 Page: 4/11 Date: 27.09.2022

22 c	config			the token's norm, i.e. a		
22 c						
22 c				normalized form of the token		
22 c	config	spacyTknAttrNorm_		text	bool	TET
	6	spacyTknAttrOrth_		verbatim text content	bool	TET
				coarse-grained part-of-speech		
22 c	config	spacyTknAttrPos_		from the Universal POS tag set	bool	TET
22 c				a length-N substring from the		
	config	spacyTknAttrPrefix_		start of the token	bool	TET
	C:	_, _, _,		smoothed log probability		
22 c	config	spacyTknAttrProb		estimate of token's word type	bool	TET
				sequential ID of the taken's		
				sequential ID of the token's lexical type, used to index into		
22.0	config	spacyTknAttrRank			hool	TET
22 (	Joining	spacyTRHACCTRAHK		tables, e.g. for word vectors	bool	161
				the rightmost token of this		
22 c	config	  spacyTknAttrRightEdge		token's syntactic descendants	bool	TET
	2011118	bpacy immicelitightenage		the sentence span that this token	5001	
22 c	config	spacyTknAttrSent		is a part of	bool	TET
	, o			a scalar value indicating the	500.	1.2.
				positivity or negativity of the		
22 c	config	spacyTknAttrSentiment		token	bool	TET
						1
				transform of the token's string to		
22 c	config	spacyTknAttrShape_		show orthographic features	bool	TET
				length-N substring from the end		
22 c	config	spacyTknAttrSuffix_		of the token	bool	TET
	config	spacyTknAttrTag_		fine-grained part-of-speech	bool	TET
				the token's slice of the parent		
22 c	config	spacyTknAttrTensor		doc's tensor	bool	TET
22 c	config	spacyTknAttrText		verbatim text content	bool	TET
				text content, with trailing space		
22 c	config	spacyTknAttrTextWithWs		character if present	bool	TET
				the vocab object of the parent		
22 c	config	spacyTknAttrVocab		doc	bool	TET
		spacyTknAttrWhitespace				
-	config	_		trailing space character if present		TET
30 fc		font	X	container for font	dict	TET
31 fo		embedded		embedded	bool	TET
31 fo		fontNo		font number in document font full name	int	TET
31 fc		fullName id		font full name font identification	str	TET
31 fc		italicAngle		font italic angle	str float	TET
31 fc		name		font name	str	TET
31 fc		type		font type	str	TET
31 fc		weight		font weight	float	TET
1	neading	heading	Х	container for headings	dict	LT

Version: 0.9.8 Page: 5/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
41	entry	ctxtLine1		1. context line	str	LT
41	entry	ctxtLine2		2. context line	str	LT
41	entry	ctxtLine3		3. context line	str	LT
41	heading	toc	х	container for toc	dict	LT
42	heading	entries		container for heading entries	list	LT
43	entry	entry		container for heading entry	dict	LT
43	entry	level		heading level	int	LT
43	entry	lineNoPage		line number in page	int	LT
43	entry	pageNo		page number in document	int	LT
	entry	regexp	Х	regular expression	str	LT
	entry	text		heading text	str	LT
50		bulleted list	Х	container for bulleted list	dict	LT
51	list	format		bullet format	str	LT
51	list	listNo		list number in document	int	LT
	list	noEntries		number entries in list	int	LT
51		pageNoFirst		page number first entry	int	LT
51		pageNoLast		page number last entry	int	LT
52		entries		container for list entries	list	LT
						1
53	entry	entry		container for bulleted list entry	dict	LT
	entry	entryNo		entry number in list	int	LT
	entry	lineNoPageFirst		first line in page	int	LT
	entry	lineNoPageLast		last line in page	int	LT
	entry	pageNo		page number in document	int	LT
	entry	paraNo		paragraph number in page	int	LT
	entry	text		list entry text	str	LT
60		numbered list	Х	container for numbered list	dict	LT
61		format		bullet format	str	LT
61		listNo		list number in document	int	LT
61		noEntries		number entries in list	int	LT
61		pageNoFirst		page number first entry	int	LT
61		pageNoLast		page number last entry	int	LT
61		regexp	х	regular expression	str	LT
62		entries		container for list entries	list	LT
-52						+
63	entry	entry		container for numbered list entry	dict	LT
	entry	entryNo		entry number in list	int	LT
	entry	lineNoPageFirst		first line in page	int	LT
	entry	lineNoPageLast		last line in page	int	LT
	entry	pageNo		page number in document	int	LT
	entry	paraNo		paragraph number in page	int	LT
	entry	text		list entry text	str	LT
	page	page		container for page	dict	TET
- 1	page page	lineNoFirst		number of first line in page	int	TET
	page page	lineNoLast		number of last line in page	int	TET
				· -		TET
	hage	Pageno			IIIL	161
71	nago	naraNoFirst			int	TET
71	page page		pageNo paraNoFirst	pageNo	pageNo page number in document number of first paragraph in	pageNo page number in document int number of first paragraph in

