lvl	container	tag	opt	description	type	source
10	document	document		container for document	root	TET
11	document	config	Х	container for configuration	dict	TET
				created at (DD.MM.YY		
11	document	createdAt		HH:MM:SS)	str	system
11	document	createdBy		created by module	str	system
11	document	fonts	Х	container for fonts	list	TET
				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
	document	noLines		number lines in document	int	TET
11	document	noLinesFooter	Х	number footer lines	int	LT
	document	noLinesHeader	Х	number header lines	int	LT
	document	noLinesToc	X	number toc lines	int	LT
	document	noListsBullet	X	number bulleted lists	int	LT
	document	noListsNumber	X	number numbered lists	int	LT
	document	noPages		number pages in document	int	TET
11	document	norages		number pages in document	1110	161
11	document	noParas		number paragraphs in document	int	TET
11	document	HOFALAS		number paragraphs in document	1110	1 - 1
11	document	noSentences	,	number sentences in document	int	cnaCv
	document	noTables	Х	number tables in document	int int	spaCy LT
	document	noWords		number words in document	int	TET
	document			container for pages	list	TET
		pages	.,	· -		TET
	document	params	Х	container for parameters	dict	
20	config	config	X	container for configuration	root	TET
		1.		include the configuration data in		
21	config	jsonInclConfig		the JSON file	bool	TET
	_			include the font data in the JSON		
21	config	jsonInclFonts		file	bool	TET
				include the heading data in the		
21	config	jsonInclHeading		JSON file	bool	TET
				include the bulleted list data in		
21	config	jsonInclListBullet		the JSON file	bool	TET
				include the numbered list data in		
21	config	jsonInclListNumber		the JSON file	bool	TET
				include the parameters in the		
				JSON file	bool	TET
21	config	jsonInclParams		JON THE		
21	config	jsonInclParams		include the table data in the		
	config	jsonInclParams jsonInclTable			bool	TET
21				include the table data in the		TET TT
21 21	config config	jsonInclTable parser		include the table data in the JSON file container for module parser	bool dict	TT
21 21	config	jsonInclTable		include the table data in the JSON file	bool	
21 21	config config	jsonInclTable parser		include the table data in the JSON file container for module parser	bool dict	TT
21 21 21	config config	jsonInclTable parser		include the table data in the JSON file container for module parser container for module tokenizer	bool dict	TT
21 21 21	config config config	jsonInclTable parser tokenizer		include the table data in the JSON file container for module parser container for module tokenizer Improves the readability of the	bool dict dict	TET
21 21 21 22	config config config	jsonInclTable parser tokenizer		include the table data in the JSON file container for module parser container for module tokenizer Improves the readability of the	bool dict dict	TET
21 21 21 22	config config config config	jsonInclTable parser  tokenizer  jsonIndent		include the table data in the JSON file container for module parser container for module tokenizer Improves the readability of the JSON file	bool dict dict int	TT TET TET

Version: 0.9.8 Page: 1/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
22	config	ltFooterMaxLines		maximum number of footers	int	TET
				Maximum Levenshtein distance		
22	config	ltHeaderMaxDistance		for a header line	int	TET
	J 5					
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	p		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
				minimum number of entries to		
22	config	ltListBulletMinEntries		determine a bulleted list	int	TET
				file with rules to determine the		
22	config	ltListBulletRuleFile		bulleted lists	bool	TET
		ltListBulletToleranceL		tolerance of vertical indentation		
22	config	lx		in percent	bool	TET
				·		
				if it is set to true, the regular		
		ltListNumberFileInclRe		expression for the numbered list		
22	config	gexp		is included in the JSON file	bool	TET
				minimum number of entries to		
22	config	ltListNumberMinEntries		determine a numbered list	bool	TET
				file with rules to determine the		
22	config	ltListNumberRuleFile		numbered lists	bool	TET
		ltListNumberToleranceL		tolerance of vertical indentation		
22	config	lx		in percent	bool	TET
				·		
				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
	J			maximum number of pages for		
				the search of the TOC (from the		
22	config	ltTocLastPage		beginning)	bool	TET
	<u> </u>					
22	config	ltTocMinEntries		minimum number of TOC entries	bool	TET
				ignore the tokens which are		
						1

Version: 0.9.8 Page: 2/11 Date: 23.09.2022

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
	config	spacyTknAttrCluster		brown cluster ID	bool	TET
22	config	spacyTknAttrDep_		syntactic dependency relation	bool	TET
	config	spacyTknAttrDoc		the parent document	bool	TET
	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
	8	2F 413/ 1111111111111111111111111111111111		pa. 6 6.7, 11 d. 17		
22	config	spacyTknAttrEntType_		named entity type	bool	TET
	56			the syntactic parent, or	2001	1
22	config	spacyTknAttrHead		"governor", of this token	bool	TET
	5511116	Space I ministerineau		the index of the token within the	2001	
22	config	spacyTknAttrI		parent document	bool	TET
	COTTING	Pracitiment		parent document	5001	161
				the character offset of the token		
วา	config	anagyTkn\++xTd		within the parent document	haal	TET
	comig	spacyTknAttrIdx		does the token consist of	bool	IEI
22	config	and artist to the sale of the			haal	TET
22	config	spacyTknAttrIsAlpha		alphabetic characters?	bool	TET

Version: 0.9.8 Page: 3/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
				does the token consist of ASCii	7,1	
22	config	spacyTknAttrIsAscii		characters?	bool	TET
	0					
22	config	spacyTknAttrIsBracket		is the token a bracket?	bool	TET
		25-00/1111100112011001				1
22	config	spacyTknAttrIsCurrency		is the token a currency symbol?	bool	TET
		spacy immeer is carrency		is the token a carrency symbol.	500.	
22	config	spacyTknAttrIsDigit		does the token consist of digits?	bool	TET
		spacyTknAttrIsLeftPunc		is the token a left punctuation	500.	
22	config	t		mark, e.g. "(" ?	bool	TET
	config	spacyTknAttrIsLower		is the token in lowercase?	bool	TET
	comig	Spacy Infinite I I Shower		is the token in lower case.	5001	
22	config	spacyTknAttrIsOov		is the token out-of-vocabulary?	bool	TET
	config	spacyTknAttrIsPunct		is the token punctuation?	bool	TET
	Coming	spacyTriacciTsFuncc		is the token punctuation:	5001	161
22	config	spacyTknAttrIsQuote		is the token a quotation mark?	bool	TET
	COIIIIg			is the token a right punctuation	5001	1 - 1
22	config	spacyTknAttrIsRightPun ct		mark, e.g. ")" ?	bool	TET
~~	COIIIIB				וטטו	161
22	config	spacyTknAttrIsSentEnd		does the token end a sentence?	bool	TET
	Cornig			does the token end a sentence:	DOOI	ILI
22	config	spacyTknAttrIsSentStar t		does the token start a sentence?	bool	TET
	Corning			does the token consist of	וטטטו	161
22	config	and arrellen Attached			bool	TET
	config	spacyTknAttrIsSpace		whitespace characters?	וטטטו	161
22	config	ano ar Tirn Attact act on		is the token part of a "stop list"?	bool	TET
	config	spacyTknAttrIsStop		is the token in titlecase?	bool	TET
		spacyTknAttrIsTitle				TET
	config	spacyTknAttrIsUpper		is the token in uppercase?	bool	161
22	config			language of the parent	bool	TET
	config	spacyTknAttrLang_		document's vocabulary	bool	TET
		m1 2.1. 7 C. 7 1		the leftmost token of this token's		T
22	config	spacyTknAttrLeftEdge		syntactic descendants	bool	TET
				base form of the token, with no	1 1	<b>T</b> CT
	config	spacyTknAttrLemma_		inflectional suffixes	bool	TET
22	config	spacyTknAttrLex	$\vdash$	the underlying lexeme	bool	TET
				assumptial ID of the Assumption		
				sequential ID of the token's		
	<b>-</b>	m1		lexical type, used to index into	L I	
22	config	spacyTknAttrLexId	$\vdash$	tables, e.g. for word vectors	bool	TET
	<b>-</b>	m1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2		does the token resemble an	L I	
22	config	spacyTknAttrLikeEmail		email address?	bool	TET
	c.	_, _, _,		does the token represent a		
22	config	spacyTknAttrLikeNum		number?	bool	TET
22	config	spacyTknAttrLikeUrl		does the token resemble a URL?	bool	TET
	config	spacyTknAttrLower_	1	lowercase form of the token text		TET
22	config	spacyTknAttrMorph		morphological analysis	bool	TET