Version: 0.9.8 Page: 6/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
7.4						
/1	page	paraNoLast		number of last paragraph in page	int	TET
71	page	sentenceNoFirst		number of first sentence in page	int	spaCy
	Page			named of mot sentence in page		эрасу
71	page	sentenceNoLast		number of last sentence in page	int	spaCy
	page	wordNoFirst		number of first word in page	int	TET
71	page	wordNoLast		number of last word in page	int	TET
79	page	lines		container for lines	list	TET
79	page	paras		container for paragraphs	list	TET
79	page	sentences		container for sentences	list	spaCy
80	params	params	Х	container for parameters	dict	TET
81	params	parser		container for module parser	dict	TT
				file directory for intermediate		
	params	directoryName		files and final result files	str	TET
	params	documentId	Х	document identification	int	TET
	params	environmentVariant		environment variant	str	TET
	params	fileNameCurr		name of input file	str	TET
	params	fileNameNext		name of output file	str	TET
82	params	fileNameOrig		original file name	str	TET
00		lette e dies er De envisee d		handing data was in ation was wined	h 1	тст
82	params	ltHeadingRequired		heading determination required bulleted list determination	bool	TET
0.2		]+I			bool	тст
82	params	ltListBulletRequired		required numbered list determination	bool	TET
00					1 1	тст
82	params	ltListNumberRequired		required	bool	TET
0.7	narama	1+mahlana mainad		table determination required	bool	TET
	params params	ltTableRequired ltTocRequired		TOC determination required	bool	TET
02	parairis	itiockequired		Toc determination required	DOOI	IEI
83	params	tokenizer		container for module tokenizer	dict	TET
	line	line		container for lines	dict	TET
101		lineNo		line number in document	int	TET
101		lineNoPage		line number in page	int	TET
101		lineNoPara		line number in paragraph	int	TET
		TITICILOT GEG		x coordinate of the lower left		1.2.
101	line	llx		corner	float	TET
101		pageNo		page number in document	int	TET
101	line	paraNo		paragraph number in document	int	TET
101	line	paraNoPage		paragraph number in page	int	TET
101	line	sentenceNo		sentence number in document	int	spaCy
101	line	tableCellNo		cell number in row	int	TET
	line	tableCellSpan		cel span	int	TET
101	line	tableNo		table number in document	int	TET
101		tableRowNo		row number in table	int	TET
	line	text		line text	str	TET

Version: 0.9.8 Page: 7/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
101	line	type		line type	str	TET, TL
				x coordinate of the upper right		
101	line	urx		corner	float	TET
101	line	wordNoFirst		number word of first word in line	int	TET
101	line	wordNoLast		number word of last word in line	int	TET
				word number in paragraph of		
101	line	wordNoParaFirst		first word in this line	int	TET
110	para	para		container for paragraph	dict	TET
111	para	lineNoFirst		number of first line in paragraph	int	TET
111	para	lineNoLast		number of last line in paragraph	int	TET
111	para	pageNo		page number in document	int	TET
111	para	paraNo		paragraph number in document	int	TET
111	para	paraNoPage		paragraph number in page	int	TET
				number of first sentence in		
111	para	sentenceNoFirst		paragraph	int	spaCy
				number of last sentence in		
111	para	sentenceNoLast		paragraph	int	spaCy
111	para	tableCellNo		cell number in row	int	TET
111	para	tableCellSpan		cel span	int	TET
111	para	tableNo		table number in document	int	TET
111	para	tableRowNo		row number in table	int	TET
						TET,
111	para	text		paragraph text	str	spaCy
				number of first word in		
111	para	wordNoFirst		paragraph	int	TET
				number of last word in		
111	para	wordNoLast		paragraph	int	TET
111	para	words		container for words	list	TET
120	sentence	sentence		container for sentences	dict	TET
121	sentence	sentenceNo		sentence number in document	int	TET
121	sentence	sentenceNoPage		sentence number in page	int	TET
121	sentence	sentenceNoPara		sentence number in paragraph	int	TET
121	sentence	text		sentence text	str	TET
				number word of first word in		
121	sentence	wordNoFirst		sentence	int	TET
				number word of last word in		
	sentence	wordNoLast		sentence	int	TET
	word	word	ı	container for word	dict	TET
	word	font	Х	font identification	str	TET
	word	lineNo		line number in document	int	TET
201	word	lineNoPage		line number in page	int	TET

Version: 0.9.8 Page: 8/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
				x coordinate of the lower left		
201	word	llx		corner	float	TET
201	word	pageNo		page number in document	int	TET
201	word	paraNo		paragraph number in document	int	TET
201	word	sentenceNo		sentence number in document	int	spaCy
	word	size	х	font size	float	TET
201	word	spacyTknAttrCluster		brown cluster ID	str	spaCy
201	word	spacyTknAttrDep_		syntactic dependency relation	str	spaCy
	word	spacyTknAttrDoc		the parent document	str	spaCy
	word			IOB code of named entity tag		<u> </u>
201	word	spacyTknAttrEntIob_		, ,	str	spaCy
				knowledge base ID that refers to		
204				the named entity this token is a		
201	word	spacyTknAttrEntKbId_		part of, if any	str	spaCy
201	word	spacyTknAttrEntType_		named entity type	str	spaCy
		2F467 :::::::002 2::017F6_		the syntactic parent, or		opu.c,
201	word	spacyTknAttrHead		"governor", of this token	str	spaCy
201	Word	Spacy ininectificad		the index of the token within the	361	эрасу
201	word	spacyTknAttrI		parent document	str	spaCy
201	Word	SpacyTRIACCIT		parent document	30	зрасу
				the character offset of the token		
201	word	  spacyTknAttrIdx		within the parent document	str	spaCy
201	word	SpacyTriiAcciTux		does the token consist of	Sti	spacy
201	word	ana ar Tha Atta Alaba		alphabetic characters?	bool	cnaCv
201	word	spacyTknAttrIsAlpha		does the token consist of ASCii	DOOI	spaCy
201	word	spacyTknAttrIsAscii		characters?	bool	spaCy
201	word	spacyTknAttrIsBracket		is the token a bracket?	bool	spaCy
201	word	spacyTknAttrIsCurrency		is the token a currency symbol?	bool	spaCy
201	word	spacyTknAttrIsDigit		does the token consist of digits?	bool	spaCy
		spacyTknAttrIsLeftPunc		is the token a left punctuation		
201	word	t		mark, e.g. "(" ?	bool	spaCy
201	word	spacyTknAttrIsLower		is the token in lowercase?	bool	spaCy
204	word	are a gradules. A black to C		is the token and of ware bullet 2	basi	cn c C
	word	spacyTknAttrIsOov		is the token out-of-vocabulary?	bool	spaCy
201	word	spacyTknAttrIsPunct		is the token punctuation?	bool	spaCy
201	word	spacyTknAttrIsQuote		is the token a quotation mark?	bool	spaCy
		spacyTknAttrIsRightPun		is the token a right punctuation		
201	word	ct		mark, e.g. ")" ?	bool	spaCy
201	word	and are Three 7 to the Tar Court Total		door the taken and a sector sec	basi	cnaC:
ZU1	word	spacyTknAttrIsSentEnd		does the token end a sentence?	bool	spaCy
201	word	spacyTknAttrIsSentStar t		does the token start a sentence?	bool	spaCy