Version: 0.9.8 Page: 4/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
				the token's norm, i.e. a		
				normalized form of the token		
22	config	spacyTknAttrNorm_		text	bool	TET
	config	spacyTknAttrOrth_		verbatim text content	bool	TET
				coarse-grained part-of-speech		
22	config	spacyTknAttrPos_		from the Universal POS tag set	bool	TET
	<u> </u>			a length-N substring from the		
22	config	spacyTknAttrPrefix_		start of the token	bool	TET
				smoothed log probability		
22	config	spacyTknAttrProb		estimate of token's word type	bool	TET
	5511118	Spacy Immeetities		estimate of token's word type	500.	
				sequential ID of the token's		
				lexical type, used to index into		
22	config	spacyTknAttrRank		tables, e.g. for word vectors	bool	TET
	COIIIIg	Spacyiniaccinalia		tables, e.g. for word vectors	וטטו	
				the rightmost token of this		
22	config				bool	TET
	config	spacyTknAttrRightEdge		token's syntactic descendants	וטטטו	IEI
22	<b>£</b> :			the sentence span that this token	la a a l	<b>T</b> CT
22	config	spacyTknAttrSent		is a part of	bool	TET
				a scalar value indicating the		
	<b>.</b> .			positivity or negativity of the		
22	config	spacyTknAttrSentiment		token	bool	TET
				transform of the token's string to		
22	config	spacyTknAttrShape_		show orthographic features	bool	TET
				length-N substring from the end		
	config	spacyTknAttrSuffix_		of the token	bool	TET
22	config	spacyTknAttrTag_		fine-grained part-of-speech	bool	TET
				the token's slice of the parent		
	config	spacyTknAttrTensor		doc's tensor	bool	TET
22	config	spacyTknAttrText		verbatim text content	bool	TET
				text content, with trailing space		
22	config	spacyTknAttrTextWithWs		character if present	bool	TET
				the vocab object of the parent		
22	config	spacyTknAttrVocab		doc	bool	TET
		spacyTknAttrWhitespace				
22	config	_		trailing space character if present	bool	TET
30	font	font	Х	container for font	root	TET
31	font	embedded		embedded	bool	TET
31	font	fontNo		font number in document	int	TET
31	font	fullName		font full name	str	TET
31	font	id		font identification	str	TET
31	font	italicAngle		font italic angle	float	TET
31	font	name		font name	str	TET
	font	type		font type	str	TET
	font	weight		font weight	float	TET
	page	page	1	container for page	root	TET

Version: 0.9.8 Page: 5/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
41	page	lineNoFirst		number of first line in page	int	TET
41	page	lineNoLast		number of last line in page	int	TET
41	page	lines		container for lines	list	TET
41	page	pageNo		page number in document	int	TET
				number of first paragraph in		
41	page	paraNoFirst		page	int	TET
41	page	paraNoLast		number of last paragraph in page	int	TET
41	page	paras		container for paragraphs	list	TET
41	page	sentenceNoFirst		number of first sentence in page	int	spaCy
41	page	sentenceNoLast		number of last sentence in page	int	spaCy
	page	sentences		container for sentences	list	spaCy
	page	wordNoFirst		number of first word in page	int	TET
	page	wordNoLast		number of last word in page	int	TET
	params	params	X	container for parameters	root	TET
	params	parser		container for module parser	dict	TT
		11 22		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
52	params	fileNameOrig		original file name	str	TET
52	params	ltFooterRequired		footer determination required	bool	TET
52	params	ltHeaderRequired		header determination required	bool	TET
52	params	ltHeadingRequired		heading determination required	bool	TET
52	params	ltListBulletRequired		bulleted list determination required	bool	TET
52	params	ltListNumberRequired		numbered list determination required	bool	TET
52	params	ltTableRequired		table determination required	bool	TET
52	params	ltTocRequired		TOC determination required	bool	TET
52	params	noPdfPages		number PDF pages	int	TET
50	narams	tokenizer		container for module tokenizer	dict	TET
100	params	line		container for lines	1	TET
100		lineNo		line number in document	root	TET
101						TET
		lineNoPage		line number in page	int	
101		lineNoPara		line number in paragraph x coordinate of the lower left	int	TET
101		llx		corner	float	TET
101	line	pageNo		page number in document	int	TET

Version: 0.9.8 Page: 6/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
101	line	paraNo		paragraph number in document	int	TET
101		<u> </u>		paragraph number in page	int	TET
101	iiie	paraNoPage		paragraph number in page	IIIC	IEI
101	line	sentenceNo		sentence number in document	int	spaCy
101	line	tableCellNo		cell number in row	int	TET
101	line	tableCellSpan		cel span	int	TET
101	line	tableNo		table number in document	int	TET
101	line	tableRowNo		row number in table	int	TET
101	line	text		line text	str	TET
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
		WordingEase		word number in paragraph of		
101	line	wordNoParaFirst		first word in this line	int	TET
	para	para		container for paragraph	root	TET
110	para	para		Paragraph	1000	
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
	para	tableCellSpan		cel span	int	TET
	para	tableNo		table number in document	int	TET
	para	tableRowNo		row number in table	int	TET
	•					TET,
111	para	text		paragraph text	str	spaCy
	<u> </u>			number of first word in		' '
111	para	wordNoFirst		paragraph	int	TET
	1			number of last word in		•
111	para	wordNoLast		paragraph	int	TET
	para	words		container for words	list	TET
	sentence	sentence		container for sentences	root	TET
120	Schlence	Jentenee			1000	
121	sentence	sentenceNo		sentence number in document	int	TET
	sentence			sentence number in page	int	TET
121	SCHICHICE	sentenceNoPage		Sentence number in page	1110	161
121	sentence	sentenceNoPara		sentence number in paragraph	int	TET

Version: 0.9.8 Page: 7/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
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		
121	sentence	wordNoLast		sentence	int	TET
	word	word	1	container for word	root	TET
	word	font	x	font identification	str	TET
	word	lineNo		line number in document	int	TET
	word	lineNoPage		line number in page	int	TET
	word	_		page number in document		TET
201	word	pageNo		page number in document	int	161
201	word	paraNo		paragraph number in document	int	TET
201	word	sentenceNo		sentence number in document	int	spaCy
201	word	size	х	font size	float	TET
201	word	spacyTknAttrCluster		brown cluster ID	str	spaCy
201	word	spacyTknAttrDep_		syntactic dependency relation	str	spaCy
201	word	spacyTknAttrDoc		the parent document	str	spaCy
201	word	spacyTknAttrEntIob_		IOB code of named entity tag	str	spaCy
				knowledge base ID that refers to		
				the named entity this token is a		
201	word	spacyTknAttrEntKbId_		part of, if any	str	spaCy
201	word	spacyTknAttrEntType_		named entity type	str	spaCy
				the syntactic parent, or		577
201	word	spacyTknAttrHead		"governor", of this token	str	spaCy
		ar or of a second		the index of the token within the		577
201	word	spacyTknAttrI		parent document	str	spaCy
		Spacy IIIIII		parent decament	30.	spacy
				the character offset of the token		
201	word	spacyTknAttrIdx		within the parent document	str	spaCy
	Word	Spacy Immeet ran		does the token consist of	301	spacy
201	word	spacyTknAttrIsAlpha		alphabetic characters?	bool	spaCy
201	Word	spacy ininterishipha		does the token consist of ASCii	5001	spacy
201	word	spacyTknAttrIsAscii		characters?	bool	spaCy
201	word	spacyiniactiisastii		Characters:	DOOI	зрасу
201	word	anagyTkn\++vTaDwaalra+		is the token a bracket?	haal	cnaCy
ZU1	vvOiu	spacyTknAttrIsBracket	1	is the token a blacket!	bool	spaCy
201	word	spacyTknAttrIsCurrency		is the token a currency symbol?	bool	spaCy
ZU1	VVOIU	BPacy Innacct Iscurrency		is the token a currency symbol!	וטטו	зрасу
201	word	  spacyTknAttrIsDigit		does the token consist of digits?	bool	spaCy
ZU1	vv Oi U		+	_	5001	зрасу
201	word	spacyTknAttrIsLeftPunc		is the token a left punctuation	hasi	cnoC:
	word	t and assemble by the Tall assert	1	mark, e.g. "(" ?	bool	spaCy
2UI	word	spacyTknAttrIsLower	1	is the token in lowercase?	bool	spaCy
201		m1 2 = 2		is the taken a last of the last of	L 1	
	word	spacyTknAttrIsOov	1	is the token out-of-vocabulary?		spaCy
201	word	spacyTknAttrIsPunct		is the token punctuation?	bool	spaCy

Version: 0.9.8 Page: 8/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
201	word	spacyTknAttrIsQuote		is the token a quotation mark?	bool	spaCy
201	Word			is the token a right punctuation	5001	зрасу
201	word	spacyTknAttrIsRightPun ct		mark, e.g. ")" ?	bool	spaCy
201	word			mark, e.g. / :	5001	зрасу
201	word	spacyTknAttrIsSentEnd		does the token end a sentence?	bool	spaCy
		spacyTknAttrIsSentStar				
201	word	t		does the token start a sentence?	bool	spaCy
				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
	1			the leftmost token of this token's		
201	word	spacyTknAttrLeftEdge		syntactic descendants	str	spaCy
204	1			base form of the token, with no		
	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
	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	ana artikn A + + - Droch		smoothed log probability	ctr	cnaCu
201	word	spacyTknAttrProb		estimate of token's word type	str	spaCy