Version: 0.9.8 Page: 9/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
				does the token consist of		
201	word	spacyTknAttrIsSpace		whitespace characters?	bool	spaCy
201	word	spacyTknAttrIsStop		is the token part of a "stop list"?	bool	spaCy
201	word	spacyTknAttrIsTitle		is the token in titlecase?	bool	spaCy
201	word	spacyTknAttrIsUpper		is the token in uppercase?	bool	spaCy
				language of the parent		
201	word	spacyTknAttrLang_		document's vocabulary	str	spaCy
				the leftmost token of this token's		
201	word	spacyTknAttrLeftEdge		syntactic descendants	str	spaCy
				base form of the token, with no		
201	word	spacyTknAttrLemma_		inflectional suffixes	str	spaCy
201	word	spacyTknAttrLex		the underlying lexeme	str	spaCy
				sequential ID of the token's		
				lexical type, used to index into		
201	word	spacyTknAttrLexId		tables, e.g. for word vectors	str	spaCy
				does the token resemble an		
201	word	spacyTknAttrLikeEmail		email address?	bool	spaCy
				does the token represent a		
201	word	spacyTknAttrLikeNum		number?	bool	spaCy
						, ,
201	word	spacyTknAttrLikeUrl		does the token resemble a URL?	bool	spaCy
201	word	spacyTknAttrLower_		lowercase form of the token text	str	spaCy
201	word	spacyTknAttrMorph		morphological analysis	str	spaCy
				the token's norm, i.e. a		
				normalized form of the token		
201	word	spacyTknAttrNorm_		text	str	spaCy
201	word	spacyTknAttrOrth_		verbatim text content	str	spaCy
				coarse-grained part-of-speech		
201	word	spacyTknAttrPos_		from the Universal POS tag set	str	spaCy
				a length-N substring from the		
201	word	spacyTknAttrPrefix_		start of the token	str	spaCy
				smoothed log probability		
201	word	spacyTknAttrProb		estimate of token's word type	str	spaCy
				sequential ID of the token's		
				lexical type, used to index into		
201	word	spacyTknAttrRank		tables, e.g. for word vectors	str	spaCy
			1			' '
				the rightmost token of this		
201	word	spacyTknAttrRightEdge		token's syntactic descendants	str	spaCy
				the sentence span that this token		' '
	word	spacyTknAttrSent		is a part of	str	spaCy

Version: 0.9.8 Page: 10/11 Date: 27.09.2022

lvl	container	tag	opt	description	type	source
				a scalar value indicating the		
				positivity or negativity of the		
201	word	spacyTknAttrSentiment		token	str	spaCy
				transform of the token's string to		
201	word	spacyTknAttrShape_		show orthographic features	str	spaCy
201	Word	spacy ininectionape_		length-N substring from the end	301	зрасу
201	word	spacyTknAttrSuffix_		of the token	str	spaCy
_	word	spacyTknAttrTag_		fine-grained part-of-speech	str	spaCy
201	Word	spacy immediag_		the token's slice of the parent	361	эрасу
201	word	spacyTknAttrTensor		doc's tensor	str	spaCy
_	word	spacyTknAttrText		verbatim text content	str	spaCy
201	Word	spacy inflacti text		text content, with trailing space	311	зрасу
201	word	spacyTknAttrTextWithWs		character if present	str	spaCy
201	Word	spacy i knacci i excwichws		the vocab object of the parent	Sti	spacy
201	word	spacyTknAttrVocab		doc	str	spaCy
		spacyTknAttrWhitespace				<u> </u>
201	word	spacy immicel mileespace		trailing space character if present	str	spaCy
	word	tableCellNo		cell number in row	int	TET
201	word	tableCellSpan		cel span	int	TET
201	word	tableNo		table number in document	int	TET
201	word	tableRowNo		row number in table	int	TET
201	word	text		word text	str	TET
201	word	type		line type	str	TET, TL
				x coordinate of the upper right		
201	word	urx		corner	float	TET
201	word	wordNo		word number in document	int	TET
201	word	wordNoLine		word number in line	int	TET
201	word	wordNoPage		word number in page	int	TET
201	word	wordNoPara		word number in paragraph	int	TET

Version: 0.9.8 Page: 11/11 Date: 27.09.2022