Version: 0.9.8 Page: 9/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
				sequential ID of the token's		
				lexical type, used to index into		
201	word	spacyTknAttrRank		tables, e.g. for word vectors	str	spaCy
				the rightmost token of this		
201	word			the rightmost token of this	ctr	cno Cu
201	word	spacyTknAttrRightEdge		token's syntactic descendants	str	spaCy
201	word	212 0 21 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1		the sentence span that this token		snaCv
201	word	spacyTknAttrSent		is a part of	str	spaCy
				a scalar value indicating the positivity or negativity of the		
201	word	ana ar Tha Att a Continont		token	ctr	cnaCv
201	word	spacyTknAttrSentiment		token	str	spaCy
				transform of the token's string to		
201	word	spacyTknAttrShape_		show orthographic features	str	spaCy
		ap and a summer and a summer a		length-N substring from the end		
201	word	spacyTknAttrSuffix_		of the token	str	spaCy
	word	spacyTknAttrTag_		fine-grained part-of-speech	str	spaCy
		2500/11111001103		the token's slice of the parent		opu.c,
201	word	spacyTknAttrTensor		doc's tensor	str	spaCy
	word	spacyTknAttrText		verbatim text content	str	spaCy
		Ep at 2 at a state of the state		text content, with trailing space		
201	word	spacyTknAttrTextWithWs		character if present	str	spaCy
				the vocab object of the parent		-   -
201	word	spacyTknAttrVocab		doc	str	spaCy
		spacyTknAttrWhitespace				-   -
201	word	Spacy Immeet with cospace		trailing space character if present	str	spaCy
	word	tableCellNo		cell number in row	int	TET
	word	tableCellSpan		cel span	int	TET
	word	tableNo		table number in document	int	TET
	word	tableRowNo		row number in table	int	TET
201	word	text		word text	str	TET
	word	type		line type	str	TET, TL
	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
	heading	heading	Х	container for headings	root	LT
311		TODO				
	lists bullet	listsBullet	Х	container for bulleted lists	root	LT
	lists bullet	lists		container for bulleted lists	list	LT
322	list	entries		container for list entries	list	LT
322		format		number format	str	LT
322	list	list		container for list	root	LT
322	list	listNo		list number in document	int	LT
322	list	noEntries		number entries in list	int	LT
322	list	pageNoFirst		page number first entry	int	LT
322	list	pageNoLast		page number last entry	int	LT

Version: 0.9.8 Page: 10/11 Date: 23.09.2022

lvl	container	tag	opt	description	type	source
323	entry	entry		container for bulleted list entry	root	LT
323	entry	entryNo		entry number in list	int	LT
323	entry	lineNoPageFirst		first line in page	int	LT
323	entry	lineNoPageLast		last line in page	int	LT
323	entry	pageNo		page number in document	int	LT
323	entry	paraNo		paragraph number in page	int	LT
323	entry	text		list entry text	str	LT
330	lists number	listsNumber	Х	container for numbered lists	root	LT
331	lists number	lists		container for numbered lists	list	LT
332	list	entries		container for list entries	list	LT
332	list	format		bullet format	str	LT
332	list	list		container for list	root	LT
332	list	listNo		list number in document	int	LT
332	list	noEntries		number entries in list	int	LT
332	list	pageNoFirst		page number first entry	int	LT
332	list	pageNoLast		page number last entry	int	LT
332	list	regexp	Х	regular expression	str	LT
333	entry	entry		container for numbered list entry	root	LT
333	entry	entryNo		entry number in list	int	LT
333	entry	lineNoPageFirst		first line in page	int	LT
333	entry	lineNoPageLast		last line in page	int	LT
333	entry	pageNo		page number in document	int	LT
333	entry	paraNo		paragraph number in page	int	LT
333	entry	text		list entry text	str	LT
340	heading	toc	х	container for toc	root	LT
341	heading	entries		container for heading entries	list	LT
342	entry	ctxtLine1		1. context line	str	LT
342	entry	ctxtLine2		2. context line	str	LT
342	entry	ctxtLine3		3. context line	str	LT
342	entry	entry		container for heading entry	root	LT
342	entry	level		heading level	int	LT
342	entry	lineNoPage		line number in page	int	LT
342	entry	pageNo		page number in document	int	LT
	entry	regexp	х	regular expression	str	LT
	entry	text		heading text	str	LT

Version: 0.9.8 Page: 11/11 Date: 23.09.2